universität
wien

MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Compound-pathway interaction fingerprints for prediction of developmental neurotoxicity “

verfasst von / submitted by

Kreil Isabel

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra pharmaciae (Mag. pharm.)

Wien 2021/ Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 605
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Pharmazie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Gerhard Ecker

Acknowledgment

First of all, I would like to express my sincere thanks to my supervisor Univ.-Prof. Mag. Dr. Ecker Gerhard for letting me be a part of his research group and for his professional guidance and patience, especially in the times of Corona.

Also, I would like to thank all members of the Pharmacoinformatics Research Group for their help and their encouragement. Even though in times of Corona the presence in the group was cut down to a minimum, I always appreciated the nice and friendly atmosphere.

Last, but for sure not least, I would like to thank my mother who was always there to encourage me throughout my studies and never stopped believing in me. Studying would have been way harder without her by my side.

Abstract

The objective of this master thesis was to determine whether interaction fingerprints with different approaches can be used to predict possible toxic effects of chemical compounds. This thesis is based on a list of compounds believed to be involved in neuronal pathways that lead to developmental neurotoxicity. The list was provided by the EuToxRisk consortium within the framework of the neonicotinoid case study. First, a pharmacophore-compound interaction fingerprint was created to see if similar interaction patterns are visible within the compounds. Secondly, compound-pathway interaction fingerprints were obtained to see if assumptions on a drug's effect on the human body can be made due to its connected pathway. Different approaches of compound-pathway fingerprints were considered to illustrate if the pathway predictions overlap. The result shows that similar interaction patterns were found in the compound-pathway interaction fingerprints, provided that only the direct targets of the case study compounds were included.

Zusammenfassung

Das Ziel dieser Masterarbeit war es, festzustellen, ob verschiedene Interaktions - Fingerprints verwendet werden können, um die möglichen toxischen Wirkungen von chemischen Verbindungen mit nur computerbasierten Ansätzen vorherzusagen. Diese Arbeit basiert auf einer Liste von Verbindungen, von denen angenommen wird, dass sie an neuronalen Pfaden beteiligt sind, die zu Entwicklungsneurotoxizität führen. Die Liste wurde vom EuToxRisk Konsortium im Rahmen der Neonicotinoid Case Study erstellt. Zunächst wurde ein pharmakophor-basierter Fingerabdruck erstellt, um zu sehen, ob ähnliche Muster innerhalb der Verbindungen sichtbar sind. Zweitens wurden verschiedene Compound-Pathway Interaction - Fingerprints erstellt, um zu sehen, ob unterschiedliche Vorangehensweisen zu überlappenden Pathway-Vorhersagen führen. Das Ergebnis zeigt, dass ähnliche Interaktionsmuster in den Compound-Pathway Interaction - Fingerprints gefunden wurden, vorausgesetzt, dass nur die direkten Targets der Case Study compounds einbezogen wurden.

Table of Contents

1. General Background

1.1. Neonicotinoids

Neonicotinoids have become very popular in the use as synthetic insecticides because of their high efficacy for insect controls and the ease of application. They are widely used in agriculture to protect crops from insects such as aphids, whiteflies and other beetles that chew on plant tissues. However, the occurrence of neonicotinoid insecticides has become a worldwide problem due to their related toxicity risk for many species. [1]

As the name infers, neonicotinoids are chemically related to nicotine and have a common mode of action that affects the insect nervous system by mainly targeting the nicotinic acetylcholine receptors (nAChRs). Neonicotinoids bind to nAChRs and mimic the action of acetylcholine by opening the ion channel which allows the entry of cations (Na+ and Ca2+) and causes excitatory neurotransmission in the central nervous system. [2]

Even though they are known for being highly specific for insect nAChRs, recent animal studies assume that they may also attack the mammalian nAChRs which can lead to neurotoxic effects. This leads to the assumption that they have a potential impact on off-target organisms, including humans. [3]

Regarding their long biological half-lives (for example the biological half-lives of clothianidin and imidacloprid in soils were a few months and two to three years) neonicotinoids might be more ubiquitous than previously assumed. [4] Consequently, scientific evidence is rapidly growing in order to predict the effects of neonicotinoids on off-target organisms. [2]

1.2. Developmental neurotoxicity

The developing brain in children and fetus is way more sensitive to damage caused by toxic substances than the brain of an adult. The immature blood/brain barrier, the increased absorption compared to the low body weight and the reduced ability to detoxify xenobiotics contributes to the higher sensitivity of the developing brain. Furthermore, the development of the central nervous system is a complex process including differentiation of progenitor cells, proliferation, synthesis of neurotransmitter, cell death and formation of receptors. These events happen within a strict timeframe which leads to different stages of vulnerability to xenobiotic exposure for each event. Thus, once the neurodevelopment is disturbed by neurotoxic chemicals, there is only a small possibility for repair and it often leads to permanent consequential damage. [5]

Even though, neurodevelopmental disorders like mental retardation, autism and attention deficit/-hyperactivity syndrome are common disabilities nowadays, there is a lack of sufficient data about developmental neurotoxicity of nearly all chemicals, including environmental chemicals like pesticides. [5]

1.2.1. Neurotoxic potential of neonicotinoids in humans

Regarding the human brain function, nAChRs are of critical relevance especially during development and also for cognition, memory and behaviour. They are found in the central and peripheral nervous systems of mammals and in neuromuscular junctions. In the cell membrane nAChRs are composed of a combination of five subunits (α, ß, δ, γ and ξ) which are arranged around a central conducting pore. The pore is opened by the binding of a ligand which leads to the diffusion of natrium [Na+] and potassium [K+] through the pore. (Figure 1) [6]

In the vertebrate brain α7, α4β2 and α3β4 appear to be the major neuronal subtypes of nAChR, whereas α3β4 predominate in peripheral ganglia. [7] The α7 subtype of nAChR is widespread in the central nervous system and a variety of peripheral tissues. Especially, this subtype seems to participate in neuronal proliferation, migration, apoptosis, differentiation, synapse formation and neural-circuit formation, during the development of the human brain. So overall, due to their expression in the vertebrate brain, the subtypes α7, α4β2 and α3β4 of the nAChRs seem to be mainly involved in the mammalian brain development and are essential for its function. [6] [8] Thus, nicotine and neonicotinoids may affect important processes during embryonic development by activating the nAChRs and can cause permanent damage. [9]

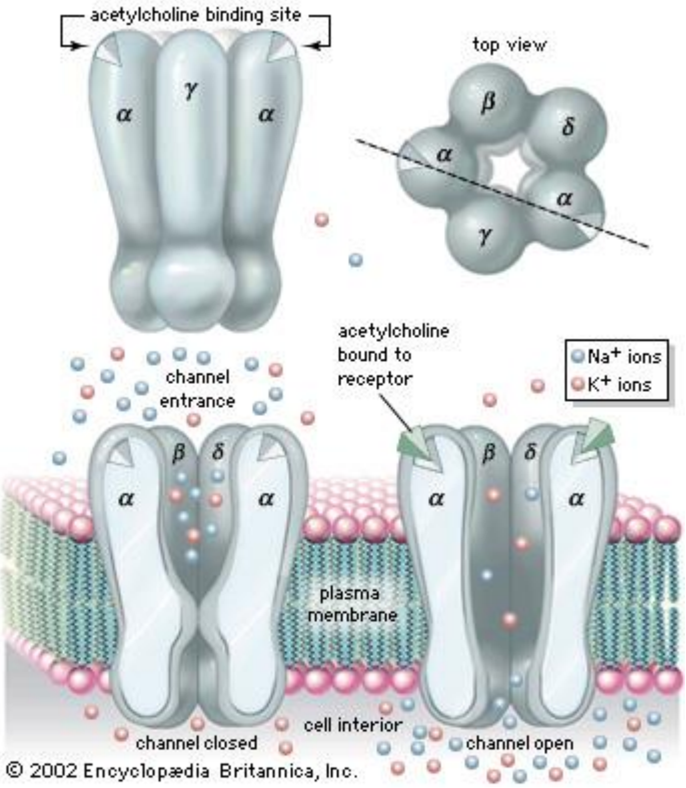


FIGURE 1 NICOTINIC ACETYLCHOLINE RECEPTOR: LIGAND-GATED ION CHANNEL COMPOSED OUT OF 5 SUBUNITS

1.3. Case Study 14 – EuToxRisk

The EuToxRisk project is an integrated European ‘flagship’ program conducting mechanism-based toxicity testing and risk assessment for the 21st century. The aim is to make a shift in toxicological testing, away from ‘black box’ animal testing, towards mechanistic and animal-free safety assessment which can be applied across industry sectors. In order to understand the complex ways that link chemical exposure to toxic outcome, EuToxRisk integrates advancements in cell biology, omics technologies, systems biology and computational modelling. The focus is mainly on two areas: repeated dose systemic toxicity and developmental/ reproductive toxicity. [10]

Within the framework of the project several case studies that are all related to a different toxic risk has been observed. In case study-14 (CS-14), which is the basis of this thesis, neonicotinoid pesticides are discussed in order to predict the potential developmental neurotoxic effects. Table 1 represents the compounds that were included in this case study.

Compounds	Molecular formula	ChEMBL-ID
2-Methylimidazole	C4H6N2	CHEMBL293391
4-Methylimidazole	C4H6N2	CHEMBL1230309
Butanone oxime	C4H9O1N1	CHEMBL2139230
Dibutyl-tin dichloride	C8H18Cl2Sn1	No ChEMBL ID
Butyl-tin trichloride	C4H9Cl3Sn1	No ChEMBL ID
Tributyl-tin chloride	C12H27Cl1Sn1	No ChEMBL ID
α-Bungarotoxin	C50H70O14	CHEMBL216458
Thiacloprid	C10H9ClN4S	CHEMBL451432
Desnitro-imidacloprid HCl	C9H11ClN4	CHEMBL309804
Imidacloprid	C9H10ClN5O2	CHEMBL406819
Thiamethoxam	C8H10ClN5O3S	CHEMBL1896251
Dinotefuran	C7H15N4O3+	CHEMBL2228155
Clothianidin	C6H8ClN5O2S	CHEMBL259727
Dihydro-beta-erythroidine hydrobromide	C16H22BrNO3	CHEMBL1319741
Nicotine	C10H14N2	CHEMBL3
2-Mercapto-benzimidazole	C7H6N2S	CHEMBL70141

N,N,N-Triethyl-2-(4-trans-stilbenoxy)-ethylammonium iodide	C22H30NOI	CHEMBL1257065
Mecamylamine HCl	C11H22ClN	CHEMBL1237082
WAY 317538	C20H25N3O2	CHEMBL1084615
2-Imidazolidinethione	C3H6N2S	CHEMBL11860
Actamiprid	C10H11ClN4	CHEMBL265941
Methyllycaconitine citrate salt	C43H58N2O17	CHEMBL510275
Imidacloprid-olefin	C9H8ClN5O2	CHEMBL95367
Epibatidine	C11H13ClN2	CHEMBL6623
AR-R 17779 HCl	C9H15ClN2O2	CHEMBL293975
Tebanicline HCl	C9H12Cl2N2O	CHEMBL430497
Bromocytisine	C11H13BrN2O	CHEMBL365956
PHA568487	C16H20N2O3	CHEMBL2086586
Tubocurarine HCl pentahydrate	C37H52Cl2N2O11	CHEMBL2063769

Table 1 Compounds of the CS-14

2. Aim of the research

The aim of this thesis was to evaluate if the potential neurotoxic effects of neonicotinoid insecticides can be predicted by different interaction fingerprints.

