

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Environmental, health, and safety assessment of
chemical alternatives during early process design:

The role of predictive modeling and streamlined techniques

GULNARA SHAVALIEVA

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Environmental, health, and safety assessment of chemical alternatives during early process design: The role of predictive modeling and streamlined techniques
GULNARA SHAVALIEVA
ISBN 978-91-7905-624-7

© GULNARA SHAVALIEVA, 2022.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr. 5090
ISSN 0346-718X

Department of Space, Earth and Environment
Division of Energy Technology
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Chalmers Reproservice
Gothenburg, Sweden 2022

Abstract

Industrial chemicals are important for many aspects of modern life, though they can be harmful to the environment and human health. Environmental or safety concerns identified during the early design and selection of chemicals could motivate choices as to safer alternatives and process setups. There is a growing interest in developing more rapid, and streamlined assessment methods to obtain a first indication of the potential impacts linked to the nature and use of industrial chemicals.

This work applies predictive modeling and streamlined techniques to estimate the potential environmental, health, and safety hazards associated with specific chemical structures. The assessment is performed during the design and selection of promising candidates for a particular process as part of the computer-aided molecular design (CAMD) and process setup. The case of phase-change solvents used for post-combustion carbon capture is examined. Furthermore, the refinement of predictive models through the incorporation of knowledge already existing in the field (prior knowledge) is investigated. A procedure for knowledge extraction from scientific articles that applies text mining is proposed.

The results show that incorporating impact assessment criteria into the CAMD facilitates the molecular design by enriching the Pareto front of candidates. The use of predictive models that estimate molecular properties, such as acute aquatic toxicity, bioconcentration, and persistency are found to support the identification of the optimal solvents for CO₂ capture. Given the role of sustainability-related properties in tasks such as CAMD, the improved performance and the interpretability of the aquatic toxicity predictive models developed here and using prior knowledge are important. The process level assessment of the phase-change solvent systems indicated that phase-change solvent alternatives could provide benefits, not only in terms of reduced energy consumption but also lower impacts on human health and the environment. However, the degradation behaviors of these compounds should be properly assessed and controlled to ensure beneficial performances compared to conventional carbon capture solvents. Overall, predictive modeling and streamlined life-cycle assessments (LCAs), as well as environmental, health, and safety evaluation methods were revealed to be valuable for defining the critical aspects that influence the potential impacts of chemicals and in supporting decisions concerning the molecular and process designs.

Keywords: *Predictive modeling, knowledge extraction, LCA, EHS, CAMD, phase-change solvents, carbon capture, acute aquatic toxicity*

List of publications

This thesis is based on the following papers, which are referred to in the thesis according to their Roman numerals:

- I. Shavaliyeva, G., Papadopoulos, A.I., Badr, S., Seferlis, P., Papadokonstantakis, S., 2018, Sustainability assessment using local lazy learning: The case of post-combustion CO₂ capture solvents, *Proceedings of the Process Systems Engineering Conference, PSE 2018, 1-5 July 2018, San Diego, USA* (peer-reviewed), doi: 10.1016/B978-0-444-64241-7.50132-4.
- II. Papadopoulos, A.I., Shavaliyeva, G., Papadokonstantakis, S., Seferlis, P., Perdomo, F.A., Galindo, A., Jackson, G., Adjiman, C.S., 2020, An approach for simultaneous computer-aided molecular design with holistic sustainability assessment: Application to phase-change CO₂ capture solvents, *Computers & Chemical Engineering, Vol. 135*, 106769, doi: 10.1016/j.compchemeng.2020.106769.
- III. Shavaliyeva, G., Kazepidis, P., Papadopoulos, A.I., Seferlis, P., Papadokonstantakis, S., 2021, Environmental, Health and Safety Assessment of Post-combustion CO₂ Capture Processes with Phase-change solvents, *Sustainable Production & Consumption, Vol. 25*, p. 60-76 doi: 10.1016/j.spc.2020.07.015.
- IV. Shavaliyeva, P., Papadokonstantakis, S., Peters, 2022, G. Knowledge mining from scientific literature for acute aquatic toxicity: classification for hybrid predictive modeling, *Proceedings of the 32nd European Symposium on Computer-Aided Engineering (ESCAPE32), June 12-15, 2022, Toulouse, France* (peer-reviewed). Accepted for publication.
- V. Shavaliyeva, G., Papadokonstantakis, S., Peters, G. 2022, Prior knowledge for predictive modeling: the case of acute aquatic toxicity, Submitted to *Journal of Chemical Information and Modeling*.

Authors' contributions

Gulnara Shavaliyeva is the principal author of **Papers I, III, IV, and V**. In collaboration with its first-listed author, she made an equivalent overall contribution to the design and execution of **Paper II**, including particular contributions to the development of the automated sustainability assessment procedure and its integration into computer-aided molecular design. Professor Stavros Papadokonstantakis was the supervisor of all the papers and contributed to discussions and editing of all the papers. Principal Researcher Dr. Athanasios I. Papadopoulos was the corresponding author for **Paper II** and contributed flowsheets and editing for **Papers I and III**. Sara Badr contributed to **Paper I** with editing and discussion. Professor Gregory Peters contributed to the discussions and editing of **Papers IV and V**.

Related publications not included in the thesis:

- Shavaliyeva, G., Papadokonstantakis, S., Kazepidis, P., Papadopoulos, A.I., Seferlis, P., 2018, Sustainability analysis of phase-change solvents for post-combustion CO₂ capture, *Chemical Engineering Transactions*, Vol. 76, p. 1045-1050, doi: 10.3303/CET1976175.
- Papadopoulos, A.I., Perdomo, F.A., Tzirakis, F., Shavaliyeva, G., Tsivintzelis, I., Kazepidis, P., Nessi, E., Papadokonstantakis, S., Seferlis, P., Galindo, A., Jackson, G., Adjiman, C.S., 2021, Molecular engineering of sustainable phase-change solvents: From digital design to scaling-up for CO₂ capture, *Chemical Engineering Journal*, Vol. 420, 127624.
- Shavaliyeva, G., Postacchini, P., Papadokonstantakis, S., 2020, Prediction of sustainability-related properties: data science methods with incorporated prior knowledge, *Proceedings of the 30th European Symposium on Computer-Aided Chemical Engineering*, Vol. 47, 1st Edition (peer-reviewed).

Acknowledgments

It is often said that the people around you shape who you are. I hope that it is true because amazing people surrounded me during my PhD, and they definitely expanded my worldview and motivated me to grow. I am deeply grateful to everyone who made my PhD time such an exciting journey.

First of all, I would like to express my sincere gratitude to my supervisor, Stavros Papadokonstantakis, for all his guidance and invaluable contributions to my work. Thank you for your care, trust, and enthusiasm. Your ability to find hundreds of possible solutions to every problem fascinates me all the time! I would also like to thank Greg Peters for valuable feedback and discussions that helped to improve my work.

I want to thank my ROLINCAP project partners for their excellent collaboration. I am particularly grateful to Athanasios Papadopoulos for all his knowledge, scientific experience, and assistance during our joint work. I would like to offer my special thanks to Sara Badr for her generous support during my first year of the PhD. I am also very grateful for the financial support provided by the European Union's Horizon 2020 Research and Innovation program.

Thank you to my IESA research group for all the thought-provoking discussions and valuable comments on my work. Thank you, colleagues, for creating a great workplace. To Christian, Holger, and Alla, thank you for your support and also for being great friends. Special thanks to Marie and Katarina for always helping me out. Thanks to my old officemates, Johan and Sofie, for being such delightful company in the office during my PhD years. To my new officemates, Munavara and Kushagra: please water the plants and do decorate the office at Christmastime.

I would also like to thank my Master's thesis students Céline Levêque, Thibaud Rosin, Maja Gustafsson and Pietro Postacchini for their excellent work.

I thank my family and friends for always being there for me. Thank you, Henrik, for being my greatest support throughout all these years. Your belief in me, love, and cheerfulness make me a better person.

Gulnara Shavalieva, Gothenburg, February 2022

Abbreviations and nomenclature

Abbreviations

AMP	2-Amino-2-methyl-1-propanol
BCF	Bioconcentration factor
BDA	1,4-Butanediamine
CAMD	Computer-aided molecular design
CCS	Carbon capture and storage
CED	Cumulative energy demand
DBA	Di-n-Butylamine
DEEA	N,N-Diethylethanolamine
DECA	N,N-Diethylcyclohexylamine
DETA	Diethylenetriamine
DMCA	Dimethylcyclohexylamine
DPA	Dipropylamine
ECA	N-Ethylcyclohexylamine
EHS	Environmental, health and safety
FF	Far field
GC	Group contribution
GWP	Global warming potential
HEPTA	Heptanamine
HEX	Heat Exchanger
HEXA	Hexylamine
HSS	Heat-stable salts
IDHL	Immediately dangerous to life or health
kNN	k-Nearest Neighbors
LCA	Life-cycle analysis
MAPA	N-Methyl-1,3-diaminopropane
MCA	Methylcyclohexylamine
MEA	Monoethanolamine
ML	Machine learning
MMEA	Monomethylethanolamine
MoA	Mode of action
MR	Molar refractivity
MSDS	Material safety data sheet
MW	Molecular weight
NDELA	N-Nitrosodiethanolamine
NF	Near field
OEL	Occupational exposure limit
P	Persistency
PEL	Permissible exposure limit

PoD	Potential of danger
PKM	Prior knowledge-based model
QSA(P)(A)R	Quantitative structure–activity(property)-(activity) relationship
RED	Relative energy difference
SBPA	N-sec-Butyl-n-propylamine
SMILE	Simplified Molecular Input Line Entry
STEL	Short-term exposure
S1N	Cyclohexylpropane-1,3-diamine
TETA	Triethylene tetramine
TPSA	Topological polar surface area
TWA	Time-weighted average

Nomenclature

AATSC0p	Average centered Broto-Moreau autocorrelation - lag 0 / weighted by polarizabilities (molecular descriptor)
EC50	Median effect concentration, the concentration of a substance producing a certain effect in 50% of the sample population
ETA_Alpha	Sum of the alpha-values of all non-hydrogen vertices of a molecule (molecular descriptor)
ETA_dEpsilon_A	A measure of the contribution of unsaturation and electronegative atom count (molecular descriptor)
ETA_EtaP_B	Branching index relative to molecular size (molecular descriptor)
GATS1i	Geary autocorrelation - lag 1 / weighted by first ionization potential (molecular descriptor)
GATS1p	Geary autocorrelation - lag 1 / weighted by polarizabilities (molecular descriptor)
LC50	Lethal concentration, the concentration of a substance that kills 50% of the sample population
Log P	Octanol–water partition coefficient
maxHBint2	Maximum E-State descriptors of strength for potential hydrogen bonds of path length 2 (molecular descriptor)
Mi	Mean first ionization potentials (scaled on carbon atom) (molecular descriptor)
Pareto front	Set of non-dominated optimal solutions from a multi-objective perspective
Predictive modeling	Technique to predict untested behaviors using mathematical and computational methods
Prior knowledge	Knowledge existing in the field
SHBd	Sum of E-States for (strong) hydrogen bond donors (molecular descriptor)
Streamlined techniques	Simplified (short-cut) techniques applying estimations when data are missing and focusing on the main aspects of the system

Table of contents

Abstract	III
List of publications.....	V
Acknowledgments.....	VII
Abbreviations and nomenclature.....	IX
Table of contents	XI
1 Introduction	1
1.1 Objective & Scope	1
1.2 Outline of the thesis	4
2 Early assessment of chemicals and process designs.....	5
2.1 Impact assessment during early process design.....	5
2.2 Estimation of missing data	6
2.3 Phase-change solvents for CO ₂ capture	8
2.4 Contribution of this work.....	10
3 Methods	13
3.1 Impact assessment at the substance level.....	13
3.2 Process level assessment.....	17
3.3 Predictive modeling	22
4 Results and Discussion	29
4.1 Role of impact assessment in the design of safer chemicals.....	29
4.2 Prior knowledge for predictive modeling	32
4.3 Process level impact assessment	40
5 Conclusions and outlook	49
5.1 Main conclusions	49
5.2 Future work	50
References	52

1 Introduction

In 2019, the European Commission (EC) initiated the European Green Deal (EGD)¹ to tackle imminent threats related to climate change and environmental degradation. One of the objectives of the EGD is to achieve zero pollution in air, water and soil by Year 2050. The main goal of the Chemicals Strategy as part of the EC's zero-pollution ambition is to protect human health and the environment by transitioning to the use of chemicals that are sustainable by design².

The “benign-by-design” concept for developing less-hazardous materials is already used by the pharmaceutical and chemical industries. Behind this concept lie the 12 Principles of Green Chemistry, which outline actions to minimize the use or production of toxic compounds and waste during synthesis, processing, and analysis of products³. The protection of human health and the environment means that the focus is on the potentially toxic activities of molecules and materials during development for the market and use⁴. However, despite our constant exposure to existing chemical substances, safety data are missing for many of these agents⁵. Apart from the existing compounds, numerous chemical structures are continuously being developed for use as catalysts, solvents, working fluids etc. Evaluation of their potential effects on human health and the environment should be required before their synthesis and use. In advanced computer-aided molecular design (CAMD) approaches, such molecules are first synthesized *in silico*, which allows for many molecular structures to be screened and assessed with respect to their technical and sustainability profiles, prior to the selection of candidates for experimental testing. It has become clear that the design phase has more freedom to influence critical decisions at an early stage. Rapid evaluation of the environmental, health, and safety (EHS) aspects of the molecular structures early during the chemical's design and selection phases might help to direct resources and effort away from the candidates that are linked to negative environmental or safety concerns⁶. Thus, there is growing interest in developing quicker, more streamlined assessment methods to obtain a first indication of the potential impacts already during the molecular design period.

Early-stage assessments of processes that employ new chemicals might ensure safer chemicals, operating conditions, and process designs prior to significant financial and physical assets being invested in large-scale systems development.

1.1 Objective & Scope

The purpose of this thesis is to perform early impact assessments of novel chemicals using predictive modeling and streamlined techniques, in order to promote the selection of more- sustainable (greener) chemical alternatives. The work addresses the following research questions:

1. In what way does the impact assessment of new molecular structures during molecular design contribute to the selection of safer and more environmentally benign chemicals? (**Paper II**)

Answering the first research question involves the development of methods to incorporate data and predictive models regarding molecular properties relevant to environmental impacts into computer-aided screening approaches (e.g., CAMD). The second research question addresses the issue of missing molecular property data for newly designed molecular structures:

2. How can the data gaps that exist for novel molecular structures be filled by predictive modeling and streamlined techniques? (**Papers I, II, and V**)

Related to the second question is:

3. How can the knowledge from scientific articles be processed more efficiently to assist the predictive modeling? (**Paper IV**)

When applied to industrial process, the behaviors of chemicals are influenced by the operational conditions of the process, as well as by other compounds present in the system. Likewise, the environmental impact of a single molecule can be affected considerably by the industrial context in which it is used. Even though the industrial context is known, it is practically impossible to include in the molecular design all the process-relevant conditions under which the substance will be applied. Thus, the impacts of the molecules during their use might differ compared to the impact of just a single substance. Therefore, the fourth research question seeks to investigate the challenges associated with the use of the newly designed molecules:

4. What are the potential behaviors and impacts of the new chemical substances when applied in the process, and how do the predictive modeling and streamlined techniques impact the assessment? (**Paper III**)

To answer research questions 1, 2, and 4, phase-change solvents for post-combustion CO₂ capture or structurally similar molecules were considered. Research question 3 addresses the need for more-efficient data acquisition and processing within the broad field of interdisciplinary research. The acute aquatic toxicity domain relevant for the EHS hazard assessment of chemicals is used as an example.

The work considers two different levels of assessment: substance and process. The substance level comprises the predictive models and the design of the more-sustainable CO₂ capture phase-change solvents. The process level assessment evaluates the potential impact of a capture system that utilizes the selected phase-change solvents (Figure 1).

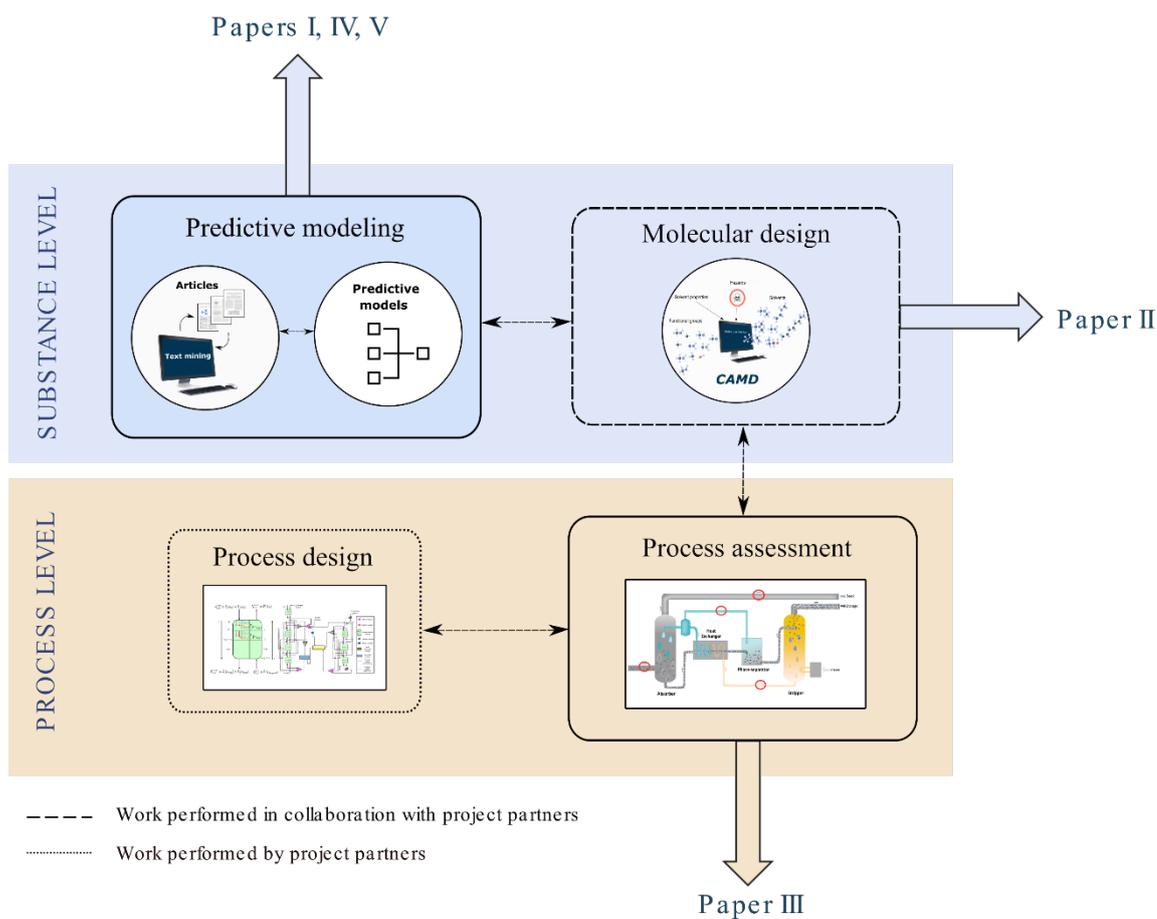


Figure 1. Scope of the work of this thesis.

Paper I examines the possibility of using a local lazy learning approach (k-nearest neighbors) to estimate the molecular properties (bioaccumulation, persistence, and acute aquatic toxicity), being a prerequisite for the impact assessment of chemicals.

Paper II presents a framework for the simultaneous assessments of life-cycle and EHS impacts (also referred to as the ‘impact assessment’) in CAMD. The effects of using impact assessment functions during CAMD compared to a procedure without such an assessment are also discussed.

Paper III evaluates the impacts associated with using selected phase-change solvents in the post-combustion CO₂ capture process. The work focuses on the potential behaviors of the solvents in relation to the operational conditions of the system and their roles in the impact assessment.

Paper IV proposes a semi-automated method for extracting useful sentences and data from published scientific articles. The method is applied to extract and classify relevant information (knowledge) about acute aquatic toxicity for subsequent use in predictive modeling.

Paper V investigates the feasibility of integrating knowledge into the predictive models. For the case of acute aquatic toxicity, the study evaluates the pure data science models and examines the use of prior knowledge at various points during the model

development (pre-analysis, during the construction of the model, and post-analysis approaches).

1.2 Outline of the thesis

The thesis summarizes the theoretical background relevant to the work and the key findings presented in the papers. The thesis is organized as follows:

Chapter 2 presents work in relation to the early assessment of chemicals. It introduces the data-driven methods for the impact assessment and provides background information on the CO₂ processes and phase-change solvent systems.

Chapter 3 describes the methods used in the work, including life-cycle assessment, EHS hazard assessment, semi-automated prior knowledge extraction, and the k-nearest neighbors data-mining approach.

Chapter 4 presents and discusses the key findings reported in the papers.

Chapter 5 lists the conclusions and suggestions for future work.

2 Early assessment of chemicals and process designs

This chapter provides background information relevant to the study. The role of the current work in terms of its contributions to the existing body of knowledge is also discussed.

2.1 Impact assessment during early process design

Life-cycle assessment (LCA) is a valuable method to estimate the environmental impacts of products, processes, and services. While the original methodology focused on existing systems, there is a need for a method that applies life-cycle principles to developing the new technologies⁷ and early process designs. The recently developed concept of prospective LCA refers to LCA studies that assess the technologies in an early development phase, so as to forecast the potential environmental consequences of their large-scale implementation⁸. Similar concepts of *ex ante* and anticipatory LCAs are being used to address the environmental impacts of technologies that are still in the R&D phase⁹. These types of early-phase studies are characterized by simplifications, conservative assumptions, and scenario analysis. Even though a few studies^{10,11} have presented more-general frameworks for conducting such LCA assessments, most early-stage LCAs have been applied to a specific domain¹²⁻¹⁶.

With additional criteria that take into account other environmental and social aspects, LCA indicators are often used to explore trade-offs between different impact categories¹⁷ and scenarios. Many sustainability assessment tools enable such an analysis based on process flowsheets. For example, the frameworks proposed by Azapagic and Perdan (2000)¹⁸, Gonzalez and Smith (2004)¹⁹, and Guillén-Gosálbez et al. (2008)²⁰ have been designed to assist industries with the implementation of more-sustainable practices. It is also common to integrate the LCA indices with a hazard assessment to allow evaluation of the EHS impacts of different industrial processes.

Frameworks created to estimate the EHS hazards during process design have been in place for decades. Classical qualitative methods, such as HAZOP (hazard and operability)²¹ and fault tree analysis (FTA)²², rely on the availability of a detailed plant layout²³. Simpler techniques that use just molecular property data and the laboratory stage of the process design have also been developed. The simpler quantitative methods are more suitable for the early-stage process assessments. While many of the details regarding the system maybe lacking, the choices (materials, equipment, operating conditions, etc.) made at this step can significantly restrict the degrees of freedom for later design stages. For instance, the Dow Fire/Explosion and Chemical Exposure²⁴ indices evaluate fire and explosion hazards and the acute toxicity risks of a chemical process using the physical and chemical properties of the materials and simplified process flow sheets²⁵. The Mond Fire, Explosion, and Toxicity Index²⁶ extends the Dow Fire/Explosion Index, to include

additional hazard considerations²⁵. The Safety Weighted Hazard Index (SWeHI) developed by Khan (2001)²⁵ identifies and ranks various hazards according to hazardous substances and operating conditions, applying a more-rigorous quantification of the impact of safety measures. The Integrated Inherent Safety Index (I2SI) method of Khan and Amyotte (2005)²⁷ refines the technique through the inclusion of environmental effects. The inherent safety concept, first introduced by Kletz (1976)²⁸ and subsequently extended by Edwards and Lawrence (1993)²⁹ and Heikkilä et al. (1999)³⁰, among others, strives to avoid and reduce hazards at the source instead of relying on safety control measures. Koller (2000)³¹ proposed an EHS method that used 11 effect categories (mobility, fire/explosion, acute toxicity, etc.) as a solution to ensure the flexibility of index methods and their applicability when data on the substances are missing. A set of effect categories represents each aspect (E, H, and S). Each effect category can be assessed using several physical or chemical properties, which are prioritized according to data quality. These priorities and the scoring methods have been updated by Sugiyama et al. (2008)³¹ and Banimostafa et al. (2012)³². Sugiyama et al. (2008) have also revised the environmental aspect by adding the short-term effects of accidental release. Patel et al. (2012)³³ have modified the approach to assessment by incorporating basic reaction mass balance information, along with data on raw material prices, greenhouse gas (GHG) emissions, and qualitative indicators.