This thesis started with a list of compounds which are related to a potential developmental neurotoxic risk. The vision was to create different fingerprints to see whether the compounds of the CS-14 show analogous interaction patterns or not.

The first approach was to create a pharmacophore-compound interaction fingerprint. For this step, the direct targets of the compounds were downloaded by using two different workflows on KNIME. Subsequently, pharmacophore models were created with LigandScout, provided there were protein-ligand complexes in the PDB available. Then the CS-14 compounds were screened against the pharmacophore models and a fingerprint was created.

The second approach was to focus on the biological pathways of the compounds by using the 'Pathway-Interactom WF' on KNIME, which was made by Barbara Füzi. With the help of this WF two different fingerprints were created – a compound-pathway interaction fingerprint and an interactome-pathway fingerprint.

Finally, the database ToxPHACTS was included to get two more compound-pathway interaction fingerprints, but this time the targets were retrieved through ToxPHACTS. This database includes - besides the targets of the origin compound - also the targets of its similar compounds. One fingerprint was created by including the targets of the origin compounds and the targets of their similar compounds. The second compound-pathway interaction fingerprint has been made by excluding the targets of the similar compounds and only using the direct targets of the CS-14 compounds.

3. Methods

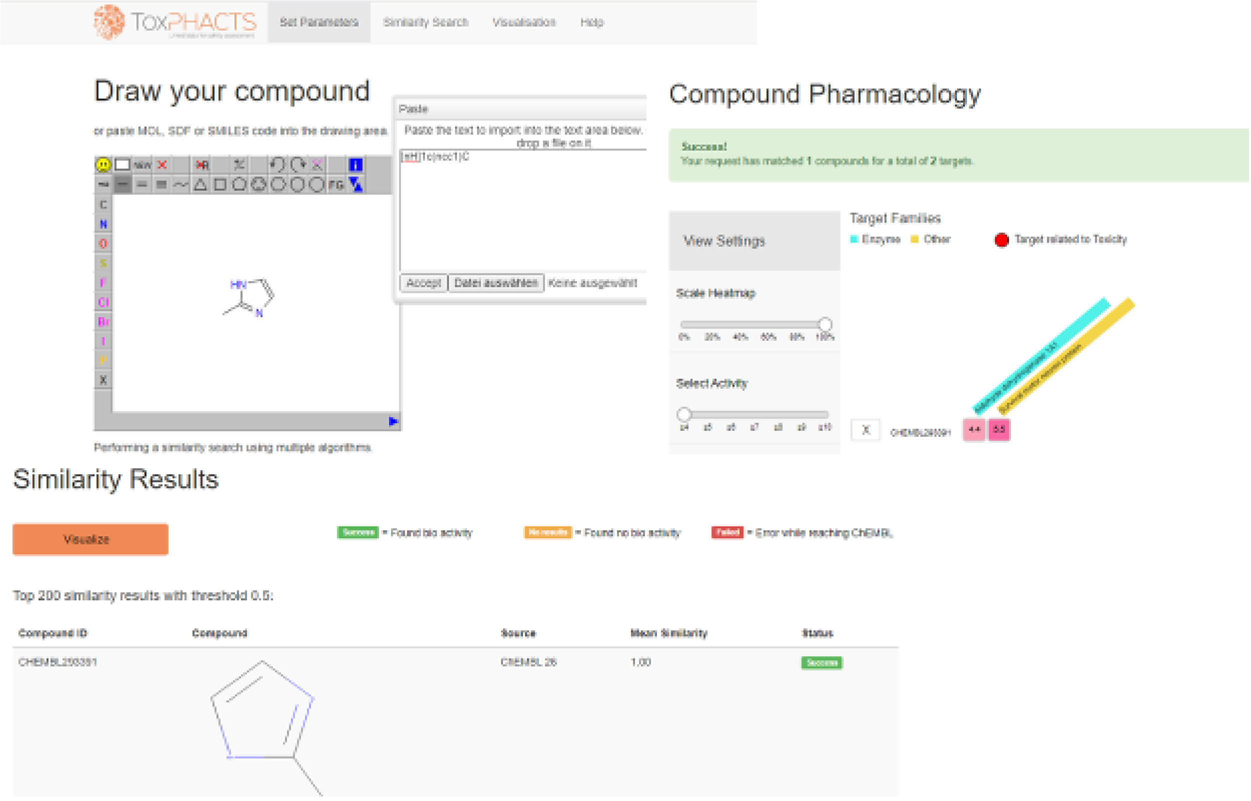
3.1. ToxPHACTS

ToxPHACTS is a software based on the analysis of large amounts of data for the early detection of side effects of new drug candidates with the aim to significantly reduce the number of animal experiments. ToxPHACTS integrates expertise in cheminformatics and semantic data integration to offer an expert system for toxicological read across. This process is based on the unique combination of similarity search, data extraction and data analysis:

- Similarity search: the target molecule is queried across the chemical space of ChEMBL by using 5 different similarity algorithms and consensus scoring

- Data extraction: the target profiles, for which bioactivity values are reported in ChEMBL, are retrieved and annotated with toxicity endpoints

- Data analysis: the results are visualized and can be seen in aggregated form [11]



Success!
Your request has matched 1 compounds for a total of 2 targets.

Top 200 similarity results with threshold 0.5:

Compound ID	Compound	Source	Mean Similarity	Status
CHEMBL250391		CHEMBL 26	1.00	Success

FIGURE 2 OVERVIEW OF TOXPHACTS

3.2. KNIME – The Konstanz Information Miner

KNIME is a is a free data mining tool which is commonly used in cheminformatics. With this open source software, it is possible to accomplish the complete mining process from data import, data processing to data analysis and visualization of the results. [12]

In KNIME the user can create workflows in order to process data. A workflow exists of a pipeline of nodes which are connected to each other and transport the input data. Each node processes the arriving data through its Import- and produces results on its Outport instances. Especially for data science KNIME workflows are from significant relevance due to its ability to process large amount of data fast. [13]

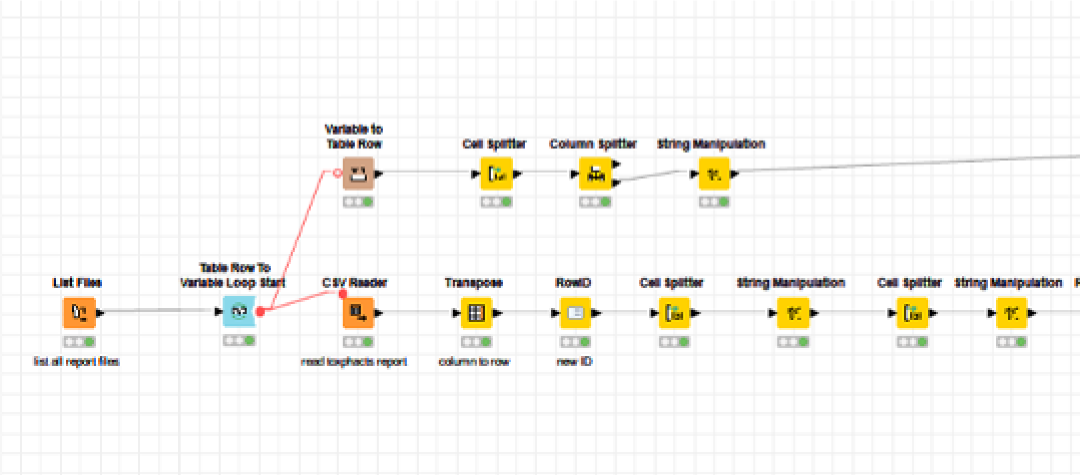


FIGURE 3 EXCERPT OF A WORKFLOW

3.3. ChEMBL Database

ChEMBL is an open chemical database which provides a wide range of information regarding binding, functional and ADMET properties for drug-like bioactive molecules. ChEMBL also provides predictive models to predict the possibility if a compound is active or inactive against a new target. This information can be helpful for the understanding of potential adverse effects of a drug. [14]

According to the status in 2017, ChEMBL provides over 1,6 million different compound structures with 14 million activity values from over 1,2 million assays. Those assays are assigned to 11 000 targets. [15]

The search tab in ChEMBL allows to search for targets or compounds, which are all linked to a unique ChEMBL-ID. The Protein Target search tab allows to browse through different protein types. Clicking on a target gives further information about other target identifiers and information on how many compounds, including their activity data, are associated with the target. [16]

Browsing through compounds can either be done by searching the name of the molecule or the search tab also provides a tool where the compound can be drawn and ChEMBL searches by identity, similarity or substructure for the compound or a similar one. For every compound which is registered in ChEMBL a compound report card exists which provides additional information such as database identifiers, names and computed properties. Additionally, bioactivity data (such as Ki, Kd, IC50 und EC50) are provided and, if compounds were tested against different targets, then also target predictions are available. [16]

3.4. LigandScout

LigandScout is a computer software that allows to create structure- or ligand-based pharmacophore models from structural data of protein-ligand complexes. The software uses different algorithms in order to perform alignments and analyze protein-ligand interactions. LigandScout provides two options on how the pharmacophore models can be derived, either the relevant chemical features are received from a known macromolecule-ligand complex (structure-based design), or the maximum common set of chemical features is searched for (ligand-based design). Afterwards, the created pharmacophore models can be screened against a virtual library to see if there are matching features. [17]

Regarding the structure based design, the software needs a PDB-ID as an input in order to be able to analyze the ligands and its macromolecular environment. To define the specific interactions that have been observed in drug-receptor interaction, pharmacophore features such as hydrogen bond donor, hydrogen bond acceptor, positive or negative ionizable chemical groups, hydrophobic interactions, aromatic rings and metal binding locations are provided by LigandScout. These pharmacophore features are represented through different icons; for example a hydrogen bond acceptor is shown as a red sphere with an arrow. (Figure 3) [18]

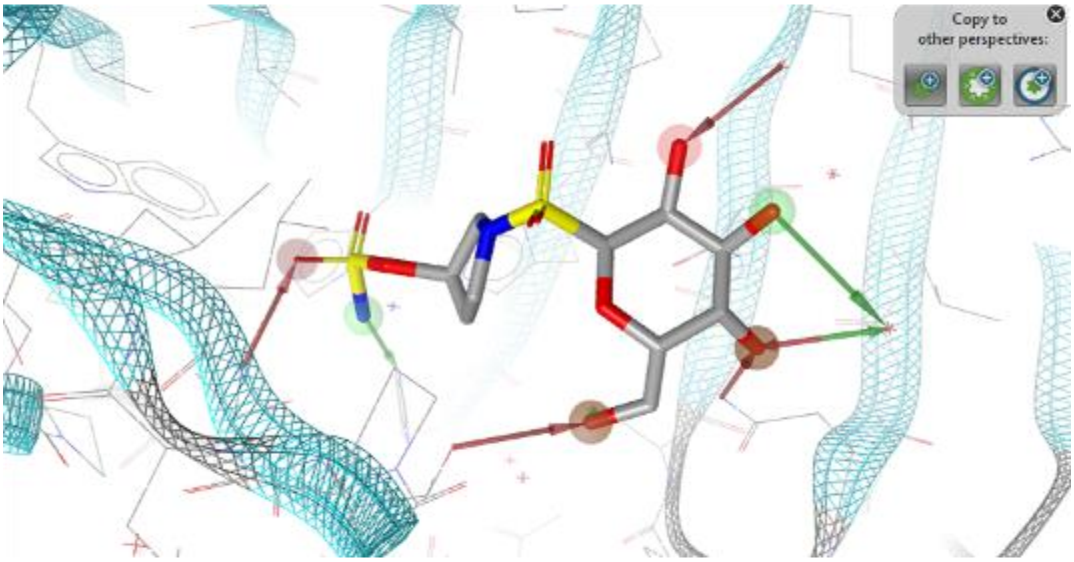


FIGURE 4 EXAMPLE FOR AN AUTOMATICALLY GENERATED PHARMACOPHORE MODEL WITH ITS PHARMACOPHORIC FEATURES

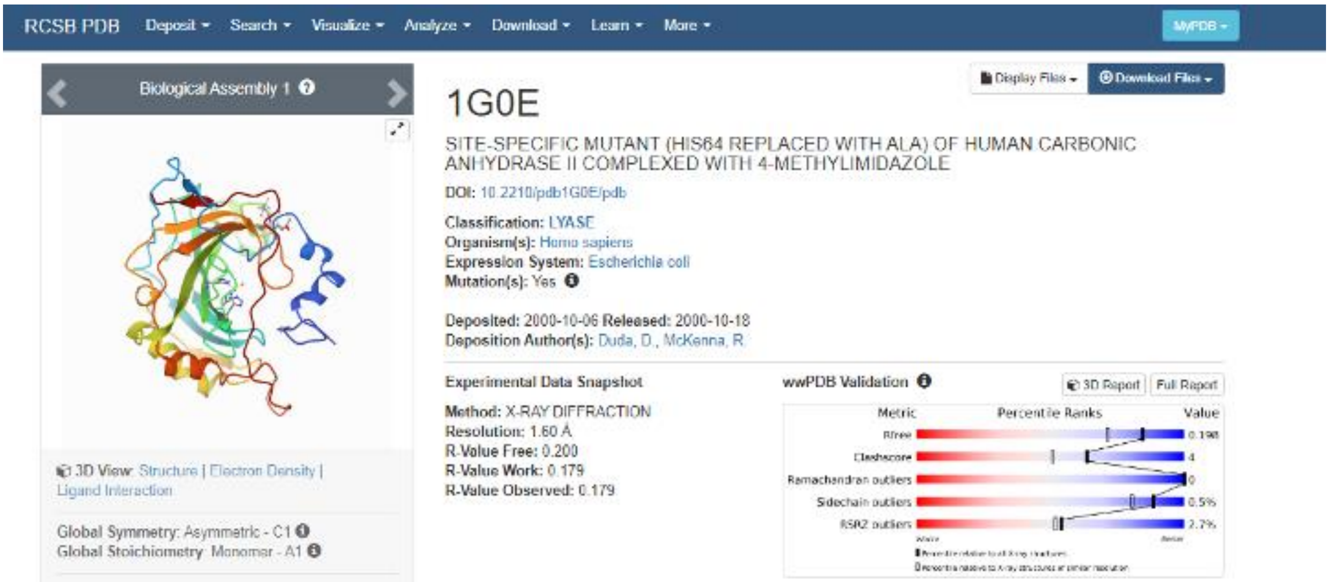
For this master thesis, the pharmacophore models have been created by using the structure based design approach and afterwards the generated pharmacophore models were screened against a library which consisted of the compounds of the CS-14.

3.5. PDB – Protein Data Bank

The Protein Data Bank (PDB), founded in 1971, is a worldwide accessible archive of three-dimensional structure data of biological macromolecules. According to the latest status, the PDB contains 175 282 protein structures. [19]

The structural data of large molecules, such as proteins and nucleic acids, is obtained by X-Ray crystallography, NMR (nuclear magnetic resonance) spectrometry and cryo-electron microscopy, which was submitted by biologists and biochemists from all over the world. [20]

Every structure which is published in the PDB receives a four-character alphanumeric identifier, its PDB-ID. Each PDB entry gives information about experimental data and information about the ligand and its macromolecule.



1G0E

SITE-SPECIFIC MUTANT (HIS64 REPLACED WITH ALA) OF HUMAN CARBONIC ANHYDRASE II COMPLEXED WITH 4-METHYLIMIDAZOLE

DOI: 10.2210/pdb1G0E/pdb

Classification: LYASE
Organism(s): Homo sapiens
Expression System: Escherichia coli
Mutation(s): Yes

Deposited: 2000-10-06 Released: 2000-10-18
Deposition Author(s): Duda, D., McKenna, R.

3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.60 Å
R-Value Free: 0.200
R-Value Work: 0.179
R-Value Observed: 0.179

wwPDB Validation

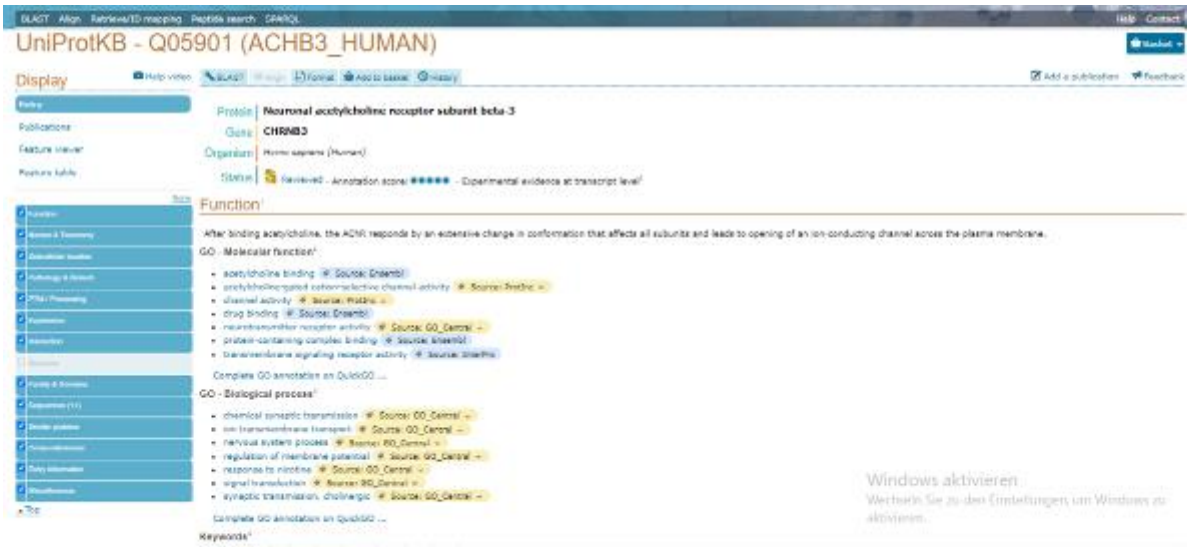
FIGURE 5 EXAMPLE OF AN ENTRY IN THE PDB

3.6. UniProt

UniProt is a freely accessible collection of protein sequences and functional information. It provides a large amount of data about the biological function of proteins submitted by research literature. [21]

In this thesis, UniprotKB/-Swiss-Prot which is one of the core databases of UniProt has been used. The aim of UniProtKB/-Swiss-Prot is to provide all known important information about one protein, such as the protein and gene name, function, subcellular location, protein-protein interactions and pattern of expression. [21]

Each protein can be identified by its UniProt-ID, which can be received through converting different identifiers (g.e. ChEMBL-ID) into UniProtI-IDs by using the Retrieve/ID mapping function on the UniProt website.



BLAST Align Retrieve/ID mapping Peptide search SPARQL
Help Contact
Basket
Display
Help video
BLAST Align Format Add to basket History
Add a publication Feedback
Publications
Feature viewer
Feature table
Top

UniProtKB - Q05901 (ACHB3_HUMAN)

Protein | Neuronal acetylcholine receptor subunit beta-3
Gene | CHRNB3
Organism | Homo sapiens (Human)
Status | Reviewed - Annotation score: ●●●●● - Experimental evidence at transcript level

Function

After binding acetylcholine, the AChR responds by an extensive change in conformation that affects all subunits and leads to opening of an ion-conducting channel across the plasma membrane.

GO - Molecular function

- acetylcholine binding Source: Ensembl
- acetylcholine-gated cation-selective channel activity Source: ProtInc
- channel activity Source: ProtInc
- drug binding Source: Ensembl
- neurotransmitter receptor activity Source: GO_Central
- protein-containing complex binding Source: Ensembl
- transmembrane signaling receptor activity Source: InterPro

Complete GO annotation on QuickGO ...

GO - Biological process

- chemical synaptic transmission Source: GO_Central
- ion transmembrane transport Source: GO_Central
- nervous system process Source: GO_Central
- regulation of membrane potential Source: GO_Central
- response to nicotine Source: GO_Central
- signal transduction Source: GO_Central
- synaptic transmission, cholinergic Source: GO_Central

Complete GO annotation on QuickGO ...