Although the early-stage assessments are often streamlined and inherently uncertain^{34,35}, they help to provide novel insights³⁶ during the advancement of new materials and technologies^{35,37}. For instance, even though CO₂ capture is an important technology for mitigating climate change, the risks associated with the use of new advanced materials and techniques should be assessed before their adoption at a larger scale. **Paper III** aims to address this issue by exploring during the early design phase the key components of the CO₂ capture technology applying novel phase-change solvents.

2.2 Estimation of missing data

Despite the unquestionable benefits of early-stage assessments, the limited availability of data hinders their use. An additional challenge is posed by processes that entail the use of chemicals. The behaviors of chemicals, which are influenced by many factors, can be difficult to predict. Furthermore, there is often a lack of information on physical and chemical properties (e.g., boiling point, solubility, toxicity) and of data relevant to LCA (e.g., inventory data, characterization factors), especially for newer chemical structures that are not yet mass-produced. Thus, streamlined techniques supported by surrogate data, qualitative models, and regression equations that are valid for specific domains are used to perform simplified impact assessments³⁸.

2.2.1 Molecular properties data

The hazard assessment of chemical processes requires information on the molecular property data, which are traditionally obtained experimentally. Molecular properties are

evaluated through predictive modelling, to reduce the need for experimental testing and fill the gaps in the missing datasets. Predictive modeling is a technique that can be used to assess the future untested behaviors of chemicals. Computational techniques facilitate the selection and optimization of compounds for synthesis and for tests related to their physicochemical properties³⁹. Such computational approaches are based on theory, experimental data, and a combination of thereof. For example, thermodynamics or quantum mechanics can be used to make a prediction. Methods such as expert and rule-based systems, quantitative structure-activity(property) relationships (QSA(P)R)s³⁹, and “additivity-principle” approaches that include group contribution (GC) methods⁴⁰ can combine theoretical aspects with patterns retrieved from existing datasets.

Expert and rule-based systems (e.g., qualitative prediction of toxicity system, DEREK⁴¹) rely on knowledge that comprises facts and heuristics⁴². The rules, generated either manually or automatically, can comprise structural alerts and reaction rules and are straightforward to apply³⁹. Another approach, the read-across method, is based on the observed similarities in molecular properties between structurally similar compounds. The read-across method assumes for the initial hypothesis that a (quantitative) structure-activity relationship exists⁴³.

Most of the QSA(P)Rs are regression-based models, which are often given in the form of a linear correlation (Eq.1) between the molecular descriptors (X) and molecular property (Y), with some coefficients (*a, b, c*)⁴⁴:

$$Y = b + aX_1 + cX_2 \dots \quad (1)$$

The GC methods use the molecular groups that constitute the molecule to make a prediction. The methods follow the so-called “additive principle”, in which the molecular property value (X) is obtained by summing the contributions of all the molecular groups⁴⁰ (Figure 2).

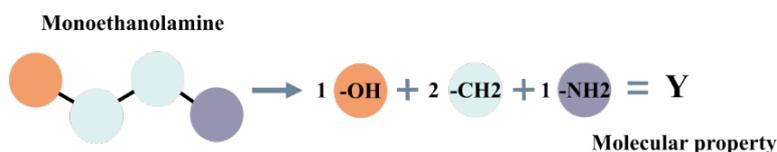


Figure 2. Schematic representation of the ‘additivity-principle’.

It was first noticed more than 100 years ago that the molecular structure determines the molecular properties of a chemical. This idea of a correlation between the structure and property of a molecule is widely used to predict molecular property data, such as boiling and flash points or enthalpy of vaporization, to name but a few. Different QSAR and GC models have been developed over the years. The QSARs and GC methods are relatively simple to apply and can easily be automated, enabling their use in approaches such as CAMD.

2.2.2 LCA data

Over the last few years, efforts have been made to create methods that can estimate the missing cradle-to-gate life-cycle impacts through applying data science methods. The first automated LCA impact methods were those of Wernet et al. (2008, 2009)^{45,46} who used neural networks, principal component analysis and molecular descriptors to develop a tool for estimating for single molecules the cumulative energy demand (CED), global warming potential (GWP), and an endpoint of Eco-indicator 99 (EI99). More recent studies have proposed an improved version based on this approach. Song et al. (2017)⁴⁷ have created deep artificial neural network (ANN) models to estimate the life-cycle impacts of chemicals. The multilayer ANNs use molecular structures to predict the life-cycle impacts of chemicals in six impact categories: cumulative energy demand; global warming (IPCC 2007); acidification (TRACI); human health (Impact 2000+); ecosystem quality (Impact 2000+); and EI99 (I,I, total)⁴⁷. The method of Calvo-Serrano et al. (2018)⁴⁸ applies a mixed-integer programming (MIP) optimization framework to construct predictive models to estimate CED, GWP, and EI99. This method expands upon that proposed by Wernet et al. (2008)⁴⁵ by including thermodynamic properties and automatic selection of the optimal chemical attributes, thereby generating better predictions. The final multi-linear models are easily combined with other software packages for molecular and process design.

2.3 Phase-change solvents for CO₂ capture

Liang Hu first patented the concept of phase-change solvents in the context of gas-liquid absorption in 2005⁴⁹. More recently the concept has found application in CO₂ capture and these solvents has attracted increased interest as third-generation absorbents (e.g., biphasic, enzymatic-enhanced, and encapsulated absorbents)⁵⁰.

Phase-change or biphasic solvents are miscible solute-solvent mixtures that are undergoing a phase liquid-liquid or liquid-solid separation that is promoted by CO₂ absorption and alterations in temperature⁵¹. CO₂-triggered solvents generate two phases as a result of changes in (i) polarity, called switchable-polarity solvents; or (ii) ionic strength, named switchable water solvents; or (iii) hydrophilicity, called switchable-hydrophilicity solvents when the solute reacts with the CO₂. Switchable-polarity solvents are usually nonaqueous mixtures. The reversible switching from low to high polarity that follows the CO₂ absorption results in a change in the solubility of the solutes in the solvents and, thereby, separation of the phases. The phase separation in switchable water solvents occurs when an ionogenic compound of an aqueous mixture becomes ionic during the dissolution of CO₂ in water. As a result, a homogeneous mixture of low ionic strength is transformed into a bicarbonate salt solution with higher ionic strength, leading to a liquid-liquid phase separation⁵¹. Switchable-hydrophilicity solvents are hydrophobic liquids that exhibit poor miscibility with water until they are exposed to CO₂. Absorption of CO₂ makes the solvent hydrophilic and fully miscible with water⁵². The desorption of CO₂ leads to separation of the phases. Among the CO₂-triggered solvent mixtures that are

suitable for CO₂ capture are the nonaqueous (i.e., amine-alcohol, amino acid-alcohol, ionic liquid-amino acid blends) and aqueous (i.e., water-amines mixtures or solutions of the salts of amino acids⁵¹) types.

Solvent phase separation triggered by changes in temperature (with so-called ‘thermomorphic solvents’) is typically obtained using mixtures of lipophilic amines and water. Phase separation occurs in such mixtures when the increase in the process temperature breaks the intermolecular hydrogen bonds between the amine and water⁵¹.

Nonaqueous solvents have lower regeneration temperatures than aqueous solvents due to the replacement of water with organic solubilizers or ionic liquids. Organic solubilizers or ionic liquids are advantageous in terms of solvent degradation and evaporation, corrosiveness, and the heat requirements for solvent regeneration. However, practical application of nonaqueous solvents for carbon capture might be restricted by the presence of water in the flue gas that is sent for treatment. The presence of water could affect the efficiency of the process, necessitating additional water management⁵¹. However, CO₂ capture with nonaqueous solvents has not yet been fully assessed and requires further research. Capture processes that use aqueous solvents have been studied extensively, with some aqueous phase-change solvents reaching technology readiness levels of 4 and 5⁵³. Table 1 gives examples of phase-change solvents and the drivers for their phase separation.

Solvent blends (e.g., salts of amino acids) that result in a solid CO₂-rich phase might create process-related challenges due to precipitation. In contrast, mixtures that exhibit liquid-liquid separation are easily incorporated into the existing absorption/desorption systems through the addition of a mechanical separation step⁵³ (Figure 3). For this reason, most of the phase-change solvents introduced as CO₂ capture solvents, which are also those investigated in this work, are solvents that exhibit liquid-liquid phase separation.

Table 1. Examples of phase-change solvents with various phase separation drivers

Solvent	Phases	Phase separation driver
CO ₂ -triggered		
Dipropylamine (DPA)	Liquid-solid	Change in the polarity ⁵⁴
MEA/iso-octanol	Liquid-liquid	Change in the polarity ⁵¹
Tetrabutylphosphonium N-trifluoromethanesulfonylleucine (ionic liquid)/water	Liquid-liquid	Change in the hydrophilicity ⁵²
N-methyldiethanolamine/water	Liquid-liquid	Change in the ionic strength (switchable water) ⁵²
Potassium taurate (amino acid salt) / water	Liquid-solid	Change in the polarity ⁵¹
MAPA/DEEA/water	Liquid-liquid	Change in the ionic strength (switchable water) ⁵²

Thermomorphic		
DPA/DMCA /water	Liquid-liquid	Change in miscibility triggered by a temperature change ⁵¹
DMX solvent	Liquid-liquid	
MCA/DMCA/AMP	Liquid-liquid	

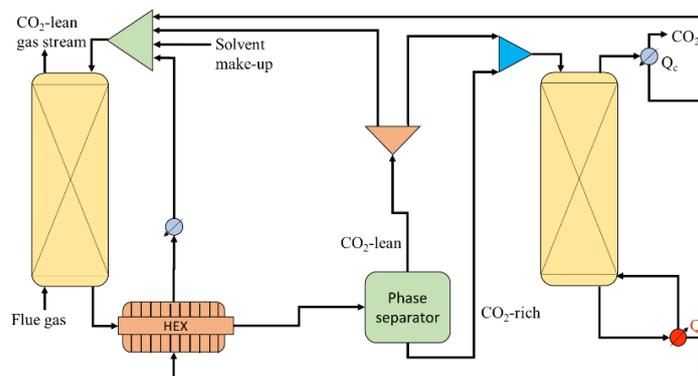


Figure 3. CO₂ capture using phase-change solvents⁵⁵

2.4 Contribution of this work

The section summarizes the contributions of this thesis to the field. The major contributions involve integrating the impact assessment into CAMD, the evaluation of novel phase-change solvents and carbon capture systems, and the use of knowledge for predictive modeling.

2.4.1 CAMD with simultaneous impact assessment

Several CAMD studies have incorporated EHS indicators^{56–59}, and a few studied considered the LCA impact constraints. For instance, Mehrkesh and Karunanithi (2014)⁶⁰ included downstream LCA impacts associated with the emissions of ionic liquid solvents. Heintz et al. (2014)⁶¹ and Weis and Visco (2010)⁶² used a single LCA score, which was estimated based on quantitative-structure property relationships (QSPRs) derived from the available data for 46 frequently used solvents. The main challenge associated with incorporating more-recent streamlined LCA methods lies in the lack of models that can be easily integrated into the CAMD framework without significantly affecting computational performance. No systematic and broad screening of new phase-change solvent alternatives has been performed, particularly not with respect to the sustainability-related properties.

The work described in **Paper II** addresses both aspects of incorporation of the EHS and LCA indices into CAMD, as well as the use of the combined approach to identify

potential phase-change solvents. An analysis of the effects of such early integration of the impact assessment into CAMD is also performed.

2.4.2 Impact assessment of phase-change CO₂ capture systems

The performances of the phase-change solvents in terms of reduced energy consumption have been widely investigated. However, little attention was given to other environmental, health, and safety aspects of these emerging materials⁵¹. The negative health and environmental impacts of amine-based capture systems have been receiving much attention in recent years. The negative effects are associated with the formation of hazardous substances, such as nitrosamines and nitramines during operation⁶³, as well as other emissions from the capture systems⁶⁴. The operating conditions of the process are reported to have a significant impact on the magnitude of the potential adverse effects²³.

The work introduced in **Paper III** aims to gain insights into the impact of the carbon capture systems caused by the use of phase-change solvents. The work identifies the potential hotspots of such systems in terms of impact and aspects of increased concern. The results can facilitate decision-making and trigger targeted experimentation at pilot and demonstration scales, towards defining the optimal conditions for the operation of the CO₂ capture systems.

2.4.3 Predictive modeling

The predictive capacity of the GC method is sufficient to estimate the thermodynamic properties, such as boiling or flash points, of the chemicals. However, there remains room for improvement of the GC and other quantitative-structure property relationship (QSPR) models for predicting the molecular parameters used in LCA and EHS assessments. The advanced models that apply data science methods (mostly machine learning, ML) often show superior performance^{65,66}. However, the results from such ML models are often difficult to interpret⁶⁵, as they are black box models requiring a large amount of training data. **Paper I** evaluates the performance of a simple ML approach for automated estimation of molecular properties, such as acute aquatic toxicity, bioconcentration, and persistency. The evaluation aims to support the integration of the impact assessment into the CAMD procedure.

Even though QSAR, GC, or ML methods are used to reduce the need for experimental testing, they are still based on chemical property data, which are often scarce and can show significant variability even for the same compound⁴⁴. For instance, toxicity data are traditionally obtained through extensive experimental testing on living organisms, which is undesirable from both the economic and ethical perspectives. In fact, EU legislation encourages the development and use of alternative *in vitro* or *in silico* approaches to replace cruel animal testing⁶⁷.

One of the ways to address these limitations is to generate hybrid models that incorporate other knowledge that already exists in the field (i.e., prior knowledge). Prior knowledge refers to data labeling, generic conclusions, functional trends in the

relationships between target and predictor variables, simplified input/output models for specific classes of chemicals, etc. The approach has been successfully applied in various fields, such as medicine⁶⁸, drug safety⁶⁹, materials⁷⁰, and image recognition⁷¹, although it has not yet been applied to the molecular properties of interest for LCA and EHS assessments. One of the main reasons for this is that prior knowledge is not systematically extracted, classified, and formulated in ways that allow it to be effectively used for hybrid modeling.

Paper IV proposes an approach for semi-automated knowledge extraction that enables the systematic extraction and classification of knowledge from a high number of scientific articles published in the field of acute aquatic toxicity. The method can be valuable in interdisciplinary research when rapid acquisition of knowledge is needed for different types of work (e.g., impact assessments).

Paper V presents extensive results regarding knowledge extraction and utilization in the field of acute aquatic toxicity. The main trends, patterns, and other important aspects of aquatic toxicity are collated and classified to facilitate further use of the knowledge. The extracted knowledge is then applied in predictive models to estimate the toxicities of chemicals that are structurally similar to the phase-change solvents. The performances of the predictive models with and without prior knowledge are evaluated.

3 Methods

This work evaluated the impact of phase-change solvents and similar to them saturated aliphatic compounds. The potential impact of capture systems applying the phase-change solvents was also assessed. The impacts were assessed by using a combination of LCA and EHS hazard assessments. The evaluations were performed at the substance level and process levels. The substance level assessment analyzed the integration of sustainability indicators into the CAMD procedure. The process level looked at the impacts of phase-change solvent carbon capture systems. The aspects of molecular property estimation were addressed using predictive models and field (prior) knowledge.

3.1 Impact assessment at the substance level

The substance level assessment aimed to identify molecules with lower EHS impacts during the CAMD, for applications as phase-change solvents for CO₂ capture. To assess the impacts of the molecules generated in CAMD, LCA and individual environment (E), health (H), and safety (S) hazard assessment indices were computed for all the molecular alternatives in each iteration of the CAMD optimization procedure. Thereafter, together with the other CAMD constraints and molecular properties, they were used to calculate the objective function. This “on-line calculation” relies on the availability of empirical and analytical predictive models for all the molecular properties included in the computation of the indices, which means that no data gaps are allowed. The integration of the LCA and E, H, and S indices calculations with the CAMD is shown in Figure 4.

The molecular structures were presented as a set of functional groups, together with the number of their appearance in the molecules. The following functional groups were considered for the CAMD approach CH₃, CH₂, CH, C, OH, CH₂NH₂, CH₂NH, CH₂N, CHNH₂, CHNH, CH₃NH, and CH₃N. As an example, one molecular structure could be decoded as follows: [CH₃, CH₂, CH₂NH₂] [1,4,1]. The number of functional groups present in each molecule ranged from 6 to 16. Up to one amine group was allowed in each functional group. This allowed for the generation of simpler molecules with fewer isomers, making it easier to interpret the results. Details of the integration procedure can be found in **Paper II**.

3.1.1 LCA indices

Within the CAMD, the LCA indices were calculated automatically using streamlined FineChem models⁴⁶, which estimate the environmental impact of producing the solvent based on its molecular structure. Since the phase-change molecules are not yet synthesized or utilized on a large scale, they are not present in the LCA inventory databases. The FineChem models estimated the LCA impact associated with the

production of 1 kg of a substance. The “cradle-to-gate” boundaries were used to consider the synthesis of the solvents. The same boundaries were applied for the processes that produced the material and energy consumed during the process. The FineChem models rely on industrial production data complemented with background data inputs from the ecoinvent database⁴⁶. Selected were the Global warming potential (GWP 100a)⁷², cumulative energy demand (CED), and an endpoint of the Eco-indicator 99 (EI99) impact assessment method⁷³; these LCA metrics are predicted more accurately by the FineChem models.

The FineChem models use the information on a number of molecular features obtained from a SMILE (Simplified Molecular Input Line Entry Specification) form of molecular alternatives generated in CAMD. These molecular features are⁴⁶:

- molecular weight
- number of functional groups (total)
- number of oxygen atoms in the keto and aldehyde groups
- number of oxygen atoms not in the keto and aldehyde groups
- number of nitrogen atoms
- number of halogen atoms
- number of aromatic or aliphatic rings
- number of ternary or quaternary carbon atoms
- number of heteroatoms in rings
- number of unique substitutes on aromatic rings

The SMILEs were formed for all the possible isomers of a molecule given as a set of functional groups and the number of times they appeared, as obtained from the CAMD (Figure 4). Thus, isomers of the same molecule can have different LCA values. For example, some isomers can have a ternary carbon, resulting in higher impact for the production of such a molecular structure. The final LCA indices returned to the main algorithm were the average values among all the isomers generated for a set of functional groups.

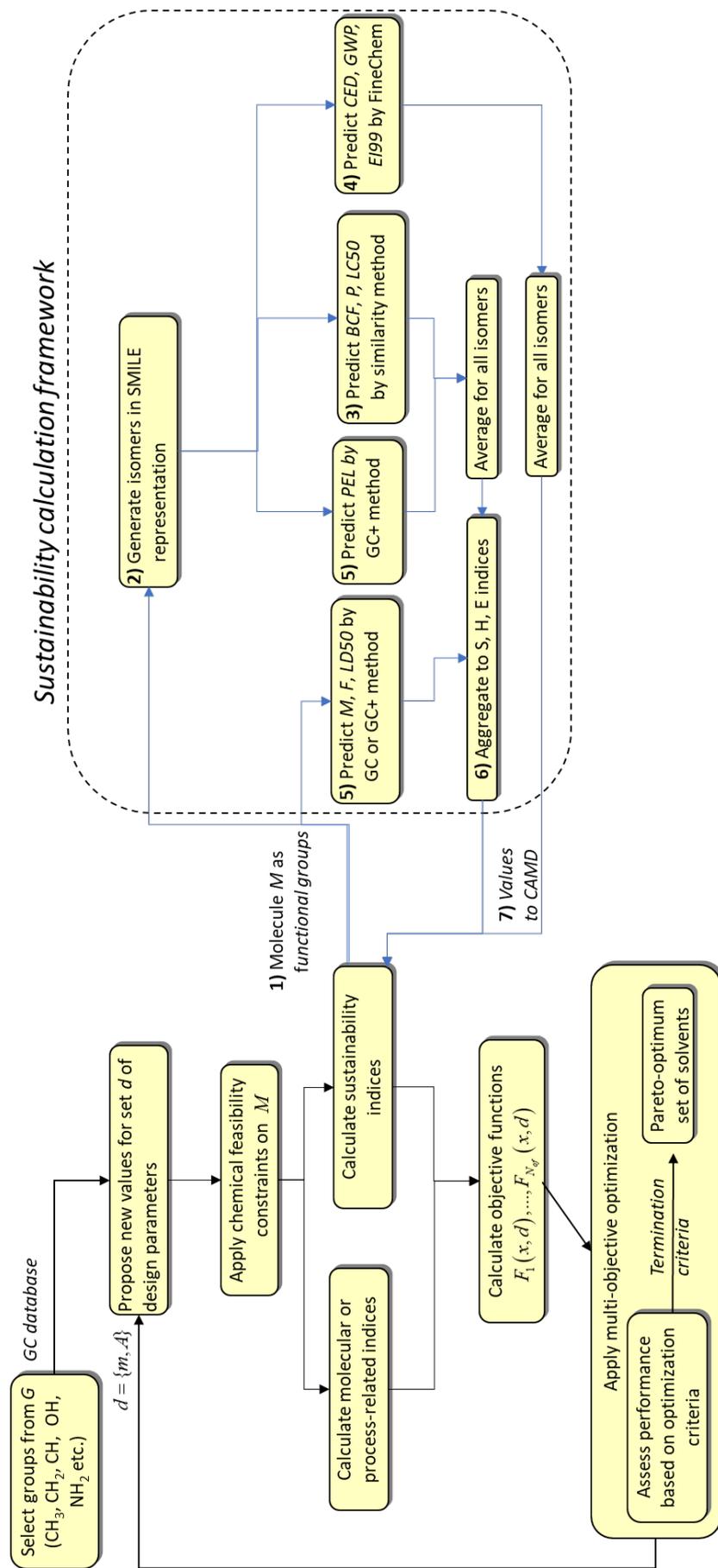


Figure 4. Simultaneous CAMD and sustainability assessment calculation steps. Source: Paper II

3.1.2 EHS indices

The E, H, and S indices were computed based on the effect categories and parameters/ molecular properties summarized in Table 2. The selection of certain parameters and molecular properties was dictated by the availability of online (automated) estimation methods easy to integrate into CAMD. The online estimation methods used for prediction are also given in Table 2. The environmental (E) aspect evaluates the potential damage if a solvent molecule is released to the environment as a consequence of everyday operation or accidents. The health (H) aspect assesses the influence of daily contact with chemicals on workers' health. The safety (S) aspect relates to the hazards of using a solvent in the process, its potential to lead to accidents, and its short-term effects on plant workers.

Various effect category indices were computed for each molecular alternative to estimate their E, H, and S scores. The effect categories are based on the physical or chemical properties of the compounds (Table 2). The S score is based on three effect categories: mobility, flammability/explosiveness, and acute toxicity. The chronic toxicity effect was used to compute the H score. The E score considers the magnitude of water-mediated effects, degradation, air-mediated effects, and accumulation⁷⁴. The E, H, and S scores were calculated via an index (I), taking a value between 0 and 1, depending on the value for the particular effect category. Figure 5 presents an example of the index-effect category scale for acute aquatic toxicity. The rankings of the parameters were derived from the works of Koller (2000)⁷⁴ and Badr (2016)²³. The final scores were computed using Eq. (2):

$$Score_i = \sum_{j=1}^n I_{i,j} \quad (2)$$

where $Score_i$ is the E, H, or S index of a substance, and $I_{i,j} = I_{i,j} + F_{i,j}$ is an effect category index in the corresponding hazard category modified by a fate index (F) according to Koller (2000)⁷⁴. The fate index is a function of molecule volatility and persistence in the environment; the higher the volatility and persistence of the substance, the higher the risk that the molecule will cause an undesired effect²³. The term n represents the number of effect categories considered for E, H, or S.

Table 2. Dangerous properties of chemicals for use in the EHS hazard assessment and their estimation methods ⁷⁵.