Keywords

Windows aktivieren
Wechseln Sie zu den Einstellungen, um Windows zu aktivieren.

FIGURE 6 OVERVIEW OF AN ENTRY IN THE UNIPROTKB

3.7. Binding MOAD – The mother of all databases

Binding MOAD (Mother Of All Databases), originally introduced in 2004, is a database with approximately 10 000 (according to the status in September, 2007) protein-ligand crystal structures. It is known for containing all biological relevant ligands and their binding-affinity data if experimentally derived. Each entry must have a resolution better than 2.5 Å and must contain a valid ligand. A biologically relevant ligand is considered if it is a peptide of 10 amino acids or less, oligonucleotide of 4 nucleotides or less, small organic molecules or a co-factor. Ligands which are not included are crystallographic additives, salts, buffers, metals, metallic catalytic centers and solvents. With Binding MOAD it is possible to only examine valid and useful ligands. [22]

3.8. PubChem

PubChem is a free database that contains information about severe chemical molecules. Due to its large amount of data, submitted by various other data sources, PubChem is popular within researchers. [23]

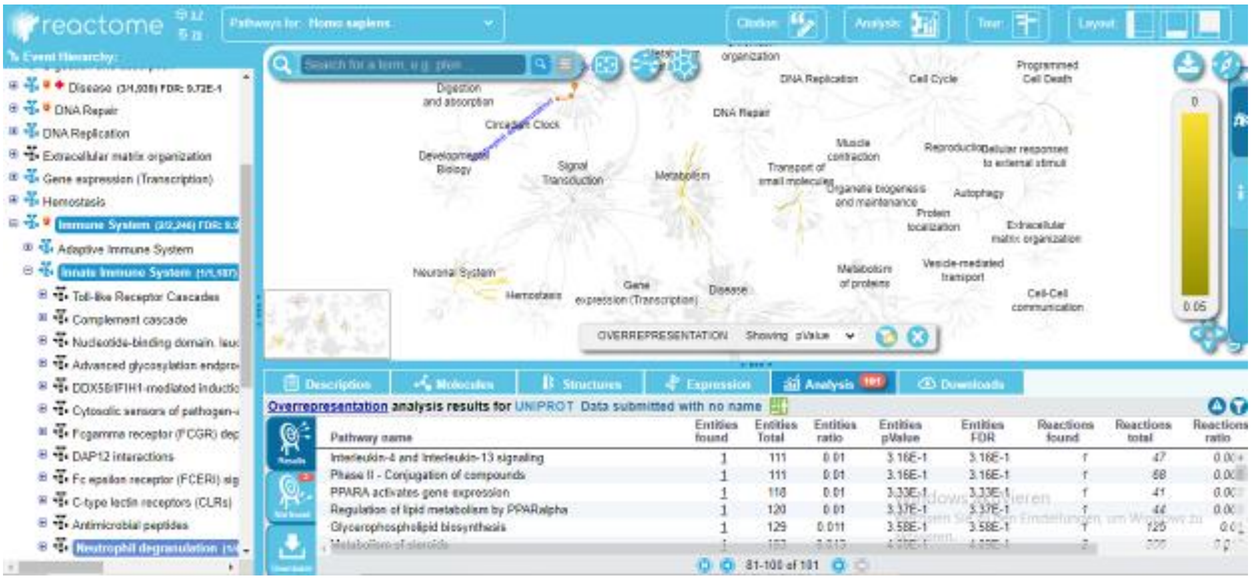
Searching for a compound in the database delivers information about its chemical structure including SMILES and InCHI strings, synonyms and other identifiers, chemical properties and other compounds that are structurally related to them.

3.9. Reactome

Reactome is an open-source database which represents human biological pathways and reactions. If a specific process has not been studied in humans, it is manually projected back to the human being from experiments performed on another species. Reactome connects the knowledge of expert biologists on how biological pathways work with several data sources.

With cross-references to external databases, such as UniProt, ChEBI or Ensembl, pathway annotations are created. Topics of pathways in Reactome are neuronal function, signaling, immune function, cell cycle, classical intermediary metabolism, apoptosis, transport, and host-virus interaction. [24]

The basic principle of Reactome is to browse through pathways, submit data to a range of data analysis tools and to visualize the pathway diagrams. After putting valid data in the Reactome browser, the pathways are visualized in an interactive overview. [25]


reactome
Pathways for: Homo sapiens
Citation
Analysis
Tour
Layout
Event Hierarchy:
Disease (31,928) FDR: 9.72E-1
DNA Repair
DNA Replication
Extracellular matrix organization
Gene expression (Transcription)
Hemostasis
Immune System (2/2,246) FDR: 9.
Adaptive Immune System
Innate Immune System (1/1,187)
Toll-like Receptor Cascades
Complement cascade
Nucleotide-binding domain, leu
Advanced glycosylation endpro
DDX58/IFIH1-mediated inductio
Cytosolic sensors of pathogen-
Fcgamma receptor (FCGR) dep
DAP12 interactions
Fc epsilon receptor (FCERI) sig
C-type lectin receptors (CLRs)
Antimicrobial peptides
Neutrophil degranulation
Search for a term, e.g. pten
Digestion and absorption
Circadian Clock
Developmental Biology
Signal Transduction
Metabolism
organization
DNA Replication
Cell Cycle
Programmed Cell Death
DNA Repair
Muscle contraction
Transport of small molecules
Reproduction
Cellular responses to external stimuli
Organelle biogenesis and maintenance
Autophagy
Protein localization
Extracellular matrix organization
Neuronal System
Hemostasis
Gene expression (Transcription)
Disease
Metabolism of proteins
Vesicle-mediated transport
Cell-Cell communication
0
0.05
OVERREPRESENTATION Showing pValue
Description
Molecules
Structures
Expression
Analysis 101
Downloads
Results
81-100 of 101

Overrepresentation analysis results for UNIPROT Data submitted with no name

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio
Interleukin-4 and Interleukin-13 signaling	1	111	0.01	3.16E-1	3.16E-1	1	47	0.00
Phase II - Conjugation of compounds	1	111	0.01	3.16E-1	3.16E-1	1	68	0.00
PPARA activates gene expression	1	118	0.01	3.33E-1	3.33E-1	1	41	0.00
Regulation of lipid metabolism by PPARalpha	1	120	0.01	3.37E-1	3.37E-1	1	44	0.00
Glycerophospholipid biosynthesis	1	129	0.011	3.58E-1	3.58E-1	1	125	0.0
Metabolism of steroids	1	153	0.013	4.09E-1	4.09E-1	2	200	0.0

FIGURE 7 OVERVIEW OF THE REACTOME DATABASE

3.10. Creating the heatmaps - RStudio

For creating the heatmaps in this thesis, the open-source integrated development environment (IDE) for R, RStudio, was used. RStudio is a programming language for statistical computing and graphics. By installing various packages, RStudio can be widely used especially when it comes to data science where a large amount of data often needs to be visualized. The use of RStudio is easily learned due to the number of prefabricated codes that are available on the internet. [26]

4. Compound-pharmacophore fingerprint

A pharmacophore model represents a set of interactions such as chemical features and interactions, that are aligned in a three-dimensional space. The spatial arrangement of chemical features illustrates the relevant interactions of a ligand-macromolecular complex. [17]
The pharmacophoric features - hydrogen bond donor or acceptor, negative or positive ionizable features, aromaticity and hydrophobicity - are assumed to be responsible for the molecular recognition of a ligand by a biological target in order to receive a specific pharmacological action. Thus, pharmacophore models offer a good concept to identify a possible ligand by its binding related structural or chemical properties. [27]

As already mentioned, there are two approaches to create pharmacophore models, one is the structure-based design and the other the ligand-based one. This thesis uses the structure-based design approach by concentrating on generating pharmacopohore models from known three-dimensional ligand-protein complexes.

4.1. Data preparation

4.1.1. KNIME

The first step was to download the targets of the CS-14 compounds by using the 'Target WF' on KNIME that has been created by Barbara Füzi. The basic principle behind this WF is to connect the ChEMBL-IDs of the compounds with the UniProt-ID of their related target protein. In order to execute the WF successfully, a valid ChEMBL-ID of the compounds needs to be available.

The ChEMBL-IDs of the CS-14 compounds have either been received by using the drawing tool in ChEMBL (SMILES code was used as an input) or, if this approach did not lead to any result, the compound has been searched in PubChem to find its ChEMBL-ID.

After putting the ChEMBL-IDs of the compounds into the WF, the WF searches through different databases (ChEMBL, DrugBank TTD, PharmGKB and IUPHAR) and the data for the target proteins is accessed through API Calls in KNIME. All the retrieved targets are based on activity values or reported targets from literature in the different databases. Afterwards, since the target proteins have their own identifier due to the different databases, the WF translates the identifiers into Uniprot-IDs. At the end of the WF a table is provided which connects the ChEMBL-ID of the compounds to their target proteins (provided a UniProt-ID for the target was available).