Effect category	Parameter/Molecular property	On-line estimation method
Safety (S)		
Mobility	Difference between the boiling point of the pure substance and the highest process temperature	Group contribution ⁷⁶
Flammability/ Explosiveness	Difference between the flashpoint of the pure substance and the highest process temperature	Group contribution ⁷⁶
Acute toxicity	<i>LD50</i> (rat) – the amount of orally ingested chemical that causes 50% of the rats to die (mg/kg body weight)	Group contribution+ ⁷⁷
Health (H)		
Chronic toxicity	Permissible exposure level (<i>PEL</i>) – a US legal exposure limit (mg/m ³) for an employee to a chemical, given as the average exposure concentration over 8 hours	Group contribution + ⁷⁷
Environment (E)		
Water-Mediated Effects	Toxicity (<i>LC50_{aq}</i>) – concentration (mg/l) of a chemical in the water that causes death/effect to 50% of most-sensitive aquatic species	Similarity ^{78,79}
Degradation in the environment	Persistence (<i>P</i>) (in days) represents the half-life of a chemical in the environment	Similarity ^{78,79}
Air-Mediated Effects	Chronic toxicity index calculated based on permissible exposure limit (<i>PEL</i>)	Group contribution + ⁷⁷
Accumulation	The bioconcentration factor (<i>BCF</i>) reflects the accumulation of a chemical in a living organism. It is calculated as the concentration of an chemical in an aquatic organism divided by the concentration of that chemical in the water.	Similarity ^{78,79}



Figure 5. Water-mediated effects index calculation based on LC50/EC50 values.

3.2 Process level assessment

The process level assessment focused on the impact of the phase-change carbon capture systems. The evaluation entailed LCA and EHS hazard assessments, supplemented with an exposure analysis. The assessment included the process-specific features of the phase-change systems and an analysis of the problems that are typical of conventional carbon capture systems, which might appear in the new systems applying

phase-change solvent alternatives. The issues could be linked to the molecular structures of the phase-change solvents, the feed flue gas quality, or the process design of the phase-change capture system. Such an analysis was required to introduce the necessary mitigation equipment and to improve the process design. The LCA and EHS hazard assessments were then used to evaluate whether the improved system could still harm the plant workers and environment. The aspects of the system that might require particular attention⁵⁵ were identified.

Figure 6 shows the process structure after introduction of the necessary mitigation equipment and the scope of the LCA and EHS hazard assessments. Under process conditions, solvent molecules tend to react with species other than CO₂ (NO_x, SO_x, O₂, soot), which enter the system with the flue gas. For example, reaction of an amine solvent with NO_x substances can lead to the formation of carcinogenic nitrosamines. Even low concentrations of nitrosamines in the air, in the range of 2.39-7.55 mg/m³²³, are harmful to human health. SO_x and soot particles affect the levels of emissions from the capture system. Thus, if a flue gas that is entering the capture unit contains high levels of NO_x and SO_x, a pre-treatment step (e.g., NO_x/SO_x scrubber) may be needed. The presence of oxygen, the high temperature of the solvent regeneration process, and the molecular structure all influence the rate of degradation of the solvent molecules. To prevent the accumulation of degradation products within the system, it is necessary to reclaim the solvent. Washing units are needed to remove the volatile degradation products and emissions of the solvent carried by the treated flue gas as it leaves the absorber. The washing unit and reclaimer waste flows are sent to the wastewater treatment unit, which is estimated by wastewater treatment models⁸⁰. The solvent lost during the process needs to be continuously replaced through increased solvent production.

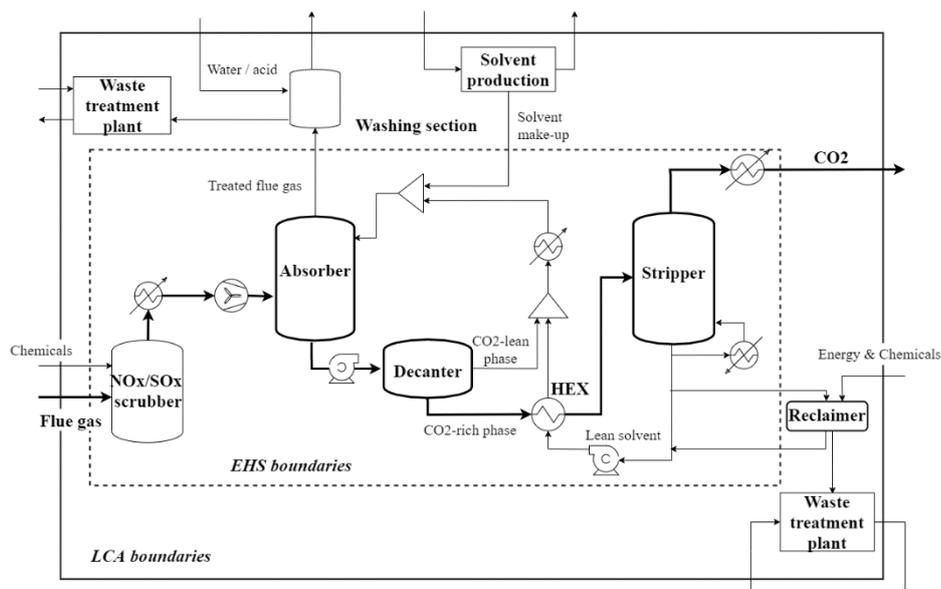


Figure 6. Process structure and scope of the LCA and EHS hazard assessments of the carbon capture process with a solvent exhibiting liquid-liquid phase separation. Note that the position of the heat exchanger (HEX) might be before or after the decanter depending on the type of phase-change solvent used. Source: Paper III.

3.2.1 LCA assessment

The process level LCA was designed to evaluate the impact of capturing 1 tonne of CO₂ (90% capture rate, 98% purity by mole) with a capture system that uses phase-change solvents under steady-state operation of the process. It was also intended to compare the results with those of similar studies using conventional solvents (i.e., MEA). The system boundary was cradle-to-gate, including only the impacts from chemical auxiliaries and energy utilities that are solvent- or process-dependent. The amounts of energy required to transport the CO₂-containing flue gas to the capture site and for the compression and transportation of the captured CO₂ for further use or storage were excluded. As delivering the flue gas to the capture site is not expected to depend upon the type of solvent or process, the delivery was assumed to be the same for all the solvent-based capture systems. The mode of transportation or usage, which mainly determines the CO₂ compression for storage or utilization, was assumed to be independent of the process or solvent type, since the solvent itself did not require specific process conditions (for example, high pressure in the absorber or stripper). Background data in the form of inventories for the production of chemical auxiliaries and energy utilities were obtained from the ecoinvent ver. 3.4 database⁸¹. The foreground data (i.e., process inventory data) were derived from process flowsheets created using experimentally derived equilibrium data.

The following impact categories were considered for the life-cycle impact assessment: Global warming potential (GWP 100a), Cumulative energy demand (CED), and a single score from the ReCiPe-2008 method based on a hierarchist weighting scheme (ReCiPe)⁸². The selection of the impact categories used in the assessment was based on the limited number of impacts estimated by the FineChem models (CED, GWP, EI99) for the solvent production process. The ReCiPe score for the solvent production was calculated based on correlation with the EI99⁸³, computed by the FineChem model. The correlation was developed after the CAMD results were obtained. Thus, only EI99 was used at the substance level in the assessment.

More information on the procedure and the data considered in the LCA calculations can be found in the *Supplementary material* to **Paper III**.

3.2.2 EHS hazard assessment

The EHS process level hazard assessment considered the inherent properties of the solvent molecule and other chemicals used in the capture system and their corresponding flows and hold-ups in the process equipment. The effect categories used for the process level assessment were computed based on the material safety data sheets (MSDS) and other sources that documented the experimentally obtained properties of the chemicals. If the experimental data were missing, values for the properties were estimated by the online methods applied during the substance level assessment. During the process level assessment, the aggregated E, H, and S scores were computed, and thereafter combined

with other sustainability indices for decision-making. The total E, H, and S scores were calculated according to Eqs. (3)–(5).

$$S = \sum_{iS} \max_F [\sum_j m_j^F I_j^{iS}] \quad (3)$$

where m_j^F is the mass flow of a specific substance j in-stream F of the CO₂ capture process system per kg of CO₂ captured under the same conditions as in the LCA (e.g., 90% recovery and 98% mole purity), and I_j^{iS} is an index of substance j for an effect category i representing the safety aspect.

$$H = \sum_H \max_j (m_j^1 I_j^H) \quad (4)$$

where m_j^1 represents 1 kg of chemical substance j . The amount of chemical here is fixed due to the long-term aspects considered as part of the health hazard, which mainly depend on the existence of hazardous substances for the design of mitigation measures rather than on the amounts of such substances³¹. I_j^H is an index of substance j for an effect category listed for the health score.

$$E = \sum_{iE} \sum_j (z \max_F (m_j^F) I_j^{iE}) + \sum_{iE} \sum_j (m_j^{out} I_j^{iE}) \quad (5)$$

where z is a fraction of the mass emitted to the environment in case of an accident, and m_j^{out} is the specific amount of chemical substance j per kg of CO₂ captured under the same conditions as in the LCA (e.g., 90% recovery and 98% mole purity), leaving the system to waste treatments units, for example, except for the product flow. I_j^{iE} is an index of substance j for an effect category - i used for the environment score.

In addition to the individual E, H, and S scores calculated for the capture systems, the potential of danger (PoD)⁷⁴ was also computed, as this evaluates the magnitude of the EHS effects in terms of a physical unit, for instance, the releasable energy content of the system (in MJ) or the volume of air/water (in m³) required to dilute a dangerous chemical⁷⁴. The overall impact is a combination of the PoDs values for all the chemicals used or formed during the capture process. The PoD was calculated using Eq.(6)⁷⁴:

$$PoD_{ij} = m_i \times 10^{a_i \times I_j - b_i} \quad (6)$$

where PoD_{ij} is the PoD for chemical i for the j^{th} effect category, I_j is an effect category index, m_i is the maximum mass of substance in the process per kg CO₂ captured under the same conditions as in the LCA (e.g., 90% recovery and 98% mole purity), and a_i and b_i are exponential parameters⁸⁴.

The PoDs were calculated so as to perform multi-criterion assessments and to evaluate the most-potent impacts of single substances in specific categories. The PoD

values allow comparisons of different chemicals in the system to identify which chemical can be attributed the strongest impact.

3.2.4 Exposure analysis

The exposure analysis evaluated the process-specific hazards (such as leakages or spills) that lead to workers being exposed to harmful substances present in the capture system. Occupational exposure concentrations (OELs), which represent the maximum airborne concentrations of chemicals to which workers can be exposed over a period of time without experiencing adverse health effects⁸⁵, have been established for many chemicals. The concentrations of chemicals that induce lethal health effects following dermal or oral exposure are given as acute toxicity limits, expressed as the LD50 (oral, dermal) or LC50 (inhalation) values⁸⁶. Information on the OELs and acute toxicity chemicals is provided by manufacturers in the MSDS or can be found in databases^{87,88} or chemical directories and guidelines. If no experimental data exist, estimation methods, such as GC for PEL (permissible exposure level) can be used.

The focal points of the analysis were exposure to solvent via leakages (long-term) and spills (short-term) and accumulation in the system of nitrosamines at concentrations exceeding the safety range of 13.7–14 mM^{23,89}, leading to harmful concentrations of the nitrosamines in the air. The concentration of solvent in the air in the case of leakage or spillage was estimated by the equations provided by Nicas (2016)⁹⁰ and Keil and Nicas (2003)⁹¹. A case of leakage assumed a constant amount of the solvent being released to the working environment at a constant evaporation rate. Very low, difficult to detect, levels of leaked solvent were assumed, resulting in prolonged periods of exposure. Thus, the resulting air concentration of the solvent was compared to the regulatory TWA (the time-weighted average limit for an 8-hour workday with repeated exposure but with no adverse effects). Different spillage volumes were assessed at decreasing-with-time emission rates in the case of a spill. Dangerous-to-health air concentrations for the solvent were assumed to be present during the first seconds of the spill. The air concentrations were compared to IDHLs (immediately dangerous to life or health) or to the inhalation toxicity (LC50) limits for humans, indicating possible adverse effects induced by the exposure. In both cases, the air concentration of the solvent was also calculated after 15 minutes of a leakage or spillage incident. This was compared to the STEL (short-term exposure) or TWA to judge whether safe concentrations were reached within that period of 15 minutes.

Unlike the solvent, the concentration profile of nitrosamines in the system is unknown. Thus, a different approach was required to assess the level of exposure to nitrosamines. To avoid a dangerous concentration of nitrosamines in the air, the potential steady-state concentration of nitrosamines⁸⁹ in the system was calculated and compared to the safe level.

More details of the methodology are given in **Paper III** and *Supplementary material* to **Paper III**.

3.3 Predictive modeling

The predictive modeling addresses aspects of the data gaps relevant to the newly designed and existing structures. The models that are developed to estimate the missing properties are based on a simple kNN (k-nearest neighbors) algorithm. Several ways to use prior knowledge for development of the models are introduced.

3.3.1 kNN

The kNN supervised ML algorithm allows estimation of a missing property value using the molecules that are structurally most-similar (nearest neighbors) with known property values (Figure 7). Together with the GC methods, the algorithm was applied to estimate the missing values for the acute aquatic toxicity, bioconcentration, and persistency of the molecular structures generated during the CAMD procedure (**Papers I and II**). The kNN algorithm was also employed to test different options for prior knowledge use towards enhancing and evaluating the acute aquatic toxicity estimation models (**Paper V**).

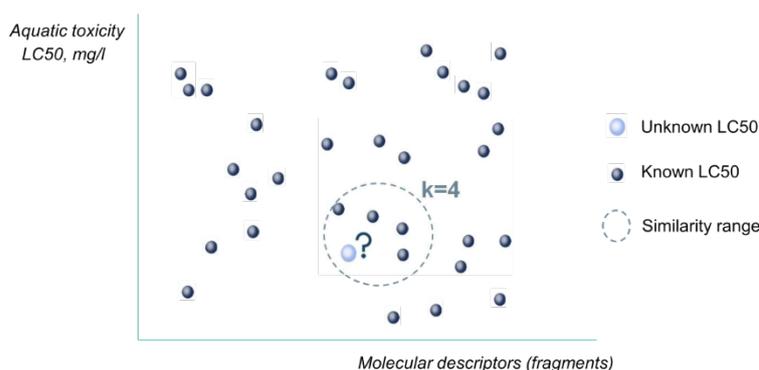


Figure 7. Schematic representation of k-nearest neighbors (kNN) algorithm approach.

The nearest neighbors were identified in two ways: (i) the Manhattan distance between the molecular descriptor vectors representing the molecules; and (ii) the Tanimoto similarity between molecular fingerprints (i.e., molecular features encoded to bit strings⁹²). The Manhattan distance between two points x and y in n -dimensional space is the sum of the absolute differences between their coordinates. Manhattan distance is defined by Eq.(7)⁹³:

$$d_M = \sum_{i=1}^n |x_i - y_i| \quad (7)$$

where x_i, y_i are the i -th descriptors of molecular vectors x and y .

The Tanimoto coefficient measures the similarity between two sets of molecular features defined by the common and total number of the sets' features. The Tanimoto coefficient is computed according Eq. (8):

$$S_{i,l} = \frac{c_{i,l}}{a_i + b_l - c_{i,l}} \quad (8)$$

where a_i and b_l are features of molecules i and l (on bits), while $c_{i,l}$ is the number of molecular features common to the two molecules. The following molecular features were considered: hydrogen bond donor, hydrogen bond acceptor, basicity, acidity, aromatic ring, and halogens (F, Cl, Br, I)⁹⁴. The features were encoded in the form of a Morgan molecular fingerprint (nBits=1,024).

The descriptors, fingerprints, and Tanimoto similarity were calculated with the help of the open-source cheminformatics tool RDKit⁹⁴. The estimation of the toxicity value during the CAMD procedure and for the fingerprint-based models without the use of prior knowledge was performed using Eq. (9)^{79,95}. The average of the neighbors' toxicity values was taken for the descriptor-based approach and all the hybrid models using prior knowledge.

$$y_{pre,i} = \sum_{l=1}^k \frac{S_{i,l}}{\sum_{j=1}^k S_{i,j}} * y_{l,db} \quad (9)$$

where $y_{pre,i}$ is designed to predict the property of molecule i . The $S_{i,l}$ term is a similarity value between molecule i for which property predictions are sought and a molecule l found in a database, for which the desired property value $y_{l,db}$ is available, and k is the optimal number of similar neighbors used in the prediction, as discovered from a cross-validation procedure (also for the descriptor-based approach).

3.3.2 Dataset

The aquatic toxicity data used in this study were retrieved from the PBT (persistence, bioaccumulation potential, toxicity) database collated by Stempel (2012)⁹⁶. The original database created by Stempel (2012) contains 94,483 chemicals. Chemicals identified as inorganics, epoxides, and peroxides and those with molecular weight >1,000 were excluded to avoid the errors encountered with prediction tools such as ECOSAR. The ECOSAR databases and the prediction were used to obtain the acute aquatic toxicity values for most of the chemicals in the original database. For approximately 2,000 chemicals, the toxicity data were obtained from the Aquire ECOTOX, Canadian Domestic Substance list, and EnviChem databases⁹⁶. "The most-sensitive species" approach was followed, i.e., the lowest effect concentration with LC50 and EC50 and duration of either 96 hours for fathead minnow or 48 hours for *Daphnia magna* was selected. For those chemicals for which no data were available in ECOSAR (7,783 molecules), the baseline toxicity was calculated based on the octanol-water partition coefficient.

Due to the variability of the data sources, data quality, and the absence of an indication of the origin of every value, the data are associated with some uncertainty and

inaccuracy. For instance, the reported accuracy level (the predicted LC50 falls within the same regulatory category, i.e., high, moderate, low, no hazard, as the measured LC50) of the ECOSAR predictions is only around 60%⁹⁷.

The reduced dataset of saturated aliphatic compounds that contain only C, H, N, and O atoms was used in the present study. These restrictions in terms of the molecular composition and structure were defined by the scope of the study being limited to only saturated aliphatic phase-change solvents. The limited number of molecular structures decreased the computational time when testing the combined CAMD with the sustainability indicators approach and different hybridization options.

3.3.3 Prior knowledge for predictive modeling

A lot of information exists for almost every domain. However, due to the amount of manual work required to process this information⁹⁸, identifying and using the relevant knowledge can be cumbersome. One of the sources of the domain's prior knowledge is scientific publications. While utilization of the knowledge that exists in scientific articles is challenging⁹⁹, the need for making such knowledge more accessible to researchers and non-professional users is growing. This part of the work explores a semi-automated system for knowledge extraction from scientific publications in the acute aquatic toxicity domain. A few ways for the use of such knowledge in predictive modeling are also evaluated.

The method used for the extraction of knowledge from scientific articles is presented in Figure 8. The knowledge extraction process starts with article collection, followed by text mining, analysis of the obtained results, and additional article screening. The semi-automation of the method aims to accelerate the processing of large volumes of primary information. The results of the automated part also provide information that guides further processing of the extracted knowledge using human judgment.

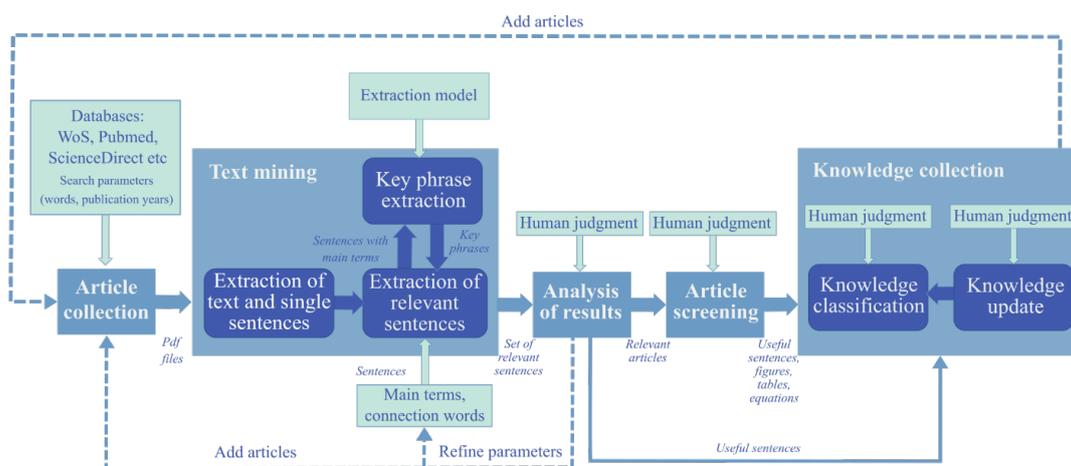


Figure 8. Knowledge extraction process for a specific domain of interest.

Article collection. Knowledge extraction was performed on scientific articles collected from the ScienceDirect, PubMed, and Web of Science databases. “Aquatic toxicity” and a period covering 21 years (2000–2020), were used as the search parameters. Papers with titles related to predictive ecotoxicity, QSARs, information on the aquatic toxicity of the separate chemical classes (groups), and modes of action (MoA) were collected. The article collection step resulted in the identification of around 400 publications. Analysis of the bibliometric information for the collected articles is presented in Figures S1-1 and S1-2 (*Supporting Information 1* for **Paper IV**).

Text mining. Automated text mining consists of three parts: extraction of article texts and single sentences, key phrases, and relevant sentences. First, the text of the articles was recognized and pre-cleaned, such that the title, abstract, and references were removed, as well as extra spaces that appeared during the text recognition and e-mails. After the text-cleaning process, single sentences were identified based on sentence terminators (dot, exclamation and question marks, quotation marks, brackets). These complete sentences led to the extraction of relevant sentences. The relevant sentences were identified based on the reader-provided input, namely the presence of preselected “main terms” and “connection words”. The main terms included the words “toxicity”, “acute”, “LC50”, and “EC50”. The following words served as the connection words (as complete words or lemmas): “increase”, “decreas”, “relat”, “correlate”, “structure”, “fragment”, “class”, “significant”, “high”, “affect”, “low”, “link”, “reason”, “determin”, “predict”, “influence”, “severe”, and “depend”. The text mining generated a list of relevant sentences for every article.

Analysis of the results and article screening. The extracted set of the relevant sentences was then evaluated manually to identify useful sentences. In this study, a sentence was considered useful if it contained information that could be used in hybrid predictive modeling (i.e., the sentence referred to aspects influencing acute aquatic toxicity values). The useful sentences were collected either as directly extracted knowledge or to identify articles and parts of the text for additional manual screening. The information retrieved in this step in the form of useful sentences, models, figures, and tables was used for structuring knowledge via the development of a classification scheme.

Knowledge collection consists of knowledge classification and updating. The information extracted from the articles published in the period 2000-2014 (225 articles) was used to develop the initial classification scheme. The information extracted from the remainder of the articles (165 articles) was used to update the classification scheme and demonstrate a procedure that combined the classification scheme with knowledge acquired in the future.

An overview of the knowledge extraction method can be found in **Paper IV**, while the methodological details and extensive results are presented in **Paper V**.

The collected knowledge was then used for predictive modeling. In this work, two types of models were constructed: with and without prior knowledge. Models that were constructed without using prior knowledge were designated as ‘standard’ models, and the

models that did utilize prior knowledge were termed ‘hybrid’ models. The standard models were developed for the purposes of comparison, to evaluate the impact of the knowledge incorporation on the performances of the models. The hybrid models used the part of the extracted knowledge that was relevant to the dataset (i.e., concerning the toxicities of aliphatic compounds for fish and crustaceans).

The prior knowledge was applied before, during, and after the kNN algorithm approach, according to different schemes presented in Figure 9. For model H0 (Figure 9a), a portion of the data was removed before the estimation was performed with the kNN algorithm. According to prior knowledge, the toxicity of a chemical increases with an increase in their molecular weight (MW). For many molecules in the dataset with MW < 300 g/mol, a positive correlation of the MW with the toxicity values was observed. In contrast, a weak correlation was noted for the compounds with MW > 300 (187 molecules), so these compounds were removed.

For model H1 (Figure 9b), a descriptor (predictor) selection based on prior knowledge was performed, such that descriptors (“MolLogP”, “AATSC0p”, “TPSA”, “ETA_dEpsilon_A”, “SHBD”, “Mi”) that were stated as having a strong influence on aquatic toxicity were selected to represent the molecular structures. The full names of the descriptors used for the model and introduced in Table 3 can be found in Table S3-7 of *Supporting Information 3* for **Paper V**. In the case of model H2 (Figure 9c), the toxicity value was estimated using either a prior knowledge-based (PKM) or a standard kNN model, depending on the octanol-water partition coefficients of the molecules (log P). Large errors were observed for molecules that had log P values lower than 1.5 and higher than 4.0 when a standard kNN model was used to predict toxicity. Thus, the standard kNN approach was complemented by a PKM model for toxicity estimations of the molecules with the specified log P values. Model H3 (Figure 9d) applies the PKM model to assist the kNN model in selecting the neighbors; molecules with the closest estimations (smallest difference between the predictions) to those obtained by the PKM model are considered the nearest neighbors. The hybridization H4 (Figure 9e) applies a modeling post-assessment technique. A set of rules derived from prior knowledge is used to evaluate the final toxicity prediction models. The rules are presented in Table 3. More details on the knowledge extraction and its use for predictive modeling can be found in **Paper V**.

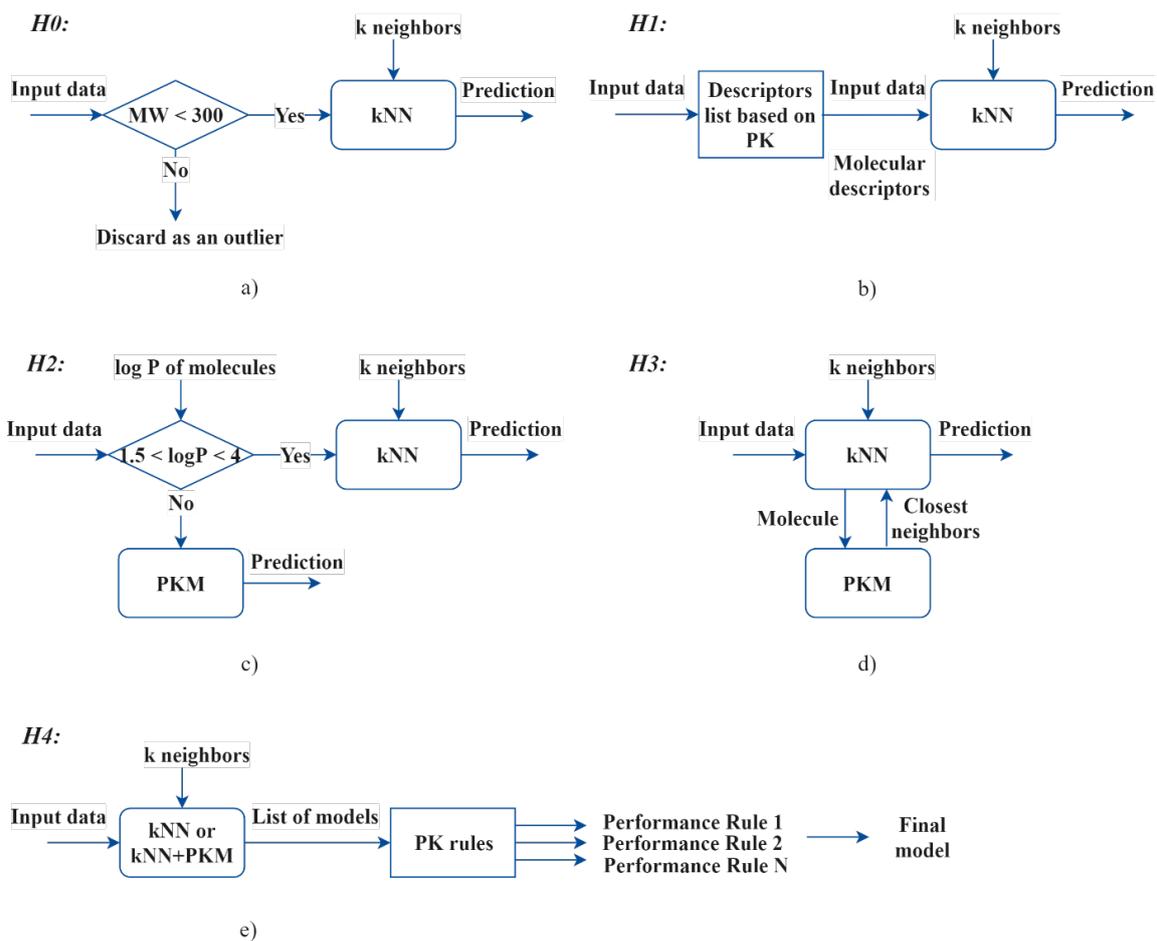


Figure 9. Schematic representation of the use of prior knowledge in the development of hybrid models. PK, prior knowledge; PKM, prior knowledge model (GC+QSAR)

Table 3. Set of rules identified for the dataset and used for evaluation of the performance of the models.

Main toxicity trends	Expressed in descriptors
Toxicity increases with hydrophobicity ^{100,101}	Toxicity increases with an increase of MolLogP (RDkit)
Toxicity increases with polarizability ¹⁰¹⁻¹⁰³	Toxicity increases with an increase of molar refractivity MR (RDkit)
	Toxicity decreases with an increase of GATS1p (PaDELPy) Toxicity increases with an increase of AATSC0p (PaDELPy)
Toxicity has a negative correlation with topological polar surface area ^{101,104}	Toxicity decreases with an increase in TPSA (RDkit)
Most of the toxic compounds act as hydrogen-bonding acceptors, while the least-toxic compounds act mainly as hydrogen-bonding donors ^{105,106}	Toxic compounds have lower SHBd (PaDELPy)
	Toxic compounds have lower maxHBint2 (PaDELPy)
There is a positive effect of unsaturation and electronegative atom count ¹⁰⁷	Toxicity decreases with an increase of ETA_dEpsilon_A (PaDELPy)
Toxicity decreases with ionization potential increase ^{102,103}	Toxicity decreases when Mi (PaDELPy) increases
	The larger the "GATS1i" (PaDELPy), the less likely the compound will be to react and generate toxicity ¹⁰³
Molecular size and bulk have positive influences on toxicity ^{101,107-109}	With an increase of MW (RDkit), the toxicity increases
	Toxicity is higher for higher values of ETA_Alpha (PaDELPy)
There is an inverse effect of branching on toxicity ^{101,104,107,110}	Toxicity decreases with an increase of ETA_EtaP_B (PaDELPy)
Toxicities of primary, secondary, and dimethyl tertiary amines increase with increasing chain length ¹¹¹	Toxicity of molecules containing N or amine group increases if the number of carbon atoms increases
Toxicity increases with the increasing alkyl chain length in ethoxylates ¹¹²	Toxicity of molecules containing the methoxy group increases if the number of carbon atoms increases
Substitution of H-atom with a methyl group (-CH ₃) on the N-atom reduces the toxicity of amine surfactants ¹¹³	The toxicity of molecules decreases with the number of N-CH ₃ fragments

4 Results and Discussion

The chapter summarizes the main results of the work of this thesis. The first part discusses aspects of the impact assessment at the substance level. The effects of integrating EHS hazard and LCA impact criteria into the CAMD procedure and the usage of prior knowledge for predictive modeling are described. The second part addresses the process level assessment. The elements with potentially significant influences on the environmental health and safety performance are identified for phase-change carbon capture systems.

4.1 Role of impact assessment in the design of safer chemicals

Pareto fronts for some of the property categories obtained during the simultaneous CAMD and integrated sustainability assessment of the molecules are visualized as Case 2 in Figure 10. Case 1 depicts the results of the CAMD without the use of the sustainability-related indices, instead performing the impact assessment of the molecules obtained through CAMD as a post-design step. The best-performing structures are presented in Table 4. The use of the sustainability criteria led to the identification of a broader set of structures in the obtained Pareto front (~100 structures), as compared to having only thermodynamic properties as objective functions (Case 1, ~40 structures). The search, however, was found to be biased towards OH-containing molecules. Although these molecules are preferable from the sustainability point-of-view, they might be less-beneficial in terms of phase-change due to their higher polarity and, consequently, greater miscibility in water as a result of the strong hydrogen bonds of the OH-group and water (Figure 11).

Nevertheless, OH-containing molecules were identified as possible solutions, in that they satisfied the constraint of the Hansen solubility parameter difference to water, indicating a phase-change behavior. Thus, molecules with the -OH group may still exhibit the phase-change behavior, albeit to a lesser extent. This idea remains to be tested with advanced group-contribution methods and/or experiments.

The increased polarity and ability to form strong hydrogen bonds reduce the mobility, boiling point, and flash point of an OH-containing molecule, thereby decreasing the hazards to health and safety. Furthermore, the increased hydrophilicity due to polarity makes OH-group-containing molecules less-toxic for aquatic life. Such an effect can be explained by the hydrophilic groups being less likely to attach to the lipids of living cells¹¹⁴. Superior performances of the OH-containing molecules (alkanolamines) in terms of GWP and the EI99 index have also been observed. These may be linked to the substitution of an alkyl for an OH- group, which reduces the number of potential isomers that contain quaternary carbon- or nitrogen-bearing centers (atoms with four substitutes). The quaternary centers are characterized by higher impact values, as they are difficult to synthesize^{115,116}. In contrast, the CED values of the alkanolamines (estimated by the

FineChem model) were found to be slightly higher than those predicted for the alkanamines. This indicates a more-energy-intensive alkanolamine production process (compared to those for alkyl-containing compounds) or it may reflect uncertainty introduced by the FineChem prediction.

Phase-change characteristics that determine the energy and cost penalties of the CO₂ capture process are the main drivers for adopting the technology. Thus, this solvent feature may need to be prioritized. In such a case, the preferred route maybe the decomposed approach, in which strong phase-change candidates are identified by means of CAMD as a first step and then a second step of further screening is performed according to the results of the impact assessment. However, the simultaneous approach that resulted in a wider set of structures in the Pareto fronts still contained structures similar to those proposed by the decomposed approach. Furthermore, new OH-containing molecules that exhibited the phase-change behavior became apparent. These kinds of molecules have not been studied to the same extent as alkane- or other types of non-OH amines, though they might be worth investigating in the future. The detailed analysis of the results obtained for both cases is presented in **Paper II**.

The process whereby the sustainability framework is integrated into the CAMD revealed the lack of reliable *in silico* methods to estimate the properties of the molecules that have missing experimental data. This prompted the idea of developing the hybrid prediction models that use knowledge existing in the field. The first step in the development of the hybrid models, namely knowledge extraction, illuminated the trends described above.

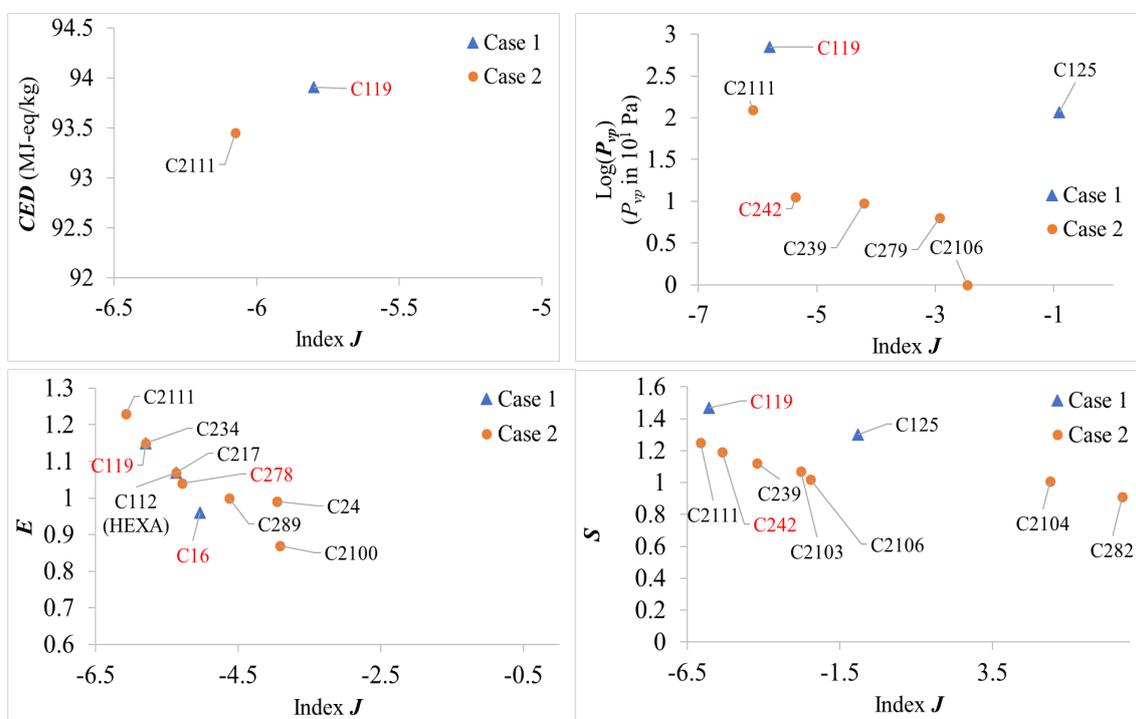


Figure 10. Pareto points of the selected properties as a function of the J index for the molecular structures obtained after CAMD in Case 1 (triangles) and Case 2 (circles). The ID numbers of the molecules shown in red indicate high-performing options. Source: Paper II.

Table 4. Best-performing options among all the structures designed in Cases 1 and 2, with *J* score lower than -5. Adapted from Paper II.

ID	Groups	Group frequency	<i>GWP</i> [kg CO ₂ - eq/kg]	<i>CED</i> [MJ- eq/kg]	<i>EI99</i> [Pt]	<i>S</i>	<i>H</i>	<i>E</i>
C2111	[CH ₃ ,CH ₂ ,CH ₂ NH ₂]	[1,6,1]	3.58	93.45	0.3	1.25	0.18	1.23
C234/ C119	[CH ₃ ,CH ₂ ,CH ₂ NH ₂]	[1,5,1]	3.53	93.91	0.29	1.47	0.26	1.15
C224/ C123 (DEEA)	[CH ₃ ,CH ₂ ,OH,CH ₂ N]	[2,3,1,1]	3.31	107.38	0.23	1.49	0.19	1.24
C217/ C112 (HEXA)	[CH ₃ ,CH ₂ ,CH ₂ NH ₂]	[1,4,1]	3.48	94.22	0.28	1.72	0.35	1.07
C242	[CH ₃ ,CH ₂ ,OH,CHNH]	[2,4,1,1]	3.24	108.11	0.24	1.19	0.26	1.27
C278	[CH ₃ ,CH ₂ ,CH,OH,CH ₂ N]	[3,3,1,1,1]	3.62	117.11	0.32	1.19	0.08	1.04
C227	[CH ₃ ,CH ₂ ,OH,CHNH]	[2,3,1,1]	3.31	107.38	0.23	1.32	0.34	1.23
C16	[CH ₃ ,CH,OH,CH ₂ N]	[3,1,1,1]	3.38	106.67	0.23	1.86	0.34	0.96
C256 (DBA)	[CH ₃ ,CH ₂ ,CH ₂ NH]	[2,5,1]	3.58	93.45	0.3	1.54	0.18	1.07
C116 (DPA)	[CH ₃ ,CH ₂ ,CH ₂ NH]	[2,3,1]	3.48	94.22	0.28	2.07	0.38	1.05

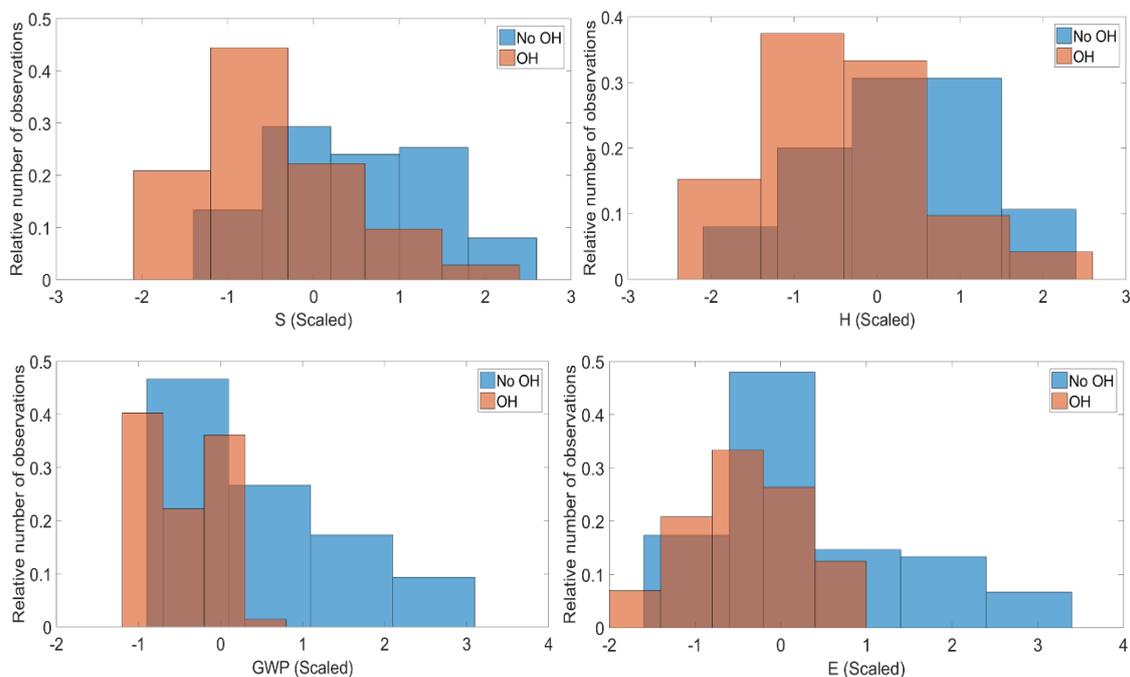


Figure 11. Histograms showing the molecular distributions for OH-containing and non-OH-containing molecules in the selected property categories. The relative number of observations is the number of observations in the bin, divided by the total number of observations. Source: Paper II.

The process of simultaneous CAMD and impact assessment was extended to include additional functional groups to design cyclic and bi-functional (with two amino groups) amine structures. This led to the identification of 120 solvent structures which were expanded into 624 isomers¹¹⁷. Among the designed structures were experimentally verified phase-change solvents such as DMCA (N,N-dimethylcyclohexanamine), MCA (N-methyl-cyclohexanamine), DPA (di-n-propylamine), HEPTA (heptanamine), and SBPA (N-sec-butyl-n-propylamine)¹¹⁸. This confirmed the efficiency of the proposed approach for designing phase-change solvents that exhibit desirable CO₂ capture performance. Cyclic molecular structures were also identified. The cyclic, multi-functional amines were reported to have higher absorption capacities than the mono-functional options, while exhibiting higher boiling points, thereby reducing the mobility and vapor losses of the solvent^{118,119}. The analysis of the cyclic structures resulted in the selection of a novel solvent alternative – S1N (cyclohexylpropane-1,3-diamine) – due to its structural resemblance to the identified cyclic candidates but with a considerably lower price. Combined with DMCA, the solvent mixture exhibited high cyclic capacity and low regeneration energy requirements, in similarity to other good phase-change solvents. Further details of the solvent selection procedure and experimental testing results can be found in the work of Papadopoulos et al. (2021)¹¹⁷.

4.2 Prior knowledge for predictive modeling

The results of the systematic extraction and collection of knowledge of the acute aquatic toxicity domain are presented in this section. The effects of the this extracted knowledge in predictive modeling are also demonstrated and analyzed.

4.2.1 Extraction and classification of knowledge

The main advantage of the partly automated literature review was a significant reduction of the text for initial reading (>85%) (Table S1-1 *Supporting Information of Paper V*). Most of the sentences extracted by the automated text-mining procedure were useful, in that they could be used for hybrid modeling or pointed out specific parts of the initial article for subsequent manual screening. Overall, the method did not require extensive knowledge of the field, as only some prior understanding was needed to define the main terms and connection words that would guide the search for relevant information.

Figure 12 presents the classification scheme based on the information collected using the proposed procedure for knowledge extraction. Despite limitations as to the number and types of articles used for the knowledge mining, many important aspects of toxicity were identified, that could be useful for use in hybrid modeling or other applications.

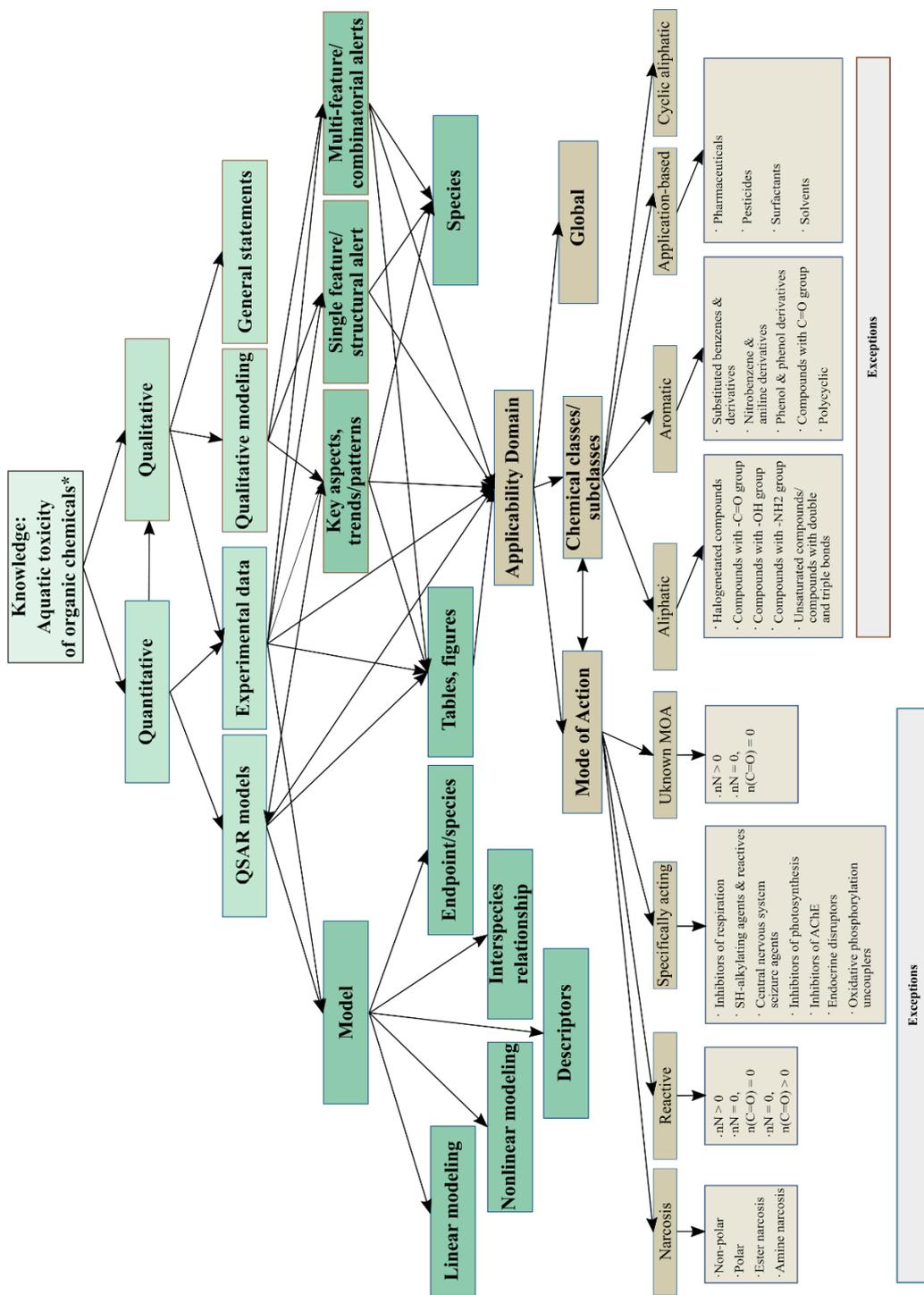


Figure 12. Knowledge classification scheme for aquatic toxicities of chemicals (*excluding inorganics, metals and metallo-organic compounds, ionic liquids, epoxides, peroxides, and mixtures). Source: Paper V.

QSAR models. The QSAR models that arose from the data acquisition approach outlined in Section 3.3.3 were developed for prediction of the toxicity value or classification of chemicals according to different toxicity levels and MoA classes. Most of the QSARs applied linear modeling (e.g., multilinear regression, principal component analysis, linear partial least squares, ordinary least square method), due to its simplicity and interpretability. The nonlinear models often showed higher levels of accuracy than the linear models developed using the same set of chemicals. The nonlinear models included, for instance, support vector machine (SVM), k-nearest neighbors (kNN), neural networks (NN), tree-based methods and gradient boosting (GB). A general outcome from the QSAR studies was that ensemble or consensus models that combined several methods outperformed the models based on a single method^{120–123}. Improved performance was also observed when similar chemicals were grouped on the basis of MoA¹²⁴ or other similarity criteria^{124,125} before developing the prediction model.