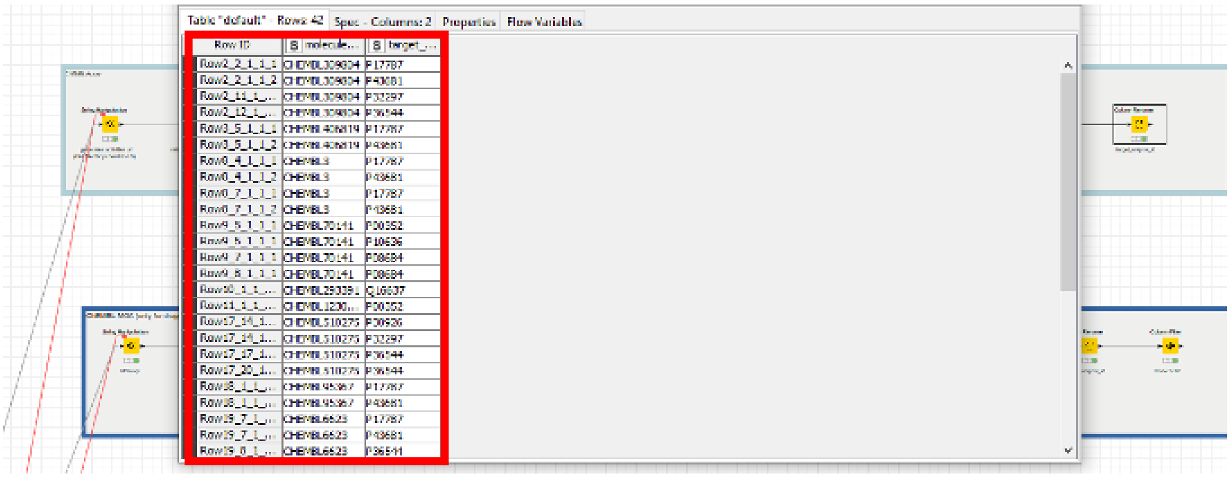


Table "default" - Rows: 42 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	S molecule...	S target_...
Row2_2_1_1_1	CHEMBL309804	P17787
Row2_2_1_1_2	CHEMBL309804	P43681
Row2_11_1_...	CHEMBL309804	P32297
Row2_12_1_...	CHEMBL309804	P36544
Row3_5_1_1_1	CHEMBL406819	P17787
Row3_5_1_1_2	CHEMBL406819	P43681
Row0_4_1_1_1	CHEMBL3	P17787
Row0_4_1_1_2	CHEMBL3	P43681
Row0_7_1_1_1	CHEMBL3	P17787
Row0_7_1_1_2	CHEMBL3	P43681
Row4_5_1_1_1	CHEMBL70141	P00352
Row4_6_1_1_1	CHEMBL70141	P10636
Row4_7_1_1_1	CHEMBL70141	P08684
Row4_8_1_1_1	CHEMBL70141	P08684
Row10_1_1_...	CHEMBL293391	Q16637
Row11_1_1_...	CHEMBL1230...	P00352
Row17_14_1...	CHEMBL510275	P30926
Row17_14_1...	CHEMBL510275	P32297
Row17_17_1...	CHEMBL510275	P36544
Row17_20_1...	CHEMBL510275	P36544
Row18_1_1_...	CHEMBL95367	P17787
Row18_1_1_...	CHEMBL95367	P43681
Row19_7_1_...	CHEMBL6523	P17787
Row19_7_1_...	CHEMBL6523	P43681
Row19_8_1_...	CHEMBL6523	P36544

FIGURE 8 CHEMBL-IDS OF THE COMPOUNDS AND THEIR CONNECTED TARGET PROTEIN

For only 12 out of the 29 compounds of the CS-14, valid results - valid is considered when the ChEMBL-ID could successfully be connected to a target protein - could be received through the WF. A total of 39 different target proteins were provided. Table 2 represents the connected compound-target pairs.

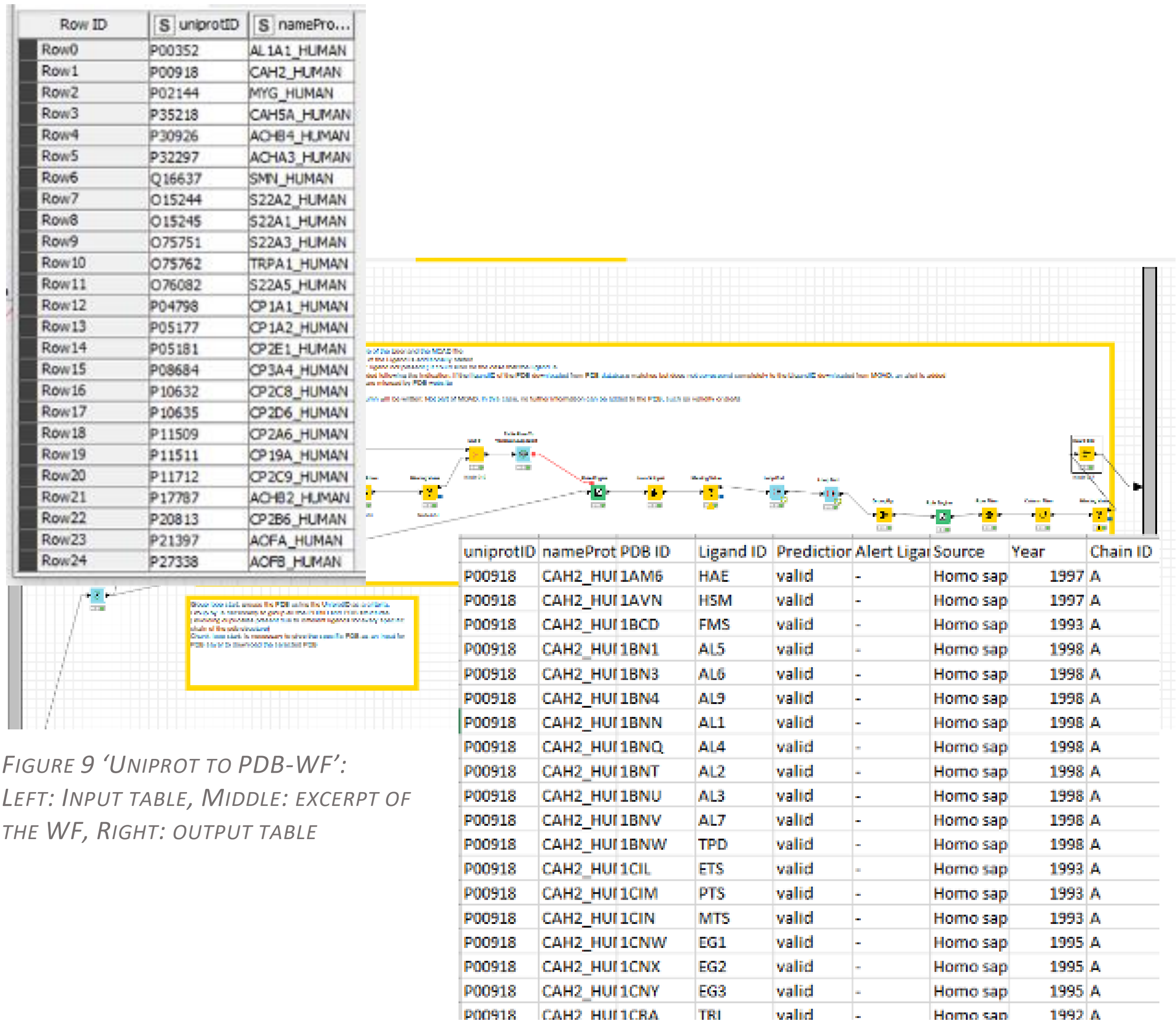
ChEMBL-ID of the compounds	**Target identifier (UniProt-ID)**
CHEMBL309804	P17787, P43681, P32297, P36544
CHEMBL406819	P17787, P43681, Q494W8
CHEMBL3	P17787, P43681, P36544, Q15822, P32297, P30532, Q15825, Q9UGM1, Q9GZZ6, Q05901, P30926, P11511, P28329, P11509,

	P05181, Q16696, P20813, P04798, P05177, P33261, P10632, P11712, P10635, P08684, P21397, P27338, O15244, O15245, O75751, O76082, Q9H015, O75762
CHEMBL70141	P00352, P10636, P08684
CHEMBL293391	Q16637
CHEMBL1230309	P00352, P00918, P02144, P35218
CHEMBL510275	P30926, P32297, P36544
CHEMBL95367	P17787, P43681
CHEMBL6623	P17787, P43681, P36544, P30926, P32297, Q15822
CHEMBL430497	P17787, P43681, P36544, P30926, P32297
CHEMBL365956	P30926, P32297
CHEMBL451432	Q494W8

Table 2 Compounds connected to their related target-ID

The second step was to get access to the PDB-IDs of the obtained targets in order to be able to create pharmacophore models. For this approach, another WF in KNIME which translates the UniProt-ID to PDB-IDs, was included. The ‘UniProt_to_PDB’ WF was created by Giulia Bianco and Riccardo Martini and provides information if ligand-protein complexes of the proteins of interest are available in the PDB. For this, the two databases, UniProt and PDB, were connected.

To start the WF, an input file which consists of at least two columns - one with the protein name and the other with its UniProt-ID – needs to be present. Subsequently, the WF translates the Uniprot-IDs into PDB-IDs by reaching access through the XML Query node. Furthermore, the database ‘Binding MOAD’ was included in order to only get ligand-protein complexes with valid ligands. Within the WF it is possible to customize the settings of some nodes in order to only receive the required information. In this thesis the information about the PDB ID, the Chain ID, the ligand ID, the name of the protein and the ligand validity prediction by MOAD was used in order to make further steps.



Row ID	S uniprotID	S namePro...
Row0	P00352	AL1A1_HUMAN
Row1	P00918	CAH2_HUMAN
Row2	P02144	MYG_HUMAN
Row3	P35218	CAH5A_HUMAN
Row4	P30926	ACHB4_HUMAN
Row5	P32297	ACHA3_HUMAN
Row6	Q16637	SMN_HUMAN
Row7	O15244	S22A2_HUMAN
Row8	O15245	S22A1_HUMAN
Row9	O75751	S22A3_HUMAN
Row10	O75762	TRPA1_HUMAN
Row11	O76082	S22A5_HUMAN
Row12	P04798	CP1A1_HUMAN
Row13	P05177	CP1A2_HUMAN
Row14	P05181	CP2E1_HUMAN
Row15	P08684	CP3A4_HUMAN
Row16	P10632	CP2C8_HUMAN
Row17	P10635	CP2D6_HUMAN
Row18	P11509	CP2A6_HUMAN
Row19	P11511	CP19A_HUMAN
Row20	P11712	CP2C9_HUMAN
Row21	P17787	ACHB2_HUMAN
Row22	P20813	CP2B6_HUMAN
Row23	P21397	AOFA_HUMAN
Row24	P27338	AOFB_HUMAN

uniprotID	nameProt	PDB ID	Ligand ID	Predictior	Alert Ligai	Source	Year	Chain ID
P00918	CAH2_HUI	1AM6	HAE	valid	-	Homo sap	1997	A
P00918	CAH2_HUI	1AVN	H5M	valid	-	Homo sap	1997	A
P00918	CAH2_HUI	1BCD	FMS	valid	-	Homo sap	1993	A
P00918	CAH2_HUI	1BN1	AL5	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BN3	AL6	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BN4	AL9	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNN	AL1	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNQ	AL4	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNT	AL2	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNU	AL3	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNV	AL7	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1BNW	TPD	valid	-	Homo sap	1998	A
P00918	CAH2_HUI	1CIL	ETS	valid	-	Homo sap	1993	A
P00918	CAH2_HUI	1CIM	PTS	valid	-	Homo sap	1993	A
P00918	CAH2_HUI	1CIN	MTS	valid	-	Homo sap	1993	A
P00918	CAH2_HUI	1CNW	EG1	valid	-	Homo sap	1995	A
P00918	CAH2_HUI	1CNX	EG2	valid	-	Homo sap	1995	A
P00918	CAH2_HUI	1CNY	EG3	valid	-	Homo sap	1995	A
P00918	CAH2_HUI	1CRA	TRI	valid	-	Homo sap	1992	A

FIGURE 9 'UNIPROT TO PDB-WF': LEFT: INPUT TABLE, MIDDLE: EXCERPT OF THE WF, RIGHT: OUTPUT TABLE

For only 14 targets of the CS-14 compounds - exited from the 39 targets that have been obtained with the first WF – valid ligand-protein complexes could be found in the PDB. A total of 372 different ligand-protein complexes were provided through the WF.