A wide variety of descriptors was used in the collected QSAR models. The majority of the QSAR studies applied single descriptors as the predictor variables, and only a few later studies worked with molecular fingerprints. The fingerprints resulted in a better performance^{106,126}. However, when the amount of training data was limited, single descriptors seemed to be more beneficial to use¹⁰⁶.

The descriptors with the strongest impacts on the acute toxicity value were found to be associated with: hydrophobic features, i.e., log Kow, log P, log D, Crippen log P, B08[C-C]; electrophilicity, i.e., ELUMO, Amax; polarizability, i.e., α , GATS1p; acceptors and donors of hydrogen bonds, i.e., Ca, NHdon Hacc, polar group descriptors; molecular size and branching, i.e., Vm, ElipVol, RDCHI; and polar surface area, i.e., TPSA. According to Gramatica et al. (2018)¹²⁷, the nX (number of halogen atoms) and nBondsM (number of multiple bonds) descriptors, which are related to halogen substitution and unsaturation, increase the PBT behaviors of chemicals. The descriptors linked to the decrease in PBT index were MAXDP2 (maximal electrotopological positive variation) and nHDonLipinski (number of donor atoms for H bonds). These two descriptors encode the ability of a chemical to form electrostatic and dipole-dipole interactions, as well as hydrogen bonds in the surrounding environment. Hossain & Roy (2018)¹²⁸ and Önlü & Saçan (2018)¹²⁹ have developed QSAR models for Contaminants of Emerging Concern (CECs), including for instance pharmaceuticals, personal care products, pesticides, and surfactants. They have concluded that the toxicities of CECs are mostly related to hydrophobicity^{128,129}, aromaticity¹²⁸, polarizability, and molecular size and shape¹²⁹. Ionization was shown to affect the bio-uptake and mechanisms of interaction with macromolecules at the target sites^{130,131}. Descriptors with a low impact can also be collected by changing the objective of the work and adding connection words that allow to identify also the low-impact descriptors.

Various species were used to obtain toxicity information. The most commonly used species in the QSAR and other types of studies were the following (with the most frequently used toxicological dose descriptors): the algal *Tetrahymena pyriformis* (*T. pyriformis*) (IGC50); crustacean *Daphnia magna* (*D. magna*) (EC50, LC50); and fish

Pimephales promelas (fathead minnow) (LC50). The *T. pyriformis* dataset is often preferred due to its size, molecular diversity, and high-quality data¹³². The interspecies quantitative structure–activity–activity relationships (QSAARs) can be regarded as a separate class of the QSAR models. These models are typically based on a small volume of data and have a linear functional form with very few predictor variables. Good performance ($R^2 > 0.7$) was observed for the species mentioned above^{132–134}. The fish-based model was considered to be superior for predicting missing toxicity data (i.e., for *T. pyriformis* and *D. magna*)¹³².

Identified toxicity alerts, trends, and patterns. The extracted data suggest a consensus among researchers that acute toxicity is defined by the mode of toxicological action and the chemical characteristics¹³⁵. The increased toxicity has often been reported to be correlated with increased lipophilicity^{101,136,137}. The toxicities of the most-toxic compounds were related to both hydrophobicity and their behaviors as hydrogen-bonding or electron-acceptors (e.g., hydrophobic nitroaromatic compounds with halogen and amino substituents^{104,105,138–140}). Khan et al. (2019)¹⁴¹ have concluded that if a hydrophobic group is necessary during the design of a drug compound, a higher polarity substitution should be preferred. Voutchkova et al. (2011)¹⁴² have suggested using a $\log P_{o/w}$ of < 2 and ΔE (LUMO-HOMO) > 9 eV to improve the probability of designing a compound with low aquatic toxicity.

Specific functional groups, such as cyano¹³⁵, isothiocyanate¹⁴³, and halogens^{136,144–146}, enhance the toxicities of molecules. However, the extent of the increase appears to be dependent upon the molecular structure and position of the group in the molecule. A study carried out on *T. pyriformis* (IGC50)¹⁴⁵ found that mono-isothiocyanate analogs with a branch hydrocarbon moiety, especially in the β -position, were less-toxic and less-reactive. Di-isothiocyanates and the allyl- and propargyl derivatives were more-toxic and more-reactive than their 1-alkyl homologs¹⁴⁷. A significant increase in toxicity was observed for compounds with a halogen in the α -position to the cyano group: $C(\#N)[CX_4]$ [F, Cl, Br, I]¹⁴⁵ and carbonyl group^{145,148}. All halogens are reported to increase toxicity. In the case of the addition of non-activated halogens, the contribution to toxicity is equivalent to its hydrophobicity. However, the contribution to toxicity is much more significant if the halogen is activated (e.g., adjacent to an ester or other unsaturated bond)^{143,145}. Analysis of the data for the bacterium *Vibrio fisheri* and ciliate *T. pyriformis* showed that branching, especially at the α - or β -carbon, inhibited the reactive center sterically, thereby decreasing the reactivities of the molecules^{143,149}.

Among the other toxicity alerts found in the literature were amino groups, the presence of additional (one or more) aromatic rings with highly electronegative substituents close to each other (5–7 Å apart)¹⁵⁰, nitro group, nitrile, disulfide, phosphoric acid derivatives, pyrazolyl group, and formamide groups¹⁵¹, ring aromaticity, sulfur, aromatic esters and vinyl moiety¹⁰⁴, double and triple bonds, and acrylate groups¹⁵², to name the most frequently encountered. As in the case of the halogenated aliphatics, the magnitude of the toxicity increase was dependent upon the position of the toxicity alert, i.e., a structural fragment having a direct influence on the toxicity of a molecule. For

example, in the substituted benzenes, the nature and position of the substituent in the ring defined the toxicity. Aromatic compounds with two hydroxyl and/or amino groups in the ortho and para orientation displayed increased toxicity owing to their ability to oxidize to more-toxic Michael-type acceptors^{145,153,154}. The ortho-substituted phenols, anilines, and nitrobenzenes exhibited lower levels of toxicity than the corresponding meta and para-substituted compounds^{139,155–158}.

Applicability domains based on the extracted knowledge. Affinity for a specific chemical class or MoA was often seen as a critical determinant for predicting and understanding chemical toxicity^{159–162}, with MoA being more challenging to determine¹⁶³. The most-covered applicability domain in MoA seems to be non-polar and polar narcosis, followed by specifically acting chemicals. Within the specifically acting chemicals domain, many studies have addressed chemicals with electro(nucleo)philic reactivity, weak acid respiratory uncouplers (WARUs), and anticholinesterase activity.

The chemical classes that were most highly represented in the collected knowledge were nitrobenzene and phenol derivatives, pesticides, pharmaceuticals, and halogenated aliphatics. Other chemical classes, such as aliphatic alcohols, amines, amides, and acids, were represented to a lesser extent, probably because their toxicity effects are instead studied in the context of a particular MoA. Compounds with double and triple bonds, such as vinyl/allyl group-containing chemicals, nitriles, propargyl alcohols, carbonyl-containing α,β -unsaturated chemicals, carbamates, and quinones have often been examined separately, likely due to their reactive nature^{110,146,164–166}.

Most of the knowledge collected under the applicability domain termed ‘global’ was related to the QSAR models^{100,120,121,125,163,167–175}. Despite the clear benefits of assigning compounds to certain chemical classes or MoAs, the different behaviors of chemicals even within the same chemical category and the inability to assign some (especially newly developed or untested) chemicals to a specific MoA class¹⁷⁶ motivated the researchers to build models that were not limited by chemical class or MoA.

The academic field of aquatic toxicity is very diverse and extensive, both from the quantitative and qualitative perspectives. On the one hand, this diversity fosters the collection of relevant information, and is therefore useful for predictive modeling. On the other hand, the wide variety of descriptors used in the studies, variable quality of the toxicity data, and limitations linked to the applicability domain, to name some important factors, make it quite challenging to apply directly the knowledge without analyzing the available information and constraints associated with its use. Thus, mapping and evaluating the domain knowledge before its application are required to facilitate the navigation of the data. Based on this mapping, some strategies for using prior knowledge in predictive modeling are described in the next section.

4.2.2 Hybrid modeling

The performances of the standard and hybrid (with the use of prior knowledge) kNN models are presented in Table 5. It is evident that the hybridization improves the coefficient of determination (R^2) and Spearman's correlation coefficient (Spr_m) for the standard descriptor-based models (DESC). The fingerprint-based models (FPNs), both standard and hybrid, show lower R^2 and Spr_m coefficients. The lower R^2 and Spr_m values can, to some extent, be explained by the PKM model being built on a different set of data before integration into the hybrid models H2 and H3.

The heatmap in Figure 13 illustrates how closely the predictions made by the models follow the rules presented in Table 3. It is clear that the dataset toxicity values (Data) show better correlations with some of the rules but not with others. The rules related to ETA_EtaP_B (branching) and methoxy group (nC methoxy) are not supported by the data. Poor correlation of the data regarding the ethoxylates might imply a different behavior of the methoxy group-containing molecules compared with the remainder of the dataset. A low diversity of the dataset in terms of branching might result in the low correlation of toxicity values with the ETA_EtaP_B descriptor. The toxicity estimations made by the standard descriptor-based models (DESC_2 and DESC_8) show similar trends to the Data, while the fingerprint-based models (FPN_2 and FPN_7) show lower levels of compliance with the rules.

The values for the hybrid models often follow the rules better than the toxicity values obtained through the standard models. A very low correlation of the H0 (dataset of molecules with $MW < 300$) model predictions with amines containing a $>N-CH_3$ fragment might indicate that these amines are the high-molecular-weight molecules that exhibit weak correlations with the MW.

The performance of the H1 models using a reduced set of predictions is similar to or slightly lower than the rest of the descriptor-based models and data. This may suggest that a limited number of descriptors can fully represent the molecules in the dataset. The involvement of the PKM model based on a GC method may explain the better correlation between the toxicities predicted by hybrid models H2 and H3 and specific fragments (i.e., number of carbon atoms in the amine- and methoxy group-containing molecules). However, models H2 and H3 exhibit worse performances in relation to considering the electronegativity and topology of molecules (lower correlations for GATS1p, TPSA, GATS1i, ETA_dEpsilon_A), as compared to the standard models. In summary, we show how the use of prior knowledge as a post-analysis step for the evaluation of an ensemble of models provides an additional measure with respect to model interpretability. This can be used along with traditional prediction accuracy measures towards the selection of a quantitative and qualitative multi-criteria model. A more detailed analysis of the results is given in **Paper V**.

Table 5. Summary of the performance of the models in terms of the coefficient of determination R^2 and Spearman correlation coefficient Spr_m . DESC and FPN specify the descriptor-based and fingerprint-based models, respectively. H0, H1, H2 H3 are the applied types of hybridization. The designations _2 to _14 indicate the numbers of closest neighbors used for the prediction.

Model	R^2	Spr_m
DESC_2	0.83	0.94
DESC_4	0.85	0.95
DESC_5	0.85	0.95
DESC_6	0.86	0.95
DESC_8	0.86	0.95
DESC_H0_3	0.95	0.98
DESC_H0_7	0.95	0.98
DESC_H1_2	0.92	0.98
FPN_2	0.70	0.84
FPN_5	0.74	0.86
FPN_7	0.74	0.86
FPN_12	0.73	0.86
FPN_14	0.73	0.86
FPN_H2_2	0.43	0.80
FPN_H2_7	0.45	0.80
FPN_H2_12	0.46	0.80
FPN_H3_12	0.52	0.76
FPN_H3_14	0.53	0.76

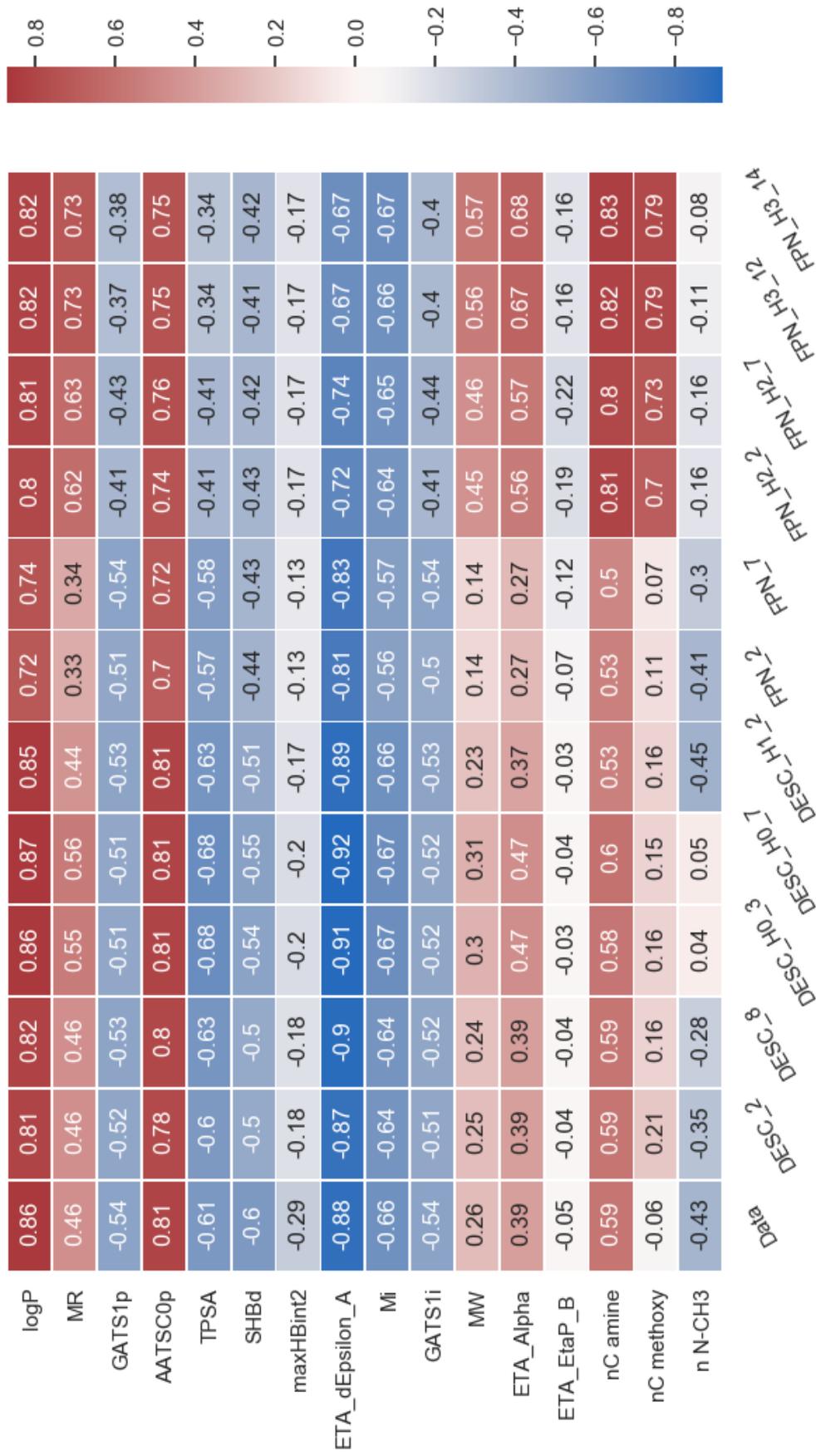


Figure 13. Spearman's correlation coefficient between the descriptors and toxicity predictions made by the models. Red: positive correlation with toxicity values; blue: negative correlation with toxicity values. Only two descriptor and fingerprint-based models (best and worst) are shown due to the similar performances of the remainder of the standard models with respect to the presented values. Source: Paper V.

4.3 Process level impact assessment

The impact assessment at the process level evaluated the performances of the phase-change solvent-based systems in mitigating the emissions from two reference plants: a quicklime plant, producing 150 tonnes of CaO per day, and a 400-MW natural gas-fired power plant. The compositions of the flue gases differed with respect to their contents of CO₂, O₂, NO_x, and SO_x. The flue gas from the quicklime plant contained 3-times (mol-%) more CO₂ and significant amounts of SO_x and NO_x, 350 and 400 mg/Nm³, respectively, as compared to the flue gas exiting the power plant, which contained a much lower level of NO_x (21 mg/Nm³) and no SO_x but twice as much O₂. Both the gases contained soot.

CO₂ capture processes with two selected solvents for the two reference plants were designed using a shortcut model of an absorption/desorption system for CO₂ capture. The model provided an assessment of the reboiler duty required in the desorber¹⁷⁷. The shortcut model accounted for the non-ideal behavior of the solvents by considering the vapor-liquid-liquid phase-equilibrium (VLE) relation for each solvent. This relation, obtained from the experimental data collected during the pilot plant testing, was used in the model to calculate the energy requirements of the desorption based on the various process operating parameters, such as the mass flowrate, the solvent lean- and rich-loading temperatures, and the pressure. The processes were optimized for the solvents based on minimization of the total cost of the process¹⁷⁷.

Two phase-change solvent alternatives were studied: a novel mixture of S1N and DMCA identified through the CAMD¹¹⁷; and the known phase-change solvent MCA¹¹⁸. S1N, which contains both primary and secondary amine groups, serves as an absorption activator in the solvent mixture, while DMCA acts as a promoter of regeneration¹¹⁷. During the phase-separation, hydrophilic S1N tends to accumulate in the water phase, while hydrophobic DMCA is concentrated in the organic phase.

In the S1N+DMCA-based capture system, the CO₂-rich flow contains significant amounts of water, whereas the CO₂-lean phase, which is recycled back to the absorber, carries half the amount of CO₂ and a small amount of water¹⁷⁸. In comparison, the CO₂-lean flow of the MCA system consists of water and a very small amount of CO₂. The S1N+DMCA and MCA solvents capture systems differ in terms of the placement of the liquid-liquid phase-separator. In the process that uses S1N+DMCA, the phase-separator is placed before the intermediate heat exchanger because the liquid-liquid phase-split occurs at 40°C. The phase-separator is placed after the intermediate heat exchanger in the MCA process because the phase-split appears at 90°C.

Both the S1N+DMCA and MCA solvents showed beneficial performances in terms of rich loading and cyclic capacity compared to the performance of MEA. The main operating parameters of the capture systems are presented in Table 6. More details on the generation of the process flowsheets can be found in the work of Kazepidis et al. (2020)¹⁷⁷ and in **Paper III**.

Table 6. Critical operating parameters of the optimized systems¹⁷⁸. Source: Paper III

	S1N+DMCA Quicklime plant	S1N+DMCA Power plant**	MCA Quicklime plant	MCA Power plant
Solvent per CO ₂ captured, kg/tonne CO ₂	11000	10000	8000	14000
$\frac{G_{mol}}{L}$, mol/l*	6.7	23.5	7.93	16.1
Temperature in the absorber, °C	45	45	73	71
Solvent regeneration temperature, °C	90	85	106	101
Reboiler duty, MJ/tonne CO ₂	2340	2083	2346	2091

*Ratio of inlet flue gas flow to circulating liquid flow rate.

**The lower solvent flow for the same plant case is a consequence of economic optimization of the flowsheets, which decreases the solvent flow in relation to its price, rather than the solvent properties. S1N is considerably more expensive than MCA.

To perform the impact assessments of the designed capture units, the material flows derived via the optimization procedure¹⁷⁸ were adjusted to incorporate important LCA and EHS aspects that were not tested experimentally, such as loss of the solvent due to degradation aerosol formation, and solvent reclaiming. Furthermore, the material flows of the auxiliary units (e.g., pre-treatment of the flue gas for the power plant case, washing of the treated flue gas, and wastewater treatment) were added.

Degradation of the solvent molecules is one of the main causes of solvent loss during the capture process. Estimation of solvent loss early during the design of the process relies on the availability of experimental studies, which are limited for novel materials such as phase-change solvents. In the cases of MCA and DMCA, experimental tests have been conducted by Zhang (2013)¹¹⁸ to assess the oxidative and thermal degradation of the molecules. According to the results, MCA degrades oxidatively at a 1.2–3.0-fold higher rate and thermally at a 2.0-fold higher rate than MEA. DMCA, which is a tertiary sterically hindered amine, is resilient to both oxidative and thermal degradation, degrading at lower rates than MEA. As there are no experimental data on the degradation behavior of S1N, the degradation behaviors of the structurally similar molecules MAPA and MCA were used for the analysis. Thus, the oxidative degradation rate were assumed to be between 3-fold and 10-fold higher than the degradation rate of the MEA.

Allylamine and propyl formamide were taken as the main degradation products of S1N. Ammonia and other amines were assumed to be the degradation products of DMCA,

and ammonia and cyclohexanone oxime were assumed to be formed during the degradation of MCA. Based on these assumptions, different scenarios were assessed.

The purge of the lean solvent stream going to the reclaimer was adjusted to keep the steady-state concentration of the non-volatile degradation products at less than 1.5 wt. % in the circulating flow. The reclaimer waste flows were estimated, assuming a 5% loss of the purged flow in the reclaimer¹⁷⁹. To remove the volatile degradation products in the flow of the treated flue gas leaving the absorber, washing units were modeled with the help of the Aspen Plus simulation software. Acid and water washes were assumed for the MCA system to neutralize ammonia. Only the water wash was modeled for the S1N+DMCA system to capture allylamine, which is volatile and highly soluble in water. The flows in the washing section were estimated based on the assumption that aerosol formation would increase the volatile emissions by one order of magnitude. Still, most of these emissions were avoided due to the use of a NO_x/SO_x wet scrubber. The NO_x/SO_x scrubber was assumed only for the power plant flue gas that contain high levels of SO_x and NO_x.

The detailed analysis of the possible degradation behaviors of the solvent molecules, the degradation scenarios considered, and all the introduced assumptions can be found in **Paper III**.

4.3.1 LCA

The results of the LCA performed for the carbon capture phase-change systems with the selected solvents are presented in Figure 14. The net CO₂ savings ranged from 0.33 to 0.47 kg of CO₂ per kg of CO₂ captured for the various investigated cases. It can be seen that the main contributor to the overall environmental impact for all the systems in all three life-cycle impact categories is the steam used for solvent regeneration, followed by the electricity required to deliver seawater for the cooling processes. However, the impacts of the remaining contributors vary depending on the system and the impact category. The flue gas of the power plant is characterized by a lower level of CO₂ (3.6 vol%), as compared to the flue gas of the quicklime plant (12.3 vol%). This explains the almost 3-fold higher impact coming from the flue gas blower and the increased amount of cooling media needed to reduce the temperature of the flue gas to the temperature required for the absorption process (40°C).

The lowest overall impact was observed for the MCA quicklime plant-based capture system, which has the smallest solvent loss due to degradation and the lowest solvent flow required to capture 1 tonne of CO₂. The MCA quicklime plant system also performs best of all the cases when the highest solvent degradation rates are applied (Figure 14, dashed lines). In contrast, the S1N+DMCA system demonstrates the most-severe variations in performance level, depending on the highest or lowest solvent degradation rates being applied for the assessment. For the case with the minimum degradation rate, the strongest impact is observed for the MCA power plant, which has the highest flow rate of solvent per tonne CO₂, while the environmental performances of the other three systems are similar. The high rate of solvent degradation has a significant

effect on the impact of the S1N+DMCA system, whereby the worst-possible degradation rate is assumed to be 10-fold higher than the degradation rate of the conventional solvent MEA. In such a case, the increase in impact is explained by the increased loss through degradation, resulting in a higher flow rate of the makeup solvent. That places an additional burden on the reclaimer dealing with the products of the degradation, as well as the reclaimer waste treatment facilities.

It can be concluded that the use of renewable electricity and industrial waste heat (i.e., at temperatures in the range of 90-100°C) can decrease the LCA impact of the phase-change capture unit by 70%-90%. The remaining impact will be dominated by the degradation behaviors of the solvent molecules. The detailed analysis of the LCA results for various solvent degradation scenarios can be found in **Paper III**.

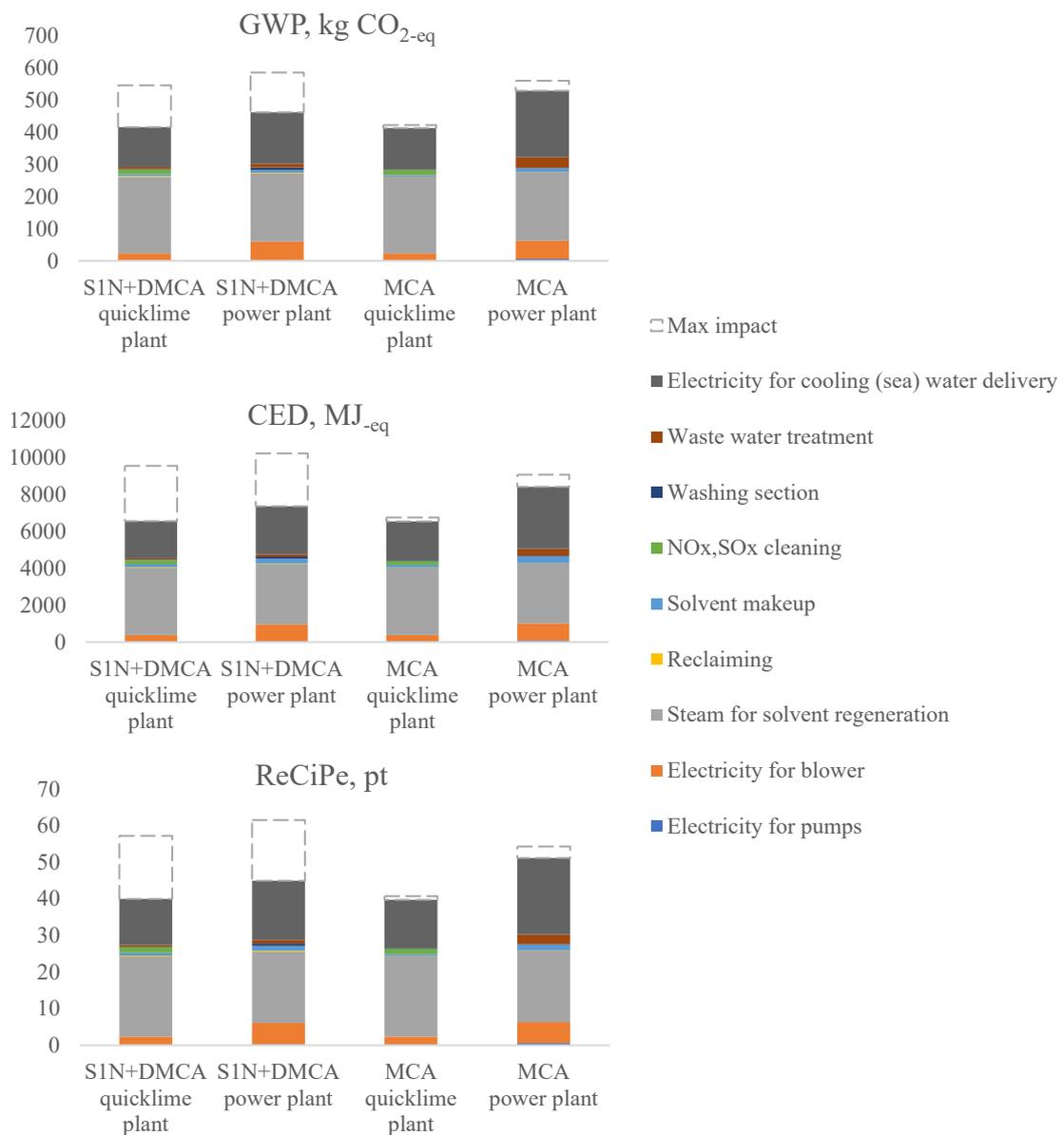


Figure 14. Contributors to the LCA impact per tonne CO₂ captured based on the scenarios with the minimum degradation rate of the solvent. Dashed lines indicate the increase in the impact when highest solvent degradation rate is applied. Adapted from Paper III.

4.3.2 EHS hazard assessment

The solvent molecules, the auxiliary process chemicals, and various compounds formed during the degradation of the solvent molecules can be hazardous for both the plant personnel and the environment. The EHS hazard impact of the chemical is defined according to the inherent properties and abilities of the chemical to cause harm to human health and the environment. It is also determined by the chemical's mass, which scales the inherent potential for hazard of the chemical. The contributions of every chemical present in the system to the potential of danger in some of the EHS subcategories are shown in Figure 15. It is clear that in terms of the mobility and water-mediated effects, the overall impact is dominated by the mass of the solvent molecules that are present in the system at the highest levels, compared to all the other chemicals. In comparison, the potential hazards associated with chronic toxicity and air-mediated effects are defined by an inherent property of the substances, namely the occupational exposure limits. That means that even low concentrations of some substances can be dangerous to human health.

The high mobility of the MCA, due to its lower boiling point compared to DMCA and S1N, explains it having the strongest impact in the Safety category. It is followed by the impact of the DMCA molecule that has a similar boiling point, and that dominates the overall impact of the S1N+DMCA carbon capture system. However, the impact of the S1N+DMCA is considerably lower due to the lower amount of mobile DMCA solvent needed to capture 1 tonne of CO₂ compared to MCA. Of the three solvent molecules, S1N has the highest boiling point, and exhibits the lowest hazard in terms of plant safety.

The large mass flows of the phase-change solvents circulating in the capture system increase the risk of the solvents being emitted to water sources, thereby affecting aquatic life. Considerable contributions to water-mediated effects are observed for sodium hypochlorite (NaOCl) and sodium nitrite (NaNO₂), substances that are present in the pretreatment of the quicklime plant flue gas. These chemicals are highly toxic to aquatic life.

High chronic toxicities for humans are expected from the allylamine, ammonia, and nitrosamines formed during the degradation of the solvent molecules, as well as from auxiliary chemicals such as sulfuric acid (H₂SO₄), NaOCl, and sodium hydroxide (NaOH) if people are frequently exposed to these substances. The health category is independent of mass, so the impacts are based on the inherent toxicity property of these chemicals.

The air-mediated effects rely on the chemicals' mass and toxicity properties. Thus, significant contributions are observed for the DMCA and S1N solvent molecules. MCA does not appear in the air-mediated effects because of its much lower chronic toxicity (PEL = 235 mg/m³) for humans, as compared to DMCA (PEL = 5 mg/m³) and S1N (PEL = 10 mg/m³). The major impact is expected to come from nitrosamines due to their known harmful effects on humans, even at very low concentrations (2.39 mg/m³²³). The highest steady-state concentrations of nitrosamines are expected to be seen for the power plant-

based capture systems due to their lower stripper temperatures and higher ratio of inlet flue gas flow to circulating liquid flow. The graphs for all the subcategories, together with the detailed analysis can be found in **Paper III**. The accumulation of nitrosamines was also separately studied in the exposure analysis part.

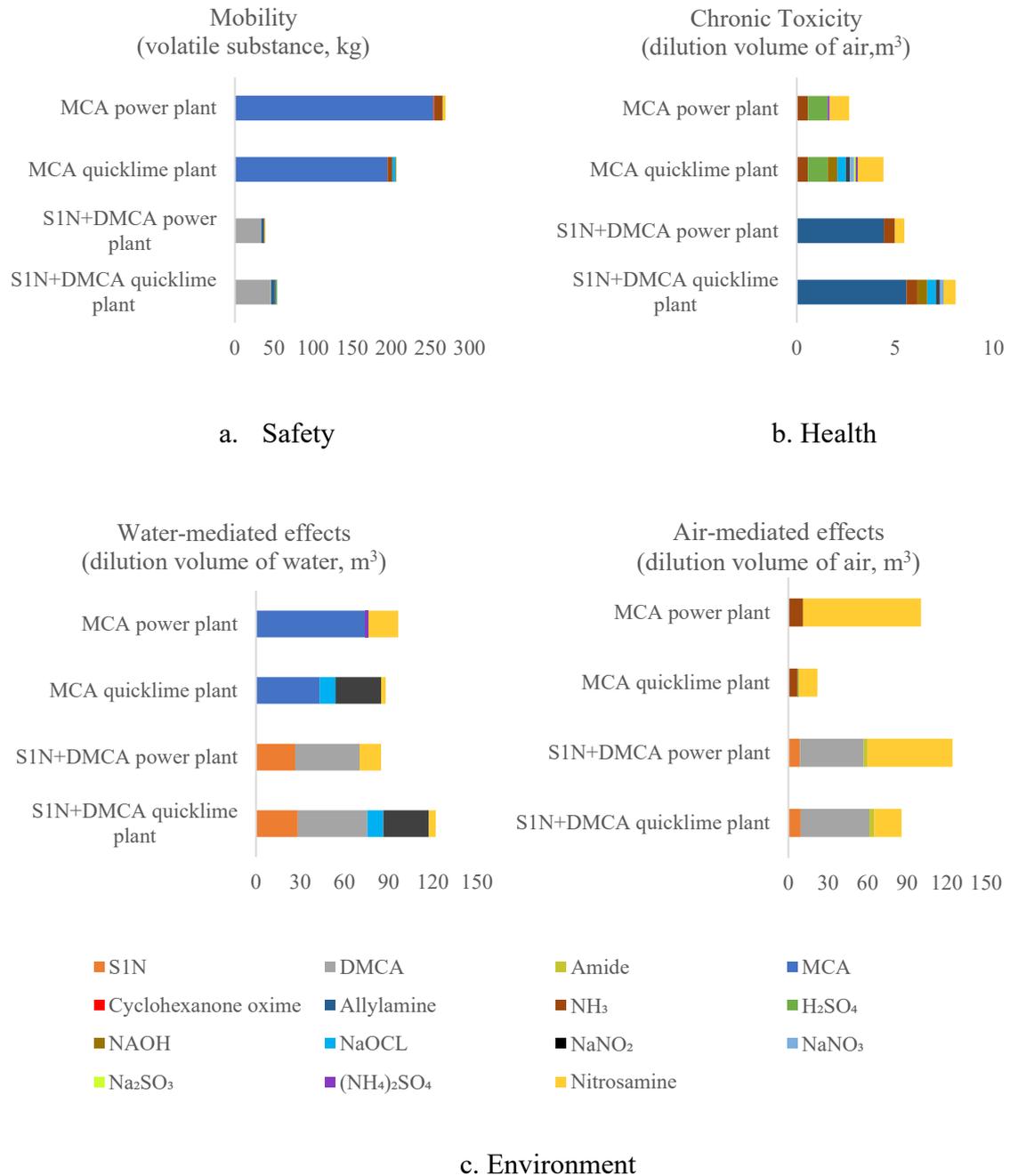


Figure 15. Examples of the potential of danger per tonne of CO₂ captured in the different categories (Health category uses a fixed mass of 1 kg) at NO_x= 5 ppm. CO₂, H₂O, O₂, and N₂ molecules are considered non-hazardous and are omitted. Note that the average values for the different scenarios are presented, and only chemicals with a contribution of more than 2% are shown. Source: Paper III

The multi-criteria assessment for the MCA and SIN+DMCA CO₂ capture systems are presented in Figure 16. The impacts of the studied carbon capture systems are normalized to the impact of the conventional MEA-based system^{23,180} (Table 7). The phase-change systems show beneficial performances in terms of the reboiler duty, the H, E, CED, and the GWP values, as compared to the MEA. However, it should be noted that the results are a first indication; additional experimental data, especially regarding solvent degradation, are required to ensure a more-accurate comparison.

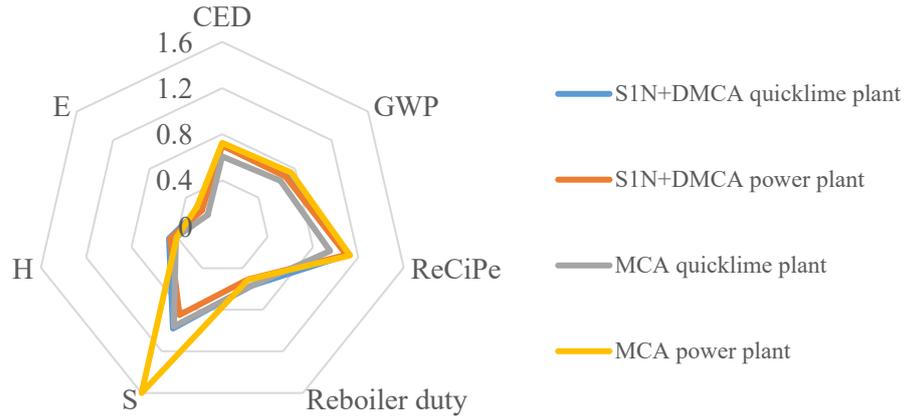


Figure 16. Multicriteria comparison of the environmental, health, and safety impacts of the different solvent systems for CO₂ capture, normalized to the impact of the conventional MEA system.

Table 7. MEA system impact data. All the values are per tonne of CO₂ captured

	Reboiler duty, GJ	S	H	E	CED, MJ _{-eq}	GWP, kg CO _{2-eq}	ReCiPe**, pt
MEA*	4.1	15,369.6	2.0	4,433.2	10,000.0	600.0	38.6

*A conventional MEA (30 wt%)-based carbon capture system at a solvent regeneration temperature of 120°C. The solvent-rich and lean-loadings are 0.49 and 0.2 mol CO₂/mol solvent, respectively¹⁸⁰.

**ReCiPe value is calculated from the EI99 value based on the correlation⁸⁸.

4.3.3 Exposure analysis

If leakages or spills of the solvent occur within the capture plant, the solvent concentration in inhaled air might pose a danger to the workers' health. The exposure analysis aims to determine whether leakages and spillage accidents pose a hazard to personnel working in close proximity to or farther away from the place of the accident. The analysis shows that for the MCA capture system, solvent leakages and spills up to 1,000 mL do not pose a hazard to humans. However, special care should be taken with spills of larger volumes, especially spills of pure MCA (Figure 17a). For the S1N+DMCA solvent system, the analysis demonstrates that leakages of pure solvents or solvent mixtures do not result in airborne concentrations of DMCA or S1N that would raise concern. However, spills of pure DMCA and solvent mixtures leaving the stripper might pose an immediate danger to workers' health (i.e., due to the resulting air concentrations of DMCA) when the spill volume is > 100 mL (Figures 17b-e).

The estimation of the steady-state concentrations of nitrosamines in the circulating flow of the solvent show that nitrosamines can reach and exceed the safety threshold values (13.7-14.0 mM^{23,89}), leading to a dangerous concentration of the nitrosamines in the working environment. For the power plant-based CO₂ capture case, this can happen already at NO_x concentrations of 5 ppm for the S1N + DMCA system and 10 ppm for the MCA system. The lower NO_x limit for the power plant-based S1N+DMCA system can be explained by the higher gas-to-liquid ratio and the slightly lower stripper temperature. The detailed analysis can be found in **Paper III**.

The process level assessment shows that phase-change solvent systems have the potential to act as a better alternative to the conventional amine solvent systems, due to the reduced reboiler duty and first indications of a lower impact on the environment in terms of short-term and long-term EHS hazard-related effects. Preference should be given to less-mobile solvents, which are clearly better with respect to plant safety. A possible drawback is that the process design of the capture systems with the phase-change solvents might promote the accumulation of carcinogenic nitrosamines, so their concentrations should be carefully monitored. In terms of life-cycle impact, the steam demand for solvent regeneration and the electricity required to deliver the cooling media are the main contributors to the impact. The use of renewable electricity and industrial waste heat (i.e., at 90°-100°C) can decrease the LCA impact of the phase-change capture plant by 70%-90%. The remaining impact will be dominated by the degradation behaviours of the solvent molecules, which means that experimental tests to evaluate the degradation behaviours of the solvents and measures to reduce the degradation rates are of the utmost importance.

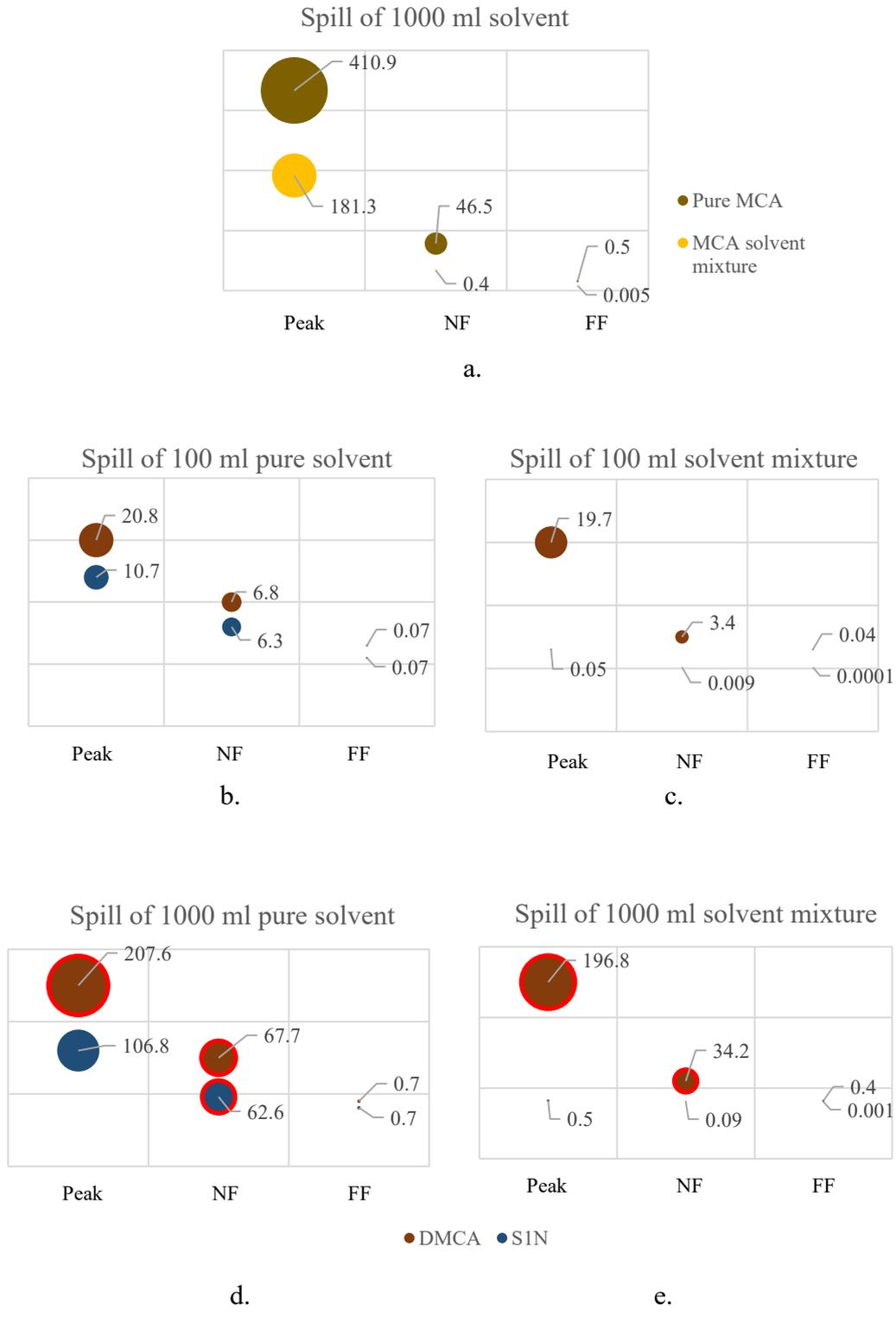


Figure 17. Air concentrations of MCA (a), and DMCA, and S1N (b-e) (in mg/m³) in the case of a spill of pure or mixed solvent. “Peak” indicates the concentration during the first seconds of the spill, NF is the concentration in the near- field zone (1 m from the center of the spill), and FF indicates the concentration in the far- field zone (10 m from the center of the spill). The circles marked with a red outline indicate values that exceed the safe limits. Source: Paper III.

5 Conclusions and outlook

The overall aim of the thesis was to evaluate the environmental, health, and safety impacts of novel chemicals during early process design. The role of predictive modeling and streamlined techniques in the early impact assessment of chemical structures were also examined. This chapter summarizes the main conclusions drawn in response to the research questions posed, and provides some suggestions for future research.

5.1 Main conclusions

The **first research question** addressed the impact assessment of new molecular structures during molecular design to identify safer and more environmentally benign chemicals. Incorporating the impact criteria into the computer-aided molecular design (CAMD) of novel phase-change solvents for CO₂ capture (simultaneous approach) led to the identification of a more-comprehensive set of candidate structures. The additional design criteria favored the OH- containing molecules, which are preferable from the sustainability point-of-view, although they are less-well-investigated in terms of their phase-change behaviors. A comparison was also made with the decomposed approach when the impact assessment was performed as a post-assessment of the CAMD-derived structural alternatives (without impacts criteria). Even though, during the decomposed approach the search was focused on CO₂ capture-related properties that facilitate the identification of beneficial phase-change solvents, the simultaneous approach resulted in the identification of several structures similar to those discovered using the decomposed technique. In addition, new alternatives were proposed. Thus, the simultaneous application of the CAMD and integrated impact assessment is advantageous for the identification of a vast number of molecular candidates, including potential alternatives that warrant further research.

The **second research question** was concerned with the problem of the data gaps that exist for novel or existing molecular structures with missing information on their properties. Predictive modeling can be applied to estimate the missing data when the empirical or commonly used prediction models, such as GC or QSARs, are unavailable. It is shown that despite the approximations introduced by predictive models that estimate some of the parameters required to perform the impact assessment, the ability of the CAMD to identify the top solvents is not affected. Furthermore, the predictive models can be advanced using domain knowledge. It is demonstrated that prior knowledge can be incorporated in various ways before, after, and during the property estimation. This work shows that prior knowledge improves the performance of the predictive models for estimating the acute aquatic toxicity value in terms of either the prediction accuracy or compliance with previously observed trends derived from scientific publications.

The extraction of knowledge from scientific articles was the topic of the **third research question**. It is shown here that the semi-automated procedure of knowledge extraction has the potential to reduce significantly the manual work required to process a

high number of scientific papers. The extracted knowledge in the field of acute aquatic toxicity appears to be diverse and extensive. Development of the knowledge classification and update schemes was found to be useful for systematic knowledge collection, evaluation, and determination of possible usage. The proposed knowledge extraction procedure can assist interdisciplinary research when rapid acquisition of knowledge is required. The knowledge extraction method and approaches for knowledge incorporation into predictive models are generic and can be used in many other knowledge domains. The knowledge extraction method can easily incorporate more resources (in terms of amount and type), while the classified knowledge allows for more hybrid alternatives, also depending on the machine learning approach used (i.e., neural networks and deep learning approaches, classification trees, random forest regression, etc.).

The **fourth and the final research** question examined the potential impact and challenges associated with the employment of the newly designed molecules. The lack of information regarding the behaviors of the molecules represents the main challenge during the impact evaluation of the process with the novel compounds. Experimental data obtained during pilot plant testing often provide information for shortcut models to estimate only the main process parameters, such as the energy requirement for the desorption process. The behaviors of the chemicals that influence the safety, health, and environmental performance have to be estimated using predictive modeling techniques and scenario analyses. The evaluation of the potential impacts of the carbon capture systems with phase-change solvents, considering some aspects of early process design and streamlined assessment uncertainties, showed that the choice and the behavior of the solvent, which are strongly influenced by the system's operating conditions, can have a significant effect on the potential impact of the system.

The predictive modeling is proven to be useful for closing the data gaps that are increasingly common in early impact assessments of chemicals. Together with streamlined techniques, the tool guides the selection of safer chemical alternatives and process designs. Ultimately, the early assessment techniques, assisted by advances in machine learning, play an important role in transitioning to the use of chemicals that are sustainable by design.

5.2 Future work

The results of early impact assessments that apply predictive and streamlined techniques are intrinsically uncertain. The level of uncertainty could be reduced by targeted experimental work, method refinement, and increasing the level of detail. Experimental work that focuses on the behaviors of phase-change solvent molecules in terms of degradation and emergence of hazardous substances during the long-term operation of the carbon capture systems would validate the results of the process level assessment.

Different types of predictive models that apply various algorithms and modes of knowledge integration can be explored. The models developed in this work were

primarily focused on acute aquatic toxicity and a limited range of chemical species. It would be interesting to study the knowledge extraction and utilization applied to other chemical classes and domains.

The procedure for knowledge extraction could be enhanced by improving the text-mining approach to reduce even further the need for the manual processing. An additional pretraining step might be introduced to generate the text-mining models that are directed towards extracting relevant pieces from the articles pertaining to the specific domain. The method can also be extended to allow the processing of the images and tables found in the papers.

Further research is required to explore an iterative approach to optimal integration between the streamlined techniques and experimentation when the streamlined methods highlight specific domains and where more experimentation is needed, and thereafter, the new data are used to update and improve the streamlined techniques.