4.1.2. Pharmacophore models (structure-based design approach)

The pharmacophore models were created by using the PDB-IDs of the known ligand-protein complexes, thus a structure-based design was pursued. The PDB information is downloaded by LigandScout and by clicking on the binding site with the ligand an automated pharmacophore model was generated which then were saved to the screening perspective. After creating all pharmacophore models for each target protein, the compounds of the CS-14 were screened against every pharmacophore in order to see if there are matching chemical or structural properties. The settings in the screening perspective were that all query features of the pharmacophore needed to be matched in order to get a ‘hit’, furthermore the best fit score was used. After screening every target against the library of the compounds, a binary table was created.

If there was a hit between a pharmacophore and a compound, it was considered as 1 in the table and if the pharmacophoric features did not match, it was considered 0. The binary fingerprint table is attached below. (Table 3)

	253 pharamcophores	35 pharmacophores	2 pharmacophores	1 pharmacophore	39 pharmacophores	18 pharmacophores	3 pharmacophores
	CAH2	AOFB	CLAT	CP1A2	CP2A6	CP2AD	CP2B6
2-Methylimidazole	0	0	0	0	0	0	0
4- Methylimidazole	1	1	0	0	0	1	0
Butanone oxime	1	1	0	0	1	0	0
Dibutyl-tin dichloride	0	0	0	0	0	0	0
Butyl-tin trichloride	0	0	0	0	0	0	0
Tributyl-tin chloride	0	0	0	0	0	0	0
α-Bungarotoxin	1	0	0	0	0	0	0
Thiacloprid	1	1	0	0	1	1	0
Desnitro-imidacloprid HCl	1	1	0	0	1	1	0
Imidacloprid	1	1	0	0	1	1	0
Thiamethoxam	1	1	1	0	1	1	0
Dinotefuran	1	0	0	0	0	0	0

Clothianidin	1	1	0	0	1	1	0
Dihydro-beta- erythroidine hydrobromide	1	0	0	0	0	1	0
Nicotine	0	0	0	0	0	1	0
2-Mercapto-benzimidazole	1	0	0	0	0	0	0
N,N,N-Triethyl-2-(4-trans-stilbenoxy)ethylammonium iodide	1	1	0	0	1	0	0
Mecamylamine HCl	0	1	0	0	0	0	0
WAY 317538	1	1	0	0	0	1	0
2-Imidazolidinethione	0	0	0	0	0	0	0
Acetamipirid	1	1	0	0	1	1	0
Methyllycaconitine citrate salt	1	1	0	0	0	0	0
Imidacloprid-olefin	1	1	0	0	1	1	1
Epibatidine	1	1	0	0	0	0	0
AR-R 17779 hydrochloride	0	0	0	0	0	0	0
Tebanicline HCl	1	1	0	1	1	0	1
Bromocytisine	0	0	0	0	0	0	0
PHA568487	1	0	0	0	0	1	0
Tubocurarine HCl	1	1	0	0	1	0	0

Table 3 Pharmacophore fingerprint

4.2. Results

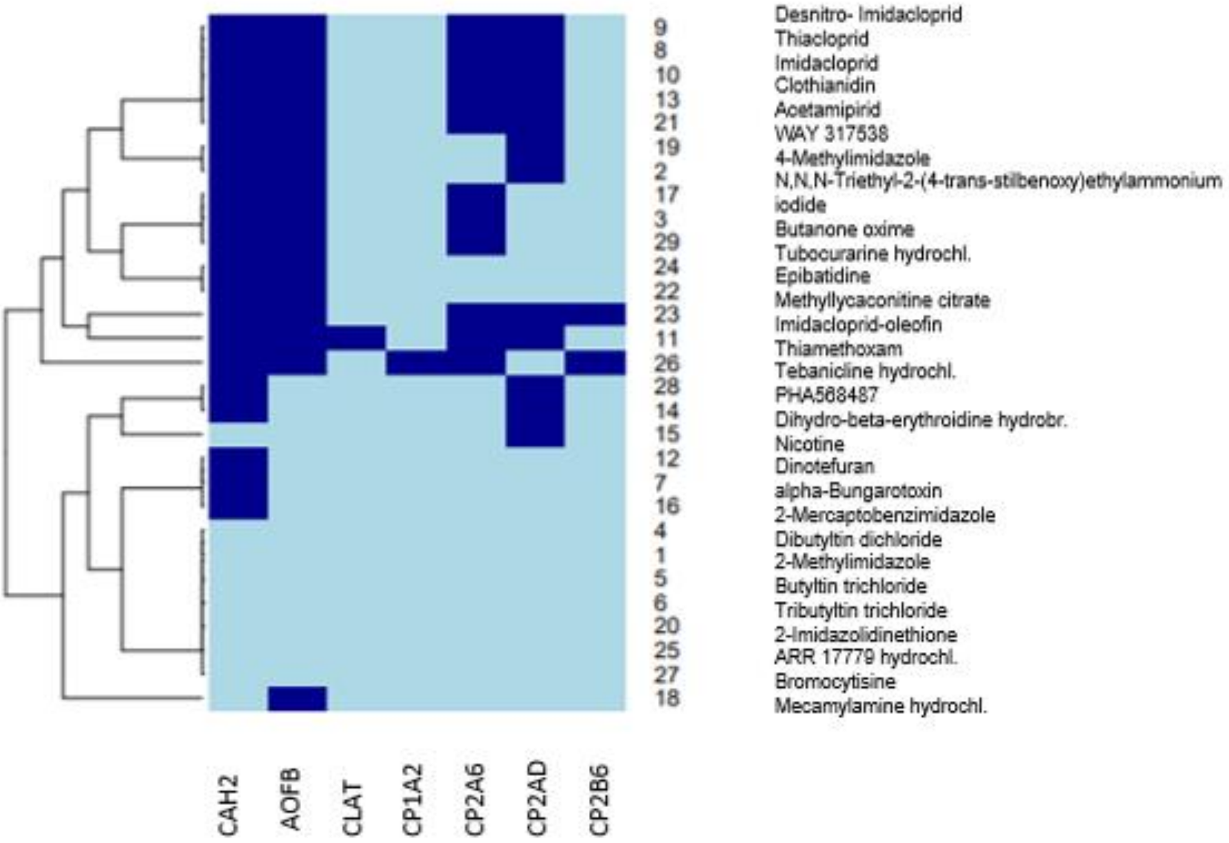


FIGURE 10 PHARMACOPHORE-COMPOUND FINGERPRINT

The heatmap presents the target proteins for which pharmacophore models were created on the x-axis and on the y-axis are the CS-14 compounds. The heatmap arranges the compounds after their similarities, thus substances with same or similar outcomes are beneath each other which means they might show similar interaction patterns. The first five compounds - Desnitro-Imidacloprid, Thiacloprid, Imidacloprid, Clothianidin and Acetamipirid - hit the same pharmacophore models and therefore build clusters. Furthermore, the pharmacophore models of CAH2 have matched nearly every compound of the CS-14, but regarding its 253 different pharmacophores, the results cannot be equally compared to the others where only 1, 2 or 3 pharmacophores have been generated.

It should also be noted that for 7 compounds no hits were recorded, including Bromocytisine, ARR 17779, 2-Imidazolidinethione, Tributyl-tin trichloride, Butyl-tin trichloride, 2-Methylimidazole and Dibutyl-tin dichloride.

Regarding the different amount of created pharmacophore models the results can hardly be compared and it also needs to be considered, that not for every target protein of the CS-14 compounds an entry in the PDB was available.

5. Compound-pathway interaction fingerprint

Pathways play an important part regarding the understanding of a drug's mechanism on a biological system. It is known that chemical substances which lead to a similar biological response in one's body, may not attack the same targets, but can still be involved in similar pathway profiles. The reason for this is that different targets can participate in the same pathways which then leads to the same pharmacological or toxic outcome. Analyzing the pathways of compounds can therefore be beneficial when it comes to predicting the therapeutic or toxic effects of new discovered drugs. [28]

In this thesis, the compound-pathway interaction fingerprints were created to show if the compounds of the CS-14 overlap in their pathway profiles and thus, prediction regarding their neurotoxic risks on the human body can be made.

5.1. Data preparation

5.1.1. KNIME

In order to identify the pathways that are potentially affected by the CS-14 compounds, another WF which was also created by Barbara Füzi has been used in this thesis.

First of all, the 'Pathway-Interactome' WF needs a table as input, where the ChEMBL-IDs of the compounds are connected to the UniProt-IDs of their target proteins. The compound-target protein pairs have been received with the WF explained in Chapter 4.1.1. and are illustrated in Table 2.

The 'Pathway-Interactome' WF makes sure that only reviewed human protein type targets are used. Furthermore, mammalian targets can also be translated based on gene names. In Figure 11 a section of the WF is shown.