References

- (1) European Commission. A European Green Deal https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en#highlights (accessed Dec 11, 2021).
- (2) European Commission. Chemicals strategy https://ec.europa.eu/growth/sectors/chemicals/chemicals-strategy_en (accessed Dec 11, 2021).
- (3) de Marco, B. A.; Rechelo, B. S.; Tótolí, E. G.; Kogawa, A. C.; Salgado, H. R. N. Evolution of Green Chemistry and Its Multidimensional Impacts: A Review. *Saudi Pharm. J.* **2019**, *27* (1), 1–8.
- (4) Crawford, S. E.; Hartung, T.; Hollert, H.; Mathes, B.; van Ravenzwaay, B.; Steger-Hartmann, T.; Studer, C.; Krug, H. F. Green Toxicology: A Strategy for Sustainable Chemical and Material Development. *Environ. Sci. Eur.* **2017**, *29* (1).
- (5) Judson, R.; Richard, A.; Dix, D. J.; Houck, K.; Martin, M.; Kavlock, R.; Dellarco, V.; Henry, T.; Holderman, T.; Sayre, P.; Tan, S.; Carpenter, T.; Smith, E. The Toxicity Data Landscape for Environmental Chemicals. *Environ. Health Perspect.* **2009**, *117* (5), 685–695.
- (6) National Research Council. *A Framework to Guide Selection of Chemical Alternatives*; National Academies Press: Washington, D.C., 2014; available at <https://www.nap.edu/catalog/18872/a-framework-to-guide-selection-of-chemical-alternatives>.
- (7) Sandén, B. A. Technology Path Assessment for Sustainable Technology Development. *Innovation* **2004**, *6* (2), 316–330.
- (8) Arvidsson, R.; Tillman, A.; Sandén, B. A.; Janssen, M.; Nordelöf, A.; Kushnir, D.; Molander, S. Environmental Assessment of Emerging Technologies: Recommendations for Prospective LCA. *J. Ind. Ecol.* **2017**, *22* (6), 1286–1294.
- (9) van der Giesen, C.; Cucurachi, S.; Guinée, J.; Kramer, G. J.; Tukker, A. A Critical View on the Current Application of LCA for New Technologies and Recommendations for Improved Practice. *J. Clean. Prod.* **2020**, *259*, 120904.
- (10) Buyle, M.; Audenaert, A.; Billen, P.; Boonen, K.; Van Passel, S. The Future of Ex-Ante LCA? Lessons Learned and Practical Recommendations. *Sustain.* **2019**, *11* (19), 1–24.
- (11) Hung, C. R.; Ellingsen, L. A.; Majeau-Bettez, G. LiSET: A Framework for Early-Stage Life Cycle Screening of Emerging Technologies. *J. Ind. Ecol.* **2020**, *24* (1), 26–37.
- (12) Cramer, J. Early Warning: Integrating Eco-Efficiency Aspects into the Product Development Process. *Environ. Qual. Manag.* **2000**, *10* (2), 1–10.
- (13) Basbagill, J.; Flager, F.; Lepech, M.; Fischer, M. Application of Life-Cycle Assessment to Early Stage Building Design for Reduced Embodied Environmental Impacts. *Build. Environ.* **2013**, *60*, 81–92.
- (14) Harris, S.; Tsalidis, G.; Corbera, J. B.; Espi Gallart, J. J.; Tegstedt, F. Application of LCA and LCC in the Early Stages of Wastewater Treatment Design: A Multiple Case Study of Brine Effluents. *J. Clean. Prod.* **2021**, *307*, 127298.
- (15) Caduff, M.; Huijbregts, M. A. J.; Koehler, A.; Althaus, H.-J.; Hellweg, S. Scaling Relationships in Life Cycle Assessment. *J. Ind. Ecol.* **2014**, *18* (3), 393–406.
- (16) Broeren, M. L. M.; Molenveld, K.; van den Oever, M. J. A.; Patel, M. K.; Worrell, E.; Shen, L. Early-Stage Sustainability Assessment to Assist with Material Selection: A Case Study for Biobased Printer Panels. *J. Clean. Prod.* **2016**, *135*, 30–41.

- (17) Moni, S. M.; Mahmud, R.; High, K.; Carbajales-Dale, M. Life Cycle Assessment of Emerging Technologies: A Review. *J. Ind. Ecol.* **2020**, *24* (1), 52–63.
- (18) Azapagic, A.; Perdan, S. Indicators of Sustainable Development for Industry. *Process Saf. Environ. Prot.* **2000**, *78* (4), 243–261.
- (19) Gonzalez, M. A.; Smith, R. L. A Methodology to Evaluate Process Sustainability. *Environ. Prog.* **2003**, *22* (4), 269–276.
- (20) Guillen-Gosalbez, G.; Caballero, J. #x ;. A.; Esteller, L. J.; Gadalla, M. Application of Life Cycle Assessment to the Structural Optimization of Process Flowsheets. *Comput. Aided Chem. Eng.* **2007**, *24*, 1163–1168.
- (21) Dunj3, J.; Fthenakis, V.; V3lchez, J. A.; Arnaldos, J. Hazard and Operability (HAZOP) Analysis. A Literature Review. *J. Hazard. Mater.* **2010**, *173* (1–3), 19–32.
- (22) Ruijters, E.; Stoelinga, M. Fault Tree Analysis: A Survey of the State-of-the-Art in Modeling, Analysis and Tools. *Comput. Sci. Rev.* **2015**, *15–16*, 29–62.
- (23) Badr, S. Sustainability Assessment of Amine-Based Post Combustion CO2 Capture, PhD thesis, ETH Zurich, Switzerland, 2016.
- (24) American Institute of Chemical Engineers (AIChE). *Dow’s Chemical Exposure Index Guide*, 1st ed.; 1998.
- (25) Khan, F. I.; Husain, T.; Abbasi, S. A. Safety Weighted Hazard Index (SWeHI). *Process Saf. Environ. Prot.* **2001**, *79* (2), 65–80.
- (26) Imperial Chemical Industries Limited., M. D. *The Mond Index : How to Identify Assess and Minimise Potential Hazards on Chemical Plant Units for New and Existing Processes*; Imperial Chemical Industries PLC, Explosion Hazards Section, Technical Dept.: Winnington, Cheshire, 1985.
- (27) Khan, F. I.; Amyotte, P. R. Integrated Inherent Safety Index (I2SI): A Tool for Inherent Safety Evaluation. *Process Saf. Prog.* **2004**, *23* (2), 136–148.
- (28) Kletz, T. A. Preventing Catastrophic Accidents. *Chem. Eng.* **1976**, *83* ((8)), 124–128.
- (29) Edwards, D. W.; Lawrence, D. Assessing the Inherent Safety of Chemical Process Routes: Is There a Relation between Plant Costs and Inherent Safety? In *Transactions of the Institution of Chemical Engineers, part B*; 1993; pp 252–258.
- (30) Heikkilä, A. *Inherent Safety in Process Plant Design. An Index-Based Approach*, 384th ed.; VTT Publications: Espoo, 1999; available at, <http://www.inf.vtt.fi/pdf/>.
- (31) Sugiyama, H.; Fischer, U.; Hungerbuhler, K.; Hirao, M. Decision Framework for Chemical Process Design Including Different Stages Environmental, Health, and Safety Assessment. *AIChE J.* **2008**, *54* (4), 1037–1053.
- (32) Banimostafa, A.; Nguyen, T. T. H.; Kikuchi, Y.; Papadokonstantakis, S.; Sugiyama, H.; Hirao, M.; Hungerbühler, K. Safety, Health, and Environmental Assessment of Bioethanol Production from Sugarcane, Corn, and Corn Stover. *Green Process. Synth.* **2012**, *1* (5).
- (33) Patel, A. D.; Meesters, K.; Den Uil, H.; De Jong, E.; Blok, K.; Patel, M. K. Sustainability Assessment of Novel Chemical Processes at Early Stage: Application to Biobased Processes. *Energy Environ. Sci.* **2012**, *5* (9), 8430–8444.
- (34) Olivetti, E.; Patanavanich, S.; Kirchain, R. Exploring the Viability of Probabilistic Under-Specification To Streamline Life Cycle Assessment. *Environ. Sci. Technol.* **2013**, *47* (10), 5208–5216.
- (35) Böckin, D. Towards Empirically Grounded Guidance for Resource Efficiency : Applying , Developing and Synthesising Environmental Assessments, PhD thesis, Chalmers University of Technology, 2021.
- (36) Thonemann, N.; Schulte, A.; Maga, D. How to Conduct Prospective Life Cycle

- Assessment for Emerging Technologies? A Systematic Review and Methodological Guidance. *Sustainability* **2020**, *12* (3), 1192.
- (37) Villares, M.; Işıldar, A.; van der Giesen, C.; Guinée, J. Does Ex Ante Application Enhance the Usefulness of LCA? A Case Study on an Emerging Technology for Metal Recovery from e-Waste. *Int. J. Life Cycle Assess.* **2017**, *22* (10), 1618–1633.
- (38) Karka, P.; Papadokostantakis, S.; Kokossis, A. Environmental Impact Assessment of Biomass Process Chains at Early Design Stages Using Decision Trees. *Int. J. Life Cycle Assess.* **2019**, *24* (9), 1675–1700.
- (39) Cumming, J. G.; Davis, A. M.; Muresan, S.; Haeberlein, M.; Chen, H. Chemical Predictive Modelling to Improve Compound Quality. *Nat. Rev. Drug Discov.* **2013**, *12* (12), 948–962.
- (40) Kolska, Z.; Zabransky, M.; Randova, A. Group Contribution Methods for Estimation of Selected Physico-Chemical Properties of Organic Compounds. In *Thermodynamics - Fundamentals and Its Application in Science*; Morales-Rodriguez, R., Ed.; InTech, 2012; Vol. i, p 13.
- (41) Marchant, C. A. Prediction of Rodent Carcinogenicity Using the DEREK System for 30 Chemicals Currently Being Tested by the National Toxicology Program. The DEREK Collaborative Group. *Environ. Health Perspect.* **1996**, *104* (suppl 5), 1065–1073.
- (42) Gevarter, W. B. *AN OVERVIEW OF EXPERT SYSTEMS*; U.S. Department of Commerce: Washington, D.C., 1982.
- (43) Escher, S. E.; Kamp, H.; Bennekou, S. H.; Bitsch, A.; Fisher, C.; Graepel, R.; Hengstler, J. G.; Herzler, M.; Knight, D.; Leist, M.; Norinder, U.; Ouédraogo, G.; Pastor, M.; Stuard, S.; White, A.; Zdrzil, B.; van de Water, B.; Kroese, D. Towards Grouping Concepts Based on New Approach Methodologies in Chemical Hazard Assessment: The Read-across Approach of the EU-ToxRisk Project. *Arch. Toxicol.* **2019**, *93* (12), 3643–3667.
- (44) Gramatica, P. A Short History Of QSAR. *QSAR Res. Unit Environ. Chem. Ecotoxicol.* **2011**.
- (45) Wernet, G.; Hellweg, S.; Fischer, U.; Papadokostantakis, S.; Hungerbühler, K. Molecular-Structure-Based Models of Chemical Inventories Using Neural Networks. *Environ. Sci. Technol.* **2008**, *42* (17), 6717–6722.
- (46) Wernet, G.; Papadokostantakis, S.; Hellweg, S.; Hungerbühler, K. Bridging Data Gaps in Environmental Assessments: Modeling Impacts of Fine and Basic Chemical Production. *Green Chem.* **2009**, *11* (11), 1826.
- (47) Song, R.; Keller, A. A.; Suh, S. Rapid Life-Cycle Impact Screening Using Artificial Neural Networks. *Environ. Sci. Technol.* **2017**, *51* (18), 10777–10785.
- (48) Calvo-Serrano, R.; González-Miquel, M.; Papadokostantakis, S.; Guillén-Gosálbez, G. Predicting the Cradle-to-Gate Environmental Impact of Chemicals from Molecular Descriptors and Thermodynamic Properties via Mixed-Integer Programming. *Comput. Chem. Eng.* **2018**, *108*, 179–193.
- (49) Zhuang, Q.; Clements, B.; Dai, J.; Carrigan, L. Ten Years of Research on Phase Separation Absorbents for Carbon Capture: Achievements and next Steps. *Int. J. Greenh. Gas Control* **2016**, *52*, 449–460.
- (50) Quan, Z.; Bruce, C.; Li, B. Emerging New Types of Absorbents for Postcombustion Carbon Capture. In *Recent Advances in Carbon Capture and Storage*; Yun, Y., Ed.; IntechOpen, 2017; Vol. i, pp 91–116.
- (51) Zhang, S.; Shen, Y.; Wang, L.; Chen, J.; Lu, Y. Phase Change Solvents for Post-Combustion CO₂ Capture: Principle, Advances, and Challenges. *Appl. Energy* **2019**, *239* (January), 876–897.

- (52) Jessop, P. G.; Mercer, S. M.; Heldebrant, D. J. CO₂-Triggered Switchable Solvents, Surfactants, and Other Materials. *Energy Environ. Sci.* **2012**, *5* (6), 7240–7253.
- (53) Papadopoulos, A. I.; Tzirakis, F.; Tsvintzelis, I.; Seferlis, P. Phase-Change Solvents and Processes for Postcombustion CO₂ Capture: A Detailed Review. *Ind. Eng. Chem. Res.* **2019**, *58* (13), 5088–5111.
- (54) Heldebrant, D. J.; Koech, P. K.; Ang, M. T. C.; Liang, C.; Rainbolt, J. E.; Yonker, C. R.; Jessop, P. G. Reversible Zwitterionic Liquids, the Reaction of Alkanol Guanidines, Alkanol Amidines, and Diamines with CO₂. *Green Chem.* **2010**, *12* (4), 713–772.
- (55) Shavaliyeva, G.; Papadokonstantakis, S.; Kazepidis, P.; Papadopoulos, A. I.; Seferlis, P. Sustainability Analysis of Phase-Change Solvents for Post-Combustion CO₂ Capture. *Chem. Eng. Trans.* **2019**, *76* (2016), 1045–1050.
- (56) Chang, S. S. L.; Kong, Y. L.; Lim, W. X.; Ooi, J.; Ng, D. K. S.; Chemmangattuvalappil, N. G. Design of Alternate Solvent for Recovery of Residual Palm Oil: Simultaneous Optimization of Process Performance with Environmental, Health and Safety Aspects. *Clean Technol. Environ. Policy* **2018**, *20* (5), 949–968.
- (57) Neoh, J. Q.; Chin, H. H.; Mah, A. X. Y.; Aboagwa, O. A.; Thangalazhy-Gopakumar, S.; Chemmangattuvalappil, N. G. Design of Bio-Oil Additives Using Mathematical Optimisation Tools Considering Blend Functionality and Sustainability Aspects. *Sustain. Prod. Consum.* **2019**, *19*, 53–63.
- (58) Ooi, J.; Promentilla, M. A. B.; Tan, R. R.; Ng, D. K. S.; Chemmangattuvalappil, N. G. A Systematic Methodology for Multi-Objective Molecular Design via Analytic Hierarchy Process. *Process Saf. Environ. Prot.* **2017**, *111*, 663–677.
- (59) Vanderveen, J. R.; Patiny, L.; Chalifoux, C. B.; Jessop, M. J.; Jessop, P. G. A Virtual Screening Approach to Identifying the Greenest Compound for a Task: Application to Switchable-Hydrophilicity Solvents. *Green Chem.* **2015**, *17* (12), 5182–5188.
- (60) Mehrkesh, A.; Karunanithi, A. T. *New Perspective on Computer Aided Molecular Design: A Life Cycle Assessment Approach*; Elsevier, 2014; Vol. 34, available at <http://dx.doi.org/10.1016/B978-0-444-63433-7.50046-8>.
- (61) Heintz, J.; Belaud, J. P.; Pandya, N.; Teles Dos Santos, M.; Gerbaud, V. Computer Aided Product Design Tool for Sustainable Product Development. *Comput. Chem. Eng.* **2014**, *71*, 362–376.
- (62) Weis, D. C.; Visco, D. P. Computer-Aided Molecular Design Using the Signature Molecular Descriptor: Application to Solvent Selection. *Comput. Chem. Eng.* **2010**, *34* (7), 1018–1029.
- (63) Karl, M.; Wright, R. F.; Berglen, T. F.; Denby, B. Worst Case Scenario Study to Assess the Environmental Impact of Amine Emissions from a CO₂ Capture Plant. *Int. J. Greenh. Gas Control* **2011**, *5* (3), 439–447.
- (64) Veltman, K.; Singh, B.; Hertwich, E. G. Human and Environmental Impact Assessment of Postcombustion CO₂ Capture Focusing on Emissions from Amine-Based Scrubbing Solvents to Air. *Environ. Sci. Technol.* **2010**, *44* (4), 1496–1502.
- (65) Zhang, L.; Zhang, H.; Ai, H.; Hu, H.; Li, S.; Zhao, J.; Liu, H. Applications of Machine Learning Methods in Drug Toxicity Prediction. *Curr. Top. Med. Chem.* **2018**, *18* (12), 987–997.
- (66) Hartung, T. Predicting Toxicity of Chemicals: Software Beats Animal Testing. *EFSA J.* **2019**, *17* (S1), 1–8.
- (67) European Commission. Environment – LIFE Programme. ANTARES | Alternative

- Non-Testing methods Assessed for REACH Substances <http://www.antaes-life.eu/index.php> (accessed Mar 14, 2018).
- (68) Craft, D.; Ferranti, D.; Krane, D. The Value of Prior Knowledge in Machine Learning of Complex Network Systems. *Bioinformatics* **2017**, *33* (22), 3610–3618.
- (69) Lysenko, A.; Sharma, A.; Boroevich, K. A.; Tsunoda, T. An Integrative Machine Learning Approach for Prediction of Toxicity-Related Drug Safety. *Life Sci. Alliance* **2018**, *1* (6), 1–14.
- (70) Chakraborty, S.; Chattopadhyay, P. P.; Ghosh, S. K.; Datta, S. Incorporation of Prior Knowledge in Neural Network Model for Continuous Cooling of Steel Using Genetic Algorithm. *Appl. Soft Comput. J.* **2017**, *58*, 297–306.
- (71) Bougie, N.; Cheng, L. K.; Ichise, R. Combining Deep Reinforcement Learning with Prior Knowledge and Reasoning. *Appl. Comput. Rev.* **2018**, *18* (2), 33–45.
- (72) Heidi K. Stranddorf; Leif Hoffmann; Anders Schmidt; Impact Categories, Normalisation and Weighting in LCA. *Environ. News* **2005**, *78*, 90.
- (73) Goedkoop, M.; Spriensma, R. The Eco-Indicator 99 - A Damage Oriented Method for Life Cycle Impact Assessment. *Assessment* **2001**, 144.
- (74) Koller, G.; Fischer, U.; Hungerbühler, K. Assessing Safety, Health, and Environmental Impact Early during Process Development. *Ind. Eng. Chem. Res.* **2000**, *39* (4), 960–972.
- (75) Papadopoulos, A. I.; Shavaliyeva, G.; Papadokonstantakis, S.; Seferlis, P.; Perdomo, F. A.; Galindo, A.; Jackson, G.; Adjiman, C. S. An Approach for Simultaneous Computer-Aided Molecular Design with Holistic Sustainability Assessment: Application to Phase-Change CO₂ Capture Solvents. *Comput. Chem. Eng.* **2020**, *135*.
- (76) Hukkerikar, A. S.; Sarup, B.; Ten Kate, A.; Abildskov, J.; Sin, G.; Gani, R. Group-Contribution + (GC +) Based Estimation of Properties of Pure Components: Improved Property Estimation and Uncertainty Analysis. *Fluid Phase Equilib.* **2012**, *321*, 25–43.
- (77) Hukkerikar, A. S.; Kalakul, S.; Sarup, B.; Young, D. M.; Sin, G.; Gani, R. Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution+ (GC +) Property Models and Uncertainty Analysis. *J. Chem. Inf. Model.* **2012**, *52* (11), 2823–2839.
- (78) Lu, J.; Peng, J.; Wang, J.; Shen, Q.; Bi, Y.; Gong, L.; Zheng, M.; Luo, X.; Zhu, W.; Jiang, H.; Chen, K. Estimation of Acute Oral Toxicity in Rat Using Local Lazy Learning. *J. Cheminform.* **2014**, *6* (1), 26.
- (79) Shavaliyeva, G.; Papadopoulos, A. I.; Badr, S.; Seferlis, P.; Papadokonstantakis, S. Sustainability Assessment Using Local Lazy Learning: The Case of Post-Combustion CO₂ Capture Solvents. In *13th International Symposium on Process Systems Engineering (PSE 2018)*; 2018; pp 823–828.
- (80) Köhler, A.; Hellweg, S.; Recan, E.; Hungerbühler, K.; Ko, A. Input-Dependent Life-Cycle Inventory Model of Industrial Wastewater-Treatment Processes in the Chemical Sector Input-Dependent Life-Cycle Inventory Model of Industrial Wastewater-Treatment Processes in the Chemical Sector. *Environ. Sci. Technol.* **2007**, No. 41, 5515–5522.
- (81) Ecoinvent 3.4 <https://www.ecoinvent.org/database/database.html>.
- (82) Goedkoop, M.; Heijungs, R.; De Schryver, A.; Struijs, J.; van Zelm, R. *ReCiPe 2008. A LCIA Method Which Comprises Harmonised Category Indicators at the Midpoint and the Endpoint Level. Report I: Characterisation.*; 2009; available at, http://www.pre-sustainability.com/download/misc/ReCiPe_main_report_final_27-02-

- 2009_web.pdf.
- (83) Dieterich, P. Short-Cut Models for Environmental Impact Metrics of Chemical Production Master 's Thesis, ETH Zurich, Switzerland, 2017.
- (84) Eissen, M.; Weiß, M.; Brinkmann, T.; Steinigeweg, S. Supporting Information Related to Comparison of Two Alternative Routes to an Enantiomerically Pure SS-Amino Acid. *Chem. Eng. Technol.* **2010**, 629–637.
- (85) JSON. Recommendation of Occupational Exposure Limits (2017-2018). *J. Occup. Health* **2017**, No. 59, 436–469.
- (86) Acute Toxicity. In *International Organization*; 1955; Vol. 9, pp 275–277.
- (87) IFA Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung. Foreign and EU limit values <https://www.dguv.de/ifa/fachinfos/occupational-exposure-limit-values/foreign-and-eu-limit-values/index.jsp> (accessed Mar 28, 2018).
- (88) ECHA (European Chemicals Agency). Information on Chemicals <https://echa.europa.eu/information-on-chemicals> (accessed Mar 28, 2018).
- (89) Fine, N. A. Nitrosamine Management in Aqueous Amines for Post-Combustion Carbon Capture, PhD thesis, The University of Texas at Austin, USA, 2015.
- (90) Nicas, M. The near Field / Far Field Model with Constant Application of Chemical Mass and Exponentially Decreasing Emission of the Mass Applied. *J. Occup. Environ. Hyg.* **2016**, 13 (7), 519–528.
- (91) Keil, C. B.; Nicas, M. Predicting Room Vapor Concentrations Due to Spills of Organic Solvents. *Am. Ind. Hyg. Assoc. J.* **2003**, 64 (4), 445–454.
- (92) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminform.* **2015**, 7 (1), 1–13.
- (93) Craw, S. *Encyclopedia of Machine Learning*; Sammut, C., Webb, G. I., Eds.; Springer US: Boston, MA, 2011; available at, <http://link.springer.com/10.1007/978-0-387-30164-8>.
- (94) Landrum, G. *RDKit Documentation. Release 2017.03.1*; 2017; available at, http://www.rdkit.org/RDKit_Docs.current.pdf.
- (95) Lu, J.; Zhang, P.; Zou, X. W.; Zhao, X. Q.; Cheng, K. G.; Zhao, Y. L.; Bi, Y.; Zheng, M. Y.; Luo, X. M. In Silico Prediction of Chemical Toxicity Profile Using Local Lazy Learning. *Comb. Chem. High Throughput Screen.* **2017**, 20 (4), 346–353.
- (96) Stempel, S. PBT Assessment under REACH by Means of Chemoinformatics. **2012**, PhD thesis, ETH Zurich, Switzerland.
- (97) Melnikov, F.; Kostal, J.; Voutchkova-Kostal, A.; Zimmerman, J. B.; Anastas, P. T. Assessment of Predictive Models for Estimating the Acute Aquatic Toxicity of Organic Chemicals. *Green Chem.* **2016**, 18 (16), 4432–4445.
- (98) Jonnalagadda, S. R.; Goyal, P.; Huffman, M. D. Automating Data Extraction in Systematic Reviews: A Systematic Review. *Syst. Rev.* **2015**, 4 (1).
- (99) Talib, R.; Hanif, M. K.; Ayesha, S.; Fatima, F. Text Mining: Techniques , Applications and Issues. In *International Journal of Advanced Computer Science and Applications*; 2016; Vol. 7, pp 414–418.
- (100) Amini, A.; Muggleton, S. H.; Lodhi, H.; Sternberg, M. J. E. A Novel Logic-Based Approach for Quantitative Toxicology Prediction. *J. Chem. Inf. Model.* **2007**, 47 (3), 998–1006.
- (101) Cassotti, M.; Ballabio, D.; Consonni, V.; Mauri, A.; Tetko, I. V.; Todeschini, R. Prediction of Acute Aquatic Toxicity toward Daphnia Magna by Using the GA- k NN Method. *Altern. to Lab. Anim.* **2014**, 42 (1), 31–41.
- (102) Qin, W. C.; Su, L. M.; Zhang, X. J.; Qin, H. W.; Wen, Y.; Guo, Z.; Sun, F. T.;