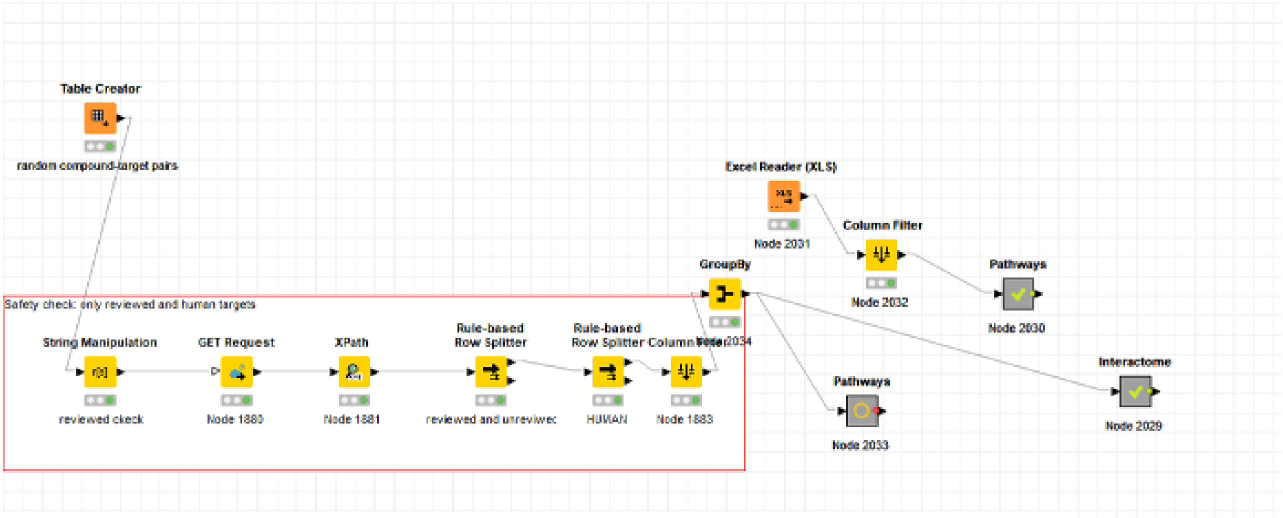
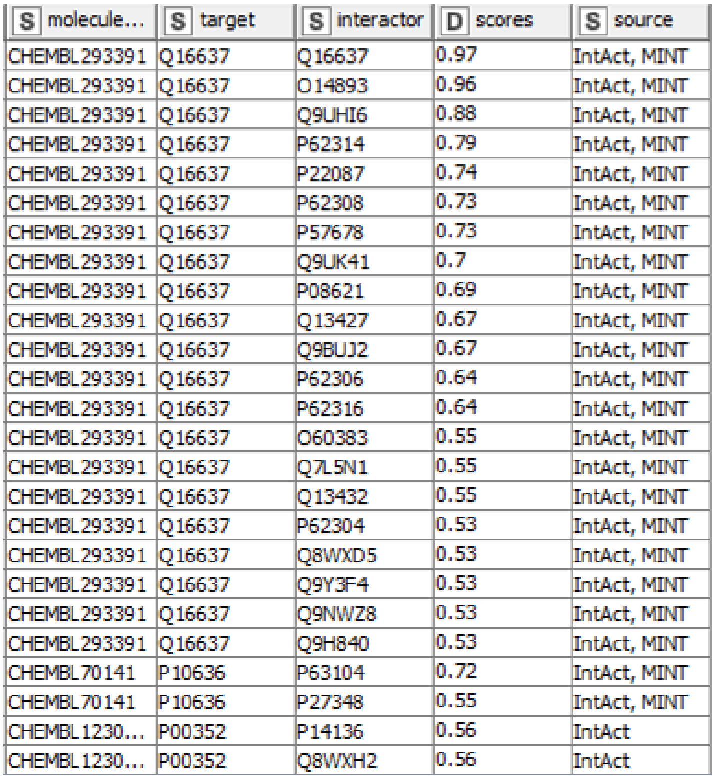


FIGURE 11 EXCERPT OF THE 'PATHWAY-INTERACTOM' WF

After checking the targets, the WF offers the possibility to either connect the target proteins to their interactome or directly to their potential pathways. In this thesis, both ways have been used to obtain two different fingerprints, one with the interactome and the other with the potential pathway of the direct targets.

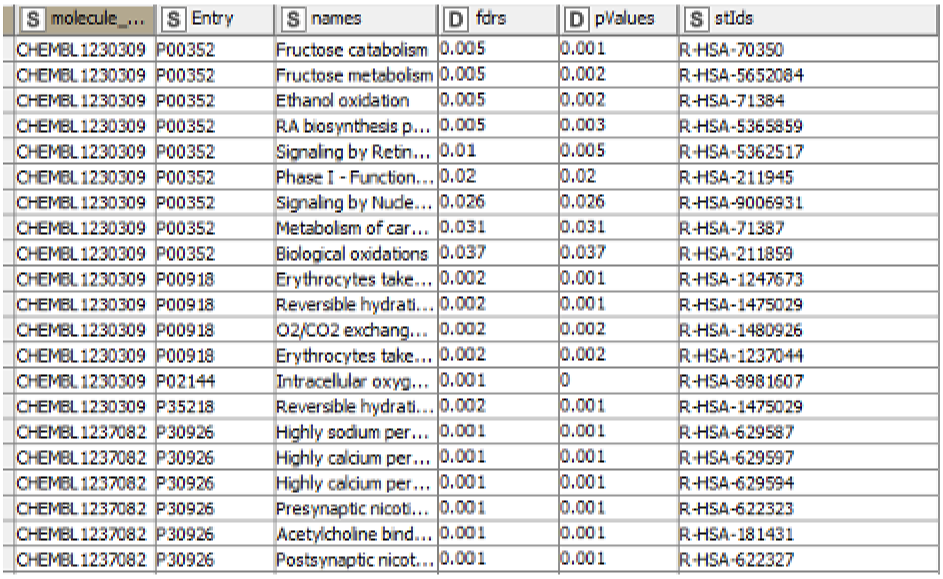
To start with the interactome part, the previously retrieved target proteins are now connected to their interactome. An interactome is described as a set of molecular interactions in a particular cell, also known as protein-protein interactions. [29] This WF examines the whole protein-protein interactions of the targets identified. For the interactome data, the databases IntAct and MINT were used. This part of the WF only searches for human and reviewed interactors. A cut off for the scores of the interactors is applied in order to receive a stable dataset for the analysis, so every interactor without a known score was excluded. An example for the output of the interactome part of the WF is given below. (Figure 12)



S molecule…	S target	S interactor	D scores	S source
CHEMBL293391	Q16637	Q16637	0.97	IntAct, MINT
CHEMBL293391	Q16637	O14893	0.96	IntAct, MINT
CHEMBL293391	Q16637	Q9UHI6	0.88	IntAct, MINT
CHEMBL293391	Q16637	P62314	0.79	IntAct, MINT
CHEMBL293391	Q16637	P22087	0.74	IntAct, MINT
CHEMBL293391	Q16637	P62308	0.73	IntAct, MINT
CHEMBL293391	Q16637	P57678	0.73	IntAct, MINT
CHEMBL293391	Q16637	Q9UK41	0.7	IntAct, MINT
CHEMBL293391	Q16637	P08621	0.69	IntAct, MINT
CHEMBL293391	Q16637	Q13427	0.67	IntAct, MINT
CHEMBL293391	Q16637	Q9BUJ2	0.67	IntAct, MINT
CHEMBL293391	Q16637	P62306	0.64	IntAct, MINT
CHEMBL293391	Q16637	P62316	0.64	IntAct, MINT
CHEMBL293391	Q16637	O60383	0.55	IntAct, MINT
CHEMBL293391	Q16637	Q7L5N1	0.55	IntAct, MINT
CHEMBL293391	Q16637	Q13432	0.55	IntAct, MINT
CHEMBL293391	Q16637	P62304	0.53	IntAct, MINT
CHEMBL293391	Q16637	Q8WXD5	0.53	IntAct, MINT
CHEMBL293391	Q16637	Q9Y3F4	0.53	IntAct, MINT
CHEMBL293391	Q16637	Q9NWZ8	0.53	IntAct, MINT
CHEMBL293391	Q16637	Q9H840	0.53	IntAct, MINT
CHEMBL70141	P10636	P63104	0.72	IntAct, MINT
CHEMBL70141	P10636	P27348	0.55	IntAct, MINT
CHEMBL1230…	P00352	P14136	0.56	IntAct
CHEMBL1230…	P00352	Q8WXH2	0.56	IntAct

FIGURE 12 TABLE CREATED WITH THE INTERACTOME PART OF THE WF

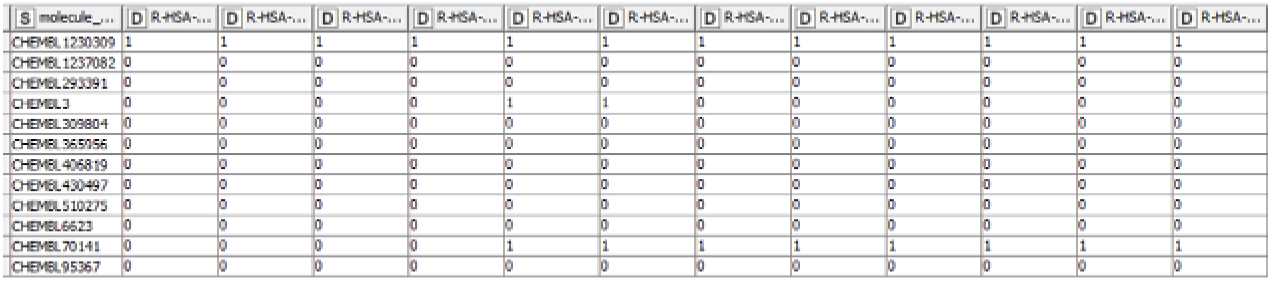
The other option of the WF is the pathway part, for which the Reactome database is used. This database does not only analyze the pathways, but also provides a visualization tool for an overrepresentation of the results. The output of the WF delivers the names of the pathways and their identifiers with statistical values such as FDR (false discovery rate) and p-values. (Figure 13)



S molecule_…	S Entry	S names	D fdrs	D pValues	S stIds
CHEMBL1230309	P00352	Fructose catabolism	0.005	0.001	R-HSA-70350
CHEMBL1230309	P00352	Fructose metabolism	0.005	0.002	R-HSA-5652084
CHEMBL1230309	P00352	Ethanol oxidation	0.005	0.002	R-HSA-71384
CHEMBL1230309	P00352	RA biosynthesis p…	0.005	0.003	R-HSA-5365859
CHEMBL1230309	P00352	Signaling by Retin…	0.01	0.005	R-HSA-5362517
CHEMBL1230309	P00352	Phase I - Function…	0.02	0.02	R-HSA-211945
CHEMBL1230309	P00352	Signaling by Nucle…	0.026	0.026	R-HSA-9006931
CHEMBL1230309	P00352	Metabolism of car…	0.031	0.031	R-HSA-71387
CHEMBL1230309	P00352	Biological oxidations	0.037	0.037	R-HSA-211859
CHEMBL1230309	P00918	Erythrocytes take…	0.002	0.001	R-HSA-1247673
CHEMBL1230309	P00918	Reversible hydrati…	0.002	0.001	R-HSA-1475029
CHEMBL1230309	P00918	O2/CO2 exchang…	0.002	0.002	R-HSA-1480926
CHEMBL1230309	P00918	Erythrocytes take…	0.002	0.002	R-HSA-1237044
CHEMBL1230309	P02144	Intracellular oxyg…	0.001	0	R-HSA-8981607
CHEMBL1230309	P35218	Reversible hydrati…	0.002	0.001	R-HSA-1475029
CHEMBL1237082	P30926	Highly sodium per…	0.001	0.001	R-HSA-629587
CHEMBL1237082	P30926	Highly calcium per…	0.001	0.001	R-HSA-629597
CHEMBL1237082	P30926	Highly calcium per…	0.001	0.001	R-HSA-629594
CHEMBL1237082	P30926	Presynaptic nicoti…	0.001	0.001	R-HSA-622323
CHEMBL1237082	P30926	Acetylcholine bind…	0.001	0.001	R-HSA-181431
CHEMBL1237082	P30926	Postsynaptic nicot…	0.001	0.001	R-HSA-622327

FIGURE 13 OUTPUT OF THE PATHWAY PART

A connection between a compound and a pathway was considered when the FDR was <0,05. In order to obtain clearer results a binary table was made, where the individual FDR score was either replaced with a 1 (if FDR<0,05) or 0 (if FDR >0,05). On the basis of the output table (Figure 14), a heatmap was created by using R.