- Sheng, L. X.; Zhao, Y. H.; Abraham, M. H. Toxicity of Organic Pollutants to Seven Aquatic Organisms: Effect of Polarity and Ionization. *SAR QSAR Environ. Res.* **2010**, *21* (5–6), 389–401.
- (103) Wang, Y.; Chen, X. A Joint Optimization QSAR Model of Fathead Minnow Acute Toxicity Based on a Radial Basis Function Neural Network and Its Consensus Modeling. *RSC Adv.* **2020**, *10* (36), 21292–21308.
- (104) Khan, K.; Baderna, D.; Cappelli, C.; Toma, C.; Lombardo, A.; Roy, K.; Benfenati, E. Ecotoxicological QSAR Modeling of Organic Compounds against Fish: Application of Fragment Based Descriptors in Feature Analysis. *Aquat. Toxicol.* **2019**, *212* (April), 162–174.
- (105) Katritzky, A. R.; Slavov, S. H.; Stoyanova-Slavova, I. S.; Kahn, I.; Karelson, M. Quantitative Structure-Activity Relationship (QSAR) Modeling of EC 50 of Aquatic Toxicities for *Daphnia Magna*. *J. Toxicol. Environ. Heal. - Part A Curr. Issues* **2009**, *72* (19), 1181–1190.
- (106) Liu, L.; Yang, H.; Cai, Y.; Cao, Q.; Sun, L.; Wang, Z.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. In Silico Prediction of Chemical Aquatic Toxicity for Marine Crustaceans via Machine Learning. *Toxicol. Res. (Camb)*. **2019**, *8* (3), 341–352.
- (107) Roy, K.; Das, R. N. QSTR with Extended Topochemical Atom (ETA) Indices. 15. Development of Predictive Models for Toxicity of Organic Chemicals against Fathead Minnow Using Second-Generation ETA Indices. *SAR QSAR Environ. Res.* **2012**, *23* (1–2), 125–140.
- (108) Khan, K.; Benfenati, E.; Roy, K. Consensus QSAR Modeling of Toxicity of Pharmaceuticals to Different Aquatic Organisms: Ranking and Prioritization of the DrugBank Database Compounds. *Ecotoxicol. Environ. Saf.* **2019**, *168* (October 2018), 287–297.
- (109) Nendza, M.; Müller, M. Discriminating Toxicant Classes by Mode of Action: 2. Physico-Chemical Descriptors. *Quant. Struct. Relationships* **2000**, *19* (6), 581–598.
- (110) Schultz, T. W.; Cronin, M. T. D.; Netzeva, T. I.; Aptula, A. O. Structure–Toxicity Relationships for Aliphatic Chemicals Evaluated with *Tetrahymena Pyriformis*. *Chem. Res. Toxicol.* **2002**, *15* (12), 1602–1609.
- (111) Toropova, A. P.; Toropov, A. A.; Veselinović, A. M.; Veselinović, J. B.; Leszczynska, D.; Leszczynski, J. Monte Carlo–Based Quantitative Structure–Activity Relationship Models for Toxicity of Organic Chemicals to *Daphnia Magna*. *Environ. Toxicol. Chem.* **2016**, *35* (11), 2691–2697.
- (112) Morrall, D. D.; Belanger, S. E.; Dunphy, J. C. Acute and Chronic Aquatic Toxicity Structure-Activity Relationships for Alcohol Ethoxylates. *Ecotoxicol. Environ. Saf.* **2003**, *56* (3), 381–389.
- (113) Liu, W.; Wang, X.; Zhou, X.; Duan, H.; Zhao, P.; Liu, W. Quantitative Structure-Activity Relationship between the Toxicity of Amine Surfactant and Its Molecular Structure. *Sci. Total Environ.* **2020**, *702*, 134593.
- (114) Mayer, P.; Reichenberg, F. Can Highly Hydrophobic Organic Substances Cause Aquatic Baseline Toxicity and Can They Contribute to Mixture Toxicity? *Environ. Toxicol. Chem.* **2006**, *25* (10), 2639–2644.
- (115) Clayden, J.; Donnard, M.; Lefranc, J.; Tetlow, D. J. Quaternary Centres Bearing Nitrogen (α -Tertiary Amines) as Products of Molecular Rearrangements. *Chem. Commun.* **2011**, *47* (16), 4624–4639.
- (116) Prusov, E. V. Construction of Quaternary Stereogenic Centers in the Total Synthesis of Natural Products. *Angew. Chemie - Int. Ed.* **2017**, *56* (46), 14356–14358.

- (117) Papadopoulos, A. I.; Perdomo, F. A.; Tzirakis, F.; Shavaliyeva, G.; Tsivintzelis, I.; Kazepidis, P.; Nessi, E.; Papadokonstantakis, S.; Seferlis, P.; Galindo, A.; Jackson, G.; Adjiman, C. S. Molecular Engineering of Sustainable Phase-Change Solvents: From Digital Design to Scaling-up for CO₂ Capture. *Chem. Eng. J.* **2021**, *420*, 127624.
- (118) Zhang, J. Study on CO₂ Capture Using Thermomorphic Biphasic Solvents with Energy-Efficient Regeneration., TU Dortmund, Germany, 2013.
- (119) Singh, P. Amine Based Solvent for CO₂ Absorption, PhD thesis, University of Twente, The Netherlands, 2017, Vol. 114.
- (120) Zhu, H.; Tropsha, A.; Fourches, D.; Varnek, A.; Papa, E.; Gramatica, P.; Öberg, T.; Dao, P.; Cherkasov, A.; Tetko, I. V. Combinatorial QSAR Modeling of Chemical Toxicants Tested against *Tetrahymena Pyriformis*. *J. Chem. Inf. Model.* **2008**, *48* (4), 766–784.
- (121) Bašić, I.; Lučić, B.; Nikolić, S.; Papeš-Šokčević, L.; Nadramija, D. Improvement of Ensemble of Multi-Regression Structure-Toxicity Models by Clustering of Molecules in Descriptor Space. *AIP Conf. Proc.* **2009**, *1148* 2 (August), 408–411.
- (122) Grigor'ev, V. Y.; Razdol'skii, A. N.; Zagrebina, A. O.; Tonkopii, V. D.; Raevskii, O. A. QSAR Classification Models of Acute Toxicity of Organic Compounds with Respect to *Daphnia Magna*. *Pharm. Chem. J.* **2014**, *48* (4), 242–245.
- (123) Cassotti, M.; Consonni, V.; Mauri, A.; Ballabio, D. Validation and Extension of a Similarity-Based Approach for Prediction of Acute Aquatic Toxicity towards *Daphnia Magna*. *SAR QSAR Environ. Res.* **2014**, *25* (12), 1013–1036.
- (124) Martin, T. M.; Young, D. M.; Lilavois, C. R.; Barron, M. G. Comparison of Global and Mode of Action-Based Models for Aquatic Toxicity. *SAR QSAR Environ. Res.* **2015**, *26* (3), 245–262.
- (125) Gini, G.; Craciun, M. V.; König, C.; Benfenati, E. Combining Unsupervised and Supervised Artificial Neural Networks to Predict Aquatic Toxicity. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 1897–1902.
- (126) He, L.; Xiao, K.; Zhou, C.; Li, G.; Yang, H.; Li, Z.; Cheng, J. Insights into Pesticide Toxicity against Aquatic Organism: QSTR Models on *Daphnia Magna*. *Ecotoxicol. Environ. Saf.* **2019**, *173* (October 2018), 285–292.
- (127) Gramatica, P.; Papa, E.; Sangion, A. QSAR Modeling of Cumulative Environmental End-Points for the Prioritization of Hazardous Chemicals. *Environ. Sci. Process. Impacts* **2018**, *20* (1), 38–47.
- (128) Hossain, K. A.; Roy, K. Chemometric Modeling of Aquatic Toxicity of Contaminants of Emerging Concern (CECs) in *Dugesia Japonica* and Its Interspecies Correlation with *Daphnia* and Fish: QSTR and QSTTR Approaches. *Ecotoxicol. Environ. Saf.* **2018**, *166* (August), 92–101.
- (129) Önlü, S.; Saçan, M. T. Toxicity of Contaminants of Emerging Concern to *Dugesia Japonica*: QSTR Modeling and Toxicity Relationship with *Daphnia Magna*. *J. Hazard. Mater.* **2018**, *351* (November 2017), 20–28.
- (130) Su, L.; Fu, L.; He, J.; Qin, W.; Sheng, L.; Abraham, M. H.; Zhao, Y. H. Comparison of *Tetrahymena Pyriformis* Toxicity Based on Hydrophobicity, Polarity, Ionization and Reactivity of Class-Based Compounds. *SAR QSAR Environ. Res.* **2012**, *23* (5–6), 537–552.
- (131) Ren, Y. Y.; Zhou, L. C.; Yang, L.; Liu, P. Y.; Zhao, B. W.; Liu, H. X. Predicting the Aquatic Toxicity Mode of Action Using Logistic Regression and Linear Discriminant Analysis. *SAR QSAR Environ. Res.* **2016**, *27* (9), 721–746.
- (132) Bouhedjar, K.; Benfenati, E.; Nacereddine, A. K. Modelling Quantitative Structure Activity–Activity Relationships (QSAARs): Auto-Pass-Pass, a New Approach to

- Fill Data Gaps in Environmental Risk Assessment under the REACH Regulation. *SAR QSAR Environ. Res.* **2020**, *31* (10), 785–801.
- (133) Dimitrov, S.; Koleva, Y.; Schultz, T. W.; Walker, J. D.; Mekenyan, O. Interspecies Quantitative Structure-Activity Relationship Model for Aldehydes: Aquatic Toxicity. *Environ. Toxicol. Chem.* **2004**, *23* (2), 463–470.
- (134) Kahn, I.; Maran, U.; Benfenati, E.; Netzeva, T. I.; Schultz, T. W.; Cronin, M. T. D. Comparative Quantitative Structure-Activity-Activity Relationships for Toxicity to *Tetrahymena Pyriformis* and *Pimephales Prometas*. *ATLA Altern. to Lab. Anim.* **2007**, *35* (1), 15–24.
- (135) Sánchez-Bayo, F. Comparative Acute Toxicity of Organic Pollutants and Reference Values for Crustaceans. I. Branchiopoda, Copepoda and Ostracoda. *Environ. Pollut.* **2006**, *139* (3), 385–420.
- (136) Lunghini, F.; Marcou, G.; Azam, P.; Enrici, M. H.; Van Miert, E.; Varnek, A. Consensus QSAR Models Estimating Acute Toxicity to Aquatic Organisms from Different Trophic Levels: Algae, *Daphnia* and Fish. *SAR QSAR Environ. Res.* **2020**, *31* (9), 655–675.
- (137) Khan, K.; Roy, K. Ecotoxicological QSAR Modelling of Organic Chemicals against *Pseudokirchneriella Subcapitata* Using Consensus Predictions Approach. *SAR QSAR Environ. Res.* **2019**, *30* (9), 665–681.
- (138) Huang, C. P.; Wang, Y. J.; Chen, C. Y. Toxicity and Quantitative Structure-Activity Relationships of Nitriles Based on *Pseudokirchneriella Subcapitata*. *Ecotoxicol. Environ. Saf.* **2007**, *67* (3), 439–446.
- (139) Tinkov, O. V.; Ognichenko, L. N.; Kuz'Min, V. E.; Gorb, L. G.; Kosinskaya, A. P.; Muratov, N. N.; Muratov, E. N.; Hill, F. C.; Leszczynski, J. Computational Assessment of Environmental Hazards of Nitroaromatic Compounds: Influence of the Type and Position of Aromatic Ring Substituents on Toxicity. *Struct. Chem.* **2016**, *27* (1), 191–198.
- (140) Galimberti, F.; Moretto, A.; Papa, E. Application of Chemometric Methods and QSAR Models to Support Pesticide Risk Assessment Starting from Ecotoxicological Datasets. *Water Res.* **2020**, *174*, 115583.
- (141) Khan, K.; Kar, S.; Sanderson, H.; Roy, K.; Leszczynski, J. Ecotoxicological Modeling, Ranking and Prioritization of Pharmaceuticals Using QSTR and i-QSTR Approaches: Application of 2D and Fragment Based Descriptors. *Mol. Inform.* **2019**, *38* (8–9).
- (142) Voutchkova, A. M.; Kostal, J.; Steinfeld, J. B.; Emerson, J. W.; Brooks, B. W.; Anastas, P.; Zimmerman, J. B. Towards Rational Molecular Design: Derivation of Property Guidelines for Reduced Acute Aquatic Toxicity. *Green Chem.* **2011**, *13* (9), 2373–2379.
- (143) Schultz, T. W.; Yarbrough, J. W.; Pilkington, T. B. Aquatic Toxicity and Abiotic Thiol Reactivity of Aliphatic Isothiocyanates: Effects of Alkyl-Size and -Shape. *Environ. Toxicol. Pharmacol.* **2007**, *23* (1), 10–17.
- (144) Roy, K.; Ghosh, G. QSTR with Extended Topochemical Atom Indices. 3. Toxicity of Nitrobenzenes To *Tetrahymena Pyriformis*. *QSAR Comb. Sci.* **2004**, *23* (23), 99–108.
- (145) Matveieva, M.; Cronin, M. T. D.; Polishchuk, P. Interpretation of QSAR Models: Mining Structural Patterns Taking into Account Molecular Context. *Mol. Inform.* **2019**, *38* (3).
- (146) Netzeva, T. I.; Pavan, M.; Worth, A. P. Review of (Quantitative) Structure-Activity Relationships for Acute Aquatic Toxicity. *QSAR Comb. Sci.* **2008**, *27* (1), 77–90.

- (147) Dobbins, L. L.; Usenko, S.; Brain, R. A.; Brooks, B. W. Probabilistic Ecological Hazard Assessment of Parabens Using *Daphnia Magna* and *Pimephales Promelas*. *Environ. Toxicol. Chem.* **2009**, *28* (12), 2744–2753.
- (148) Zhang, S.; Wang, N.; Su, L.; Xu, X.; Li, C.; Qin, W.; Zhao, Y. MOA-Based Linear and Nonlinear QSAR Models for Predicting the Toxicity of Organic Chemicals to *Vibrio Fischeri*. *Environ. Sci. Pollut. Res.* **2020**, *27* (9), 9114–9125.
- (149) DeWeese, A. D.; Schultz, T. W. Structure-Activity Relationships for Aquatic Toxicity To *Tetrahymena*: Halogen-Substituted Aliphatic Esters. *Environ. Toxicol.* **2001**, *16* (1), 54–60.
- (150) Stoyanova-Slavova, I. B.; Slavov, S. H.; Pearce, B.; Buzatu, D. A.; Beger, R. D.; Wilkes, J. G. Partial Least Square and K-Nearest Neighbor Algorithms for Improved 3D Quantitative Spectral Data-Activity Relationship Consensus Modeling of Acute Toxicity. *Environmental Toxicology and Chemistry*. 2014, pp 1271–1282.
- (151) Li, F.; Fan, D.; Wang, H.; Yang, H.; Li, W.; Tang, Y.; Liu, G. In Silico Prediction of Pesticide Aquatic Toxicity with Chemical Category Approaches. *Toxicol. Res. (Camb)*. **2017**, *6* (6), 831–842.
- (152) Martin, T. M.; Grulke, C. M.; Young, D. M.; Russom, C. L.; Wang, N. Y.; Jackson, C. R.; Barron, M. G. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. *J. Chem. Inf. Model.* **2013**, *53* (9), 2229–2239.
- (153) Aptula, A. O.; Roberts, D. W.; Cronin, M. T. D.; Schultz, T. W. Chemistry–Toxicity Relationships for the Effects of Di- and Trihydroxybenzenes to *Tetrahymena Pyriformis*. *Chem. Res. Toxicol.* **2005**, *18* (5), 844–854.
- (154) Bajot, F.; Cronin, M. T. D.; Roberts, D. W.; Schultz, T. W. Reactivity and Aquatic Toxicity of Aromatic Compounds Transformable to Quinone-Type Michael Acceptors. *SAR QSAR Environ. Res.* **2011**, *22* (1–2), 51–65.
- (155) Abe, T.; Saito, H.; Niikura, Y.; Shigeoka, T.; Nakano, Y. Embryonic Development Assay with *Daphnia Magna*: Application to Toxicity of Aniline Derivatives. *Chemosphere* **2001**, *45* (4–5), 487–495.
- (156) Ramos, E. U.; Vaal, M. A.; Hermens, J. L. M. Interspecies Sensitivity in the Aquatic Toxicity of Aromatic Amines. *Environ. Toxicol. Pharmacol.* **2002**, *11* (3–4), 149–158.
- (157) Roy, K.; Ghosh, G. QSTR with Extended Topochemical Atom Indices. 2. Fish Toxicity of Substituted Benzenes. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (2), 559–567.
- (158) Tugcu, G.; Ertürk, M. D.; Saçan, M. T. On the Aquatic Toxicity of Substituted Phenols to *Chlorella Vulgaris*: QSTR with an Extended Novel Data Set and Interspecies Models. *J. Hazard. Mater.* **2017**, *339*, 122–130.
- (159) Kienzler, A.; Barron, M. G.; Belanger, S. E.; Beasley, A.; Embry, M. R. Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. *Environ. Sci. Technol.* **2017**, *51* (17), 10203–10211.
- (160) Fogel, G. B.; Cheung, M. Derivation of Quantitative Structure-Toxicity Relationships for Ecotoxicological Effects of Organic Chemicals: Evolving Neural Networks and Evolving Rules. *2005 IEEE Congr. Evol. Comput. IEEE CEC 2005. Proc.* **2005**, *1*, 274–281.
- (161) Nendza, M.; Müller, M.; Wenzel, A. Classification of Baseline Toxicants for QSAR Predictions to Replace Fish Acute Toxicity Studies. *Environ. Sci. Process. Impacts* **2017**, *19* (3), 429–437.
- (162) Colombo, A.; Benfenati, E.; Karelson, M.; Maran, U. The Proposal of Architecture

- for Chemical Splitting to Optimize QSAR Models for Aquatic Toxicity. *Chemosphere* **2008**, 72 (5), 772–780.
- (163) Papa, E.; Villa, F.; Gramatica, P. Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in Pimephales p Romelas (Fathead Minnow) . *J. Chem. Inf. Model.* **2005**, 45 (5), 1256–1266.
- (164) Valavanidis, A.; Vlahogianni, T.; Dassenakis, M.; Scoullou, M. Molecular Biomarkers of Oxidative Stress in Aquatic Organisms in Relation to Toxic Environmental Pollutants. *Ecotoxicol. Environ. Saf.* **2006**, 64 (2), 178–189.
- (165) Chen, C. Y.; Kuo, K. L.; Fan, J. W. Toxicity of Propargylic Alcohols on Green Alga - Pseudokirchneriella Subcapitata. *J. Environ. Monit.* **2012**, 14 (1), 181–186.
- (166) Payne, M. P.; Button, W. G. Prediction of Acute Aquatic Toxicity in Tetrahymena Pyriformis - “Eco-Derek”, a Knowledge-Based System Approach. *SAR QSAR Environ. Res.* **2013**, 24 (6), 439–460.
- (167) Pavan, M.; Netzeva, T. I.; Worth, A. P. Validation of a QSAR Model for Acute Toxicity. *SAR QSAR Environ. Res.* **2006**, 17 (2), 147–171.
- (168) Eroglu, E.; Palaz, S.; Oltulu, O.; Turkmen, H.; Ozaydin, C. Comparative QSTR Study Using Semi-Empirical and First Principle Methods Based Descriptors for Acute Toxicity of Diverse Organic Compounds to the Fathead Minnow. *Int. J. Mol. Sci.* **2007**, 8 (12), 1265–1283.
- (169) Aalizadeh, R.; von der Ohe, P. C.; Thomaidis, N. S. Prediction of Acute Toxicity of Emerging Contaminants on the Water Flea Daphnia Magna by Ant Colony Optimization–Support Vector Machine QSTR Models. *Environ. Sci. Process. Impacts* **2017**, 19 (3), 438–448.
- (170) Jia, Q.; Zhao, Y.; Yan, F.; Wang, Q. QSAR Model for Predicting the Toxicity of Organic Compounds to Fathead Minnow. *Environ. Sci. Pollut. Res.* **2018**, 25 (35), 35420–35428.
- (171) Faucon, J. C.; Bureau, R.; Faisant, J.; Briens, F.; Rault, S. Prediction of the Daphnia Acute Toxicity from Heterogeneous Data. *Chemosphere* **2001**, 44 (3), 407–422.
- (172) Tao, S.; Xi, X.; Xu, F.; Li, B.; Cao, J.; Dawson, R. A Fragment Constant QSAR Model for Evaluating the EC50 Values of Organic Chemicals to Daphnia Magna. *Environ. Pollut.* **2002**, 116 (1), 57–64.
- (173) Stanton, D. T.; Dimitrov, S.; Grancharov, V.; Mekenyan, O. G. Charged Partial Surface Area (CPSA) Descriptors QSAR Applications. *SAR QSAR Environ. Res.* **2002**, 13 (2), 341–351.
- (174) Devillers, J. A New Strategy for Using Supervised Artificial Neural Networks in QSAR. *SAR QSAR Environ. Res.* **2005**, 16 (5), 433–442.
- (175) Bowen, K. R.; Flanagan, K. B.; Acree, W. E.; Abraham, M. H. Correlating Toxicities of Organic Compounds to Select Protozoa Using the Abraham Model. *Sci. Total Environ.* **2006**, 369 (1–3), 109–118.
- (176) Wu, X.; Zhang, Q.; Hu, J. QSAR Study of the Acute Toxicity to Fathead Minnow Based on a Large Dataset. *SAR QSAR Environ. Res.* **2016**, 27 (2), 147–164.
- (177) Kazepidis, P.; Papadopoulos, A. I.; Seferlis, P.; Perdomo, F. A.; Galindo, A.; Jackson, G.; Adjiman, C. S. Optimal Design of Post Combustion CO₂ Capture Processes Based on Phase-Change Solvents. *Comput. Aided Chem. Eng.* **2019**, 46, 463–468.
- (178) Kazepidis, P.; Papadopoulos, A. .; Tzirakis, F.; Tsivintzelis, I.; Seferlis, P. Optimum Design of Post-Combustion CO₂ Capture Processes Using Phase-Change Solvents. **2020**, No. Work in preparation.
- (179) Reid, L.; Sexton, A.; Fisher, K.; Dombrowski, K.; Youngerman, J.; Rochelle, G.;

- Chen, E.; Nielsen, P.; Singh, P.; Davison, J.; Beitler, C. Evaluation of Amine Reclaimer Operation and Waste Disposal From Post-Combustion Co₂ Capture. **2016**.
- (180) Damartzis, T.; Papadopoulos, A. I.; Seferlis, P.; Papadokostantakis, S.; Badr, S. *CAPSOL. Design Technologies for Multi-Scale Innovation and Integration in Post-Combustion CO₂ Capture: From Molecules to Unit Operations and Integrated Plants. Deliverable 2.6: Rigorous Design and Control of Absorption/Desorption Flowsheets*; 2014.

The reference list does not include all the articles used for the knowledge extraction procedure.