S molecule_...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...	D R-HSA-...
CHEMBL1230309	1	1	1	1	1	1	1	1	1	1	1	1
CHEMBL1237082	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL293391	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL3	0	0	0	0	1	1	0	0	0	0	0	0
CHEMBL309804	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL365956	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL406819	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL430497	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL510275	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL6623	0	0	0	0	0	0	0	0	0	0	0	0
CHEMBL70141	0	0	0	0	1	1	1	1	1	1	1	1
CHEMBL95367	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 14 COMPOUNDS AND THEIR ASSOCIATED PATHWAYS

5.2. Results

5.2.1. *Compound-pathway interaction fingerprint*

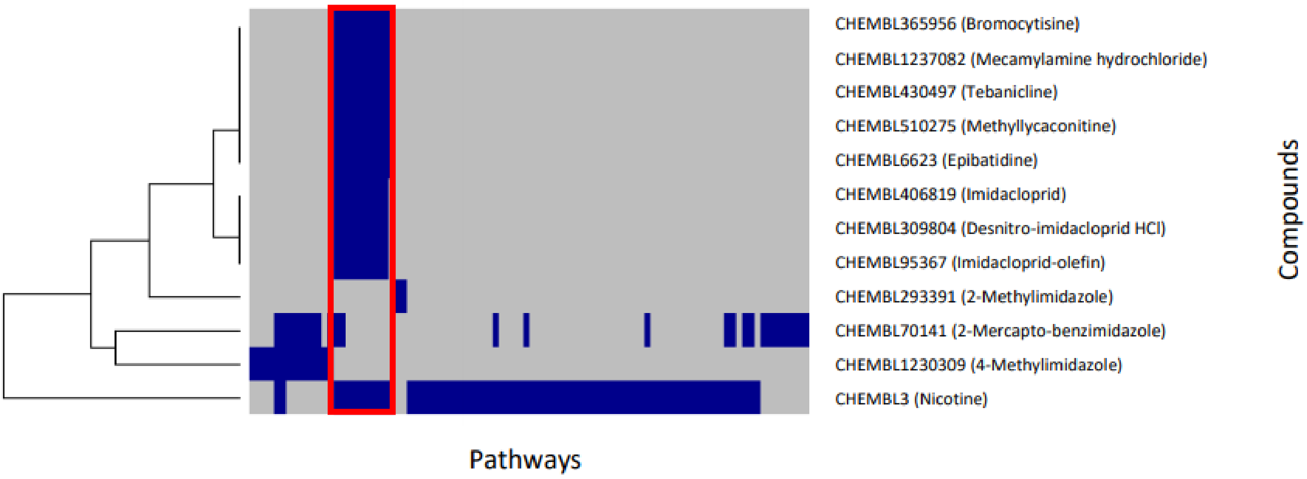


FIGURE 15 COMPOUND-PATHWAY FINGERPRINT (RED FRAME HIGHLIGHTS THE NEURONAL SYSTEM)

The heatmap represents the pathways on the x-axis and the compounds of the CS-14 on the y-axis. A valid ChEMBL-ID of the compounds and an UniProt-ID of their

connected target protein had to be present for a successful analysis. Only 12 compounds out of 29 were successfully connected to a pathway.
In total, 92 different pathways were discovered with which the examined 12 compounds seem to interact with.

The blue colored fields in the heatmap illustrate a connection between compound and pathway and the grey area assumes that there is no connection. Analyzing the heatmap shows that clusters were built among the compounds. Bromocytisine, Mecamylamine, Tebanicline, Epibatidine build a cluster by following the exact same pathways which all attack points in the neuronal system. Similar interaction patterns are as well shown among the compounds Imidacloprid, Desnitroimidacloprid and Imidacloprid-olefin.

The red frame in the heatmap highlights the pathways that are related to the neuronal system, which is seen to be mostly affected by the compounds. It is evident, that nearly every out of the 12 analyzed compounds (except 2-Methylimidazole and 4-Methylimidazole) follows at least one pathway that is connected to the neuronal system.

Nicotine seems to modulate many different pathways in the human body including metabolism, disease pathways, transport of small molecules, and neuronal pathways. Besides that, 2-Mercaptobenzimidazole is also involved in various pathways other than the neuronal system, including signal transduction pathways, apoptosis, and metabolism.

Contrary to the other examined compounds, 2-Methylimidazole and 4-Methylimidazole do not affect the neuronal system and both follow different pathways despite their structural similarity. Whereas 2-Methylimidazole is involved in the metabolism of RNA, 4-Methylimidazole interacts with metabolism, transport of small molecules, and signal transduction pathways.

Considering the fact that neonicotinoids may not only interact with the insect but also the mammalian nAChR, it is necessary to know which pathways are activated by binding to the nAChRs. Regarding the expression in the brain, the most relevant subtypes of this receptor are α7, α3β4 and α4β2. [6] Therefore only for these three subtypes it has been analyzed which pathways they are belonging to in the human body. For this, the ChEMBL-IDs of the different subtypes of the nAChR were converted into their UniProt-IDs by using the Retrieve/ ID mapping tool on UniProtKB. (Table 4) Afterwards the pathways that are modulated by those receptor subtypes were analyzed.

ChEMBL ID	UniProt ID	Protein Name
CHEMBL1907594	P30926	Neuronal acetylcholine receptor; alpha3/beta4
CHEMBL2492	P36544	Neuronal acetylcholine receptor protein alpha-7 subunit
CHEMBL1907589	No entry	Neuronal acetylcholine receptor; alpha4/beta2

TABLE 4 OVERVIEW OF THE NACHR-SUBTYPES INCLUDING THEIR NAME, CHEMBL-ID AND UNIPROT-ID

As shown in Table 4, no entry in the UniProt database is available for the nAChR-subtype α4β2, thus no pathway connection can be made. For the other two subtypes (α7, α3β4) was it possible to track the pathways by putting their UniProt-ID into the analyzing tool of the Reactome database. Both subtypes are predominantly linked to the neuronal system, the exact pathway names are shown in Table 5.

Pathway names	nAChR-subtype
Neuronal system	α7, α3β4
Highly calcium permeable postsynaptic nicotinic acetylcholine receptors	α7, α3β4
Acetylcholine binding and downstream events	α7, α3β4
Postsynaptic nicotinic acetylcholine receptors	α7, α3β4
Neurotransmitter receptors and postsynaptic signal transmission	α7, α3β4
Transmission across Chemical Synapses	α7, α3β4
Highly sodium permeable postsynaptic acetylcholine nicotinic receptors	α3β4

Highly calcium permeable nicotinic acetylcholine receptors	α3β4
Presynaptic nicotinic acetylcholine receptors	α3β4
Neutrophil degranulation (immune system)	α3β4

TABLE 5 PATHWAYS CONNECTED TO THE NACHR SUBTYPES (RED COLORED: NEURONAL PATHWAYS, BLUE COLORED: PART OF IMMUNE SYSTEM)

The pathways that seem to be modulated by the nAChR subtypes correlate to the pathways that were connected to the majority of the CS-14 compounds. Mostly the pathways in the neuronal system are affected. The red colored pathways present the ways that are regulated by the α7 and/or α3β4 subtypes of the nAChR. On the right column of Table 6 the compounds are listed for which a connection with the pathway were assumed.

Pathways	**Compounds**
Neuronal system	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, 2-Mercaptobenzimidazole
Highly calcium permeable postsynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid
Acetylcholine binding and downstream events	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid
Postsynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid
Neurotransmitter receptors and postsynaptic signal transmission	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, 2-Mercaptobenzimidazole

Transmission across Chemical Synapses	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, 2-Mercaptobenzimidazole
Highly sodium permeable postsynaptic acetylcholine nicotinic receptors	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine
Highly calcium permeable nicotinic acetylcholine receptors	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine
Presynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Imidacloprid-oleofin, Imidacloprid, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine
Neurotransmitter clearance	Nicotine
Neurotransmitter release cycle	Nicotine
Norepinephrine Neurotransmitter Release Cycle	Nicotine
Enzymatic degradation of Dopamine by monoamine oxidase	Nicotine
Metabolism of serotonin	Nicotine
Serotonin clearance from the synaptic cleft	Nicotine
Enzymatic degradation of dopamine by COMT	Nicotine
Dopamine clearance from the synaptic cleft	Nicotine
Acetylcholine Neurotransmitter Release Cycle	Nicotine
Activation of AMPK downstream of NMDARs	2-Mercaptobenzimidazole
Post NMDA receptor activation events	2-Mercaptobenzimidazole
Activation of NMDA receptors and postsynaptic events	2-Mercaptobenzimidazole
Metabolism	Nicotine, 2-Mercaptobenzimidazole, 4-Methylimidazole
Transport of small molecules	Nicotine, 4-Methylimidazole

Disease Pathways	Nicotine
Apoptosis	2-Mercaptobenzimidazole
Metabolism of RNA pathways	2-Methylimidazole
Immune system	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine
Signal Transduction pathways	2-Mercaptobenzimidazole, 4-Methylimidazole

TABLE 6 PATHWAYS CONNECTED TO THE COMPOUNDS (WORDS IN BOLD INDICATE MAIN PATHWAYS)

The results in Table 6 were visualized into an interactive overview of the different pathways (Figure 16). The overrepresentation analysis shows that three major pathway systems were mainly connected to the examined compounds. These include the neuronal system, metabolism, and the transport of small molecules. The yellow colour indicates the p-value of a pathway, the brighter the colour the smaller the p-value. A small p-value supports the credibility that a connection between the compound and the pathway can be assumed.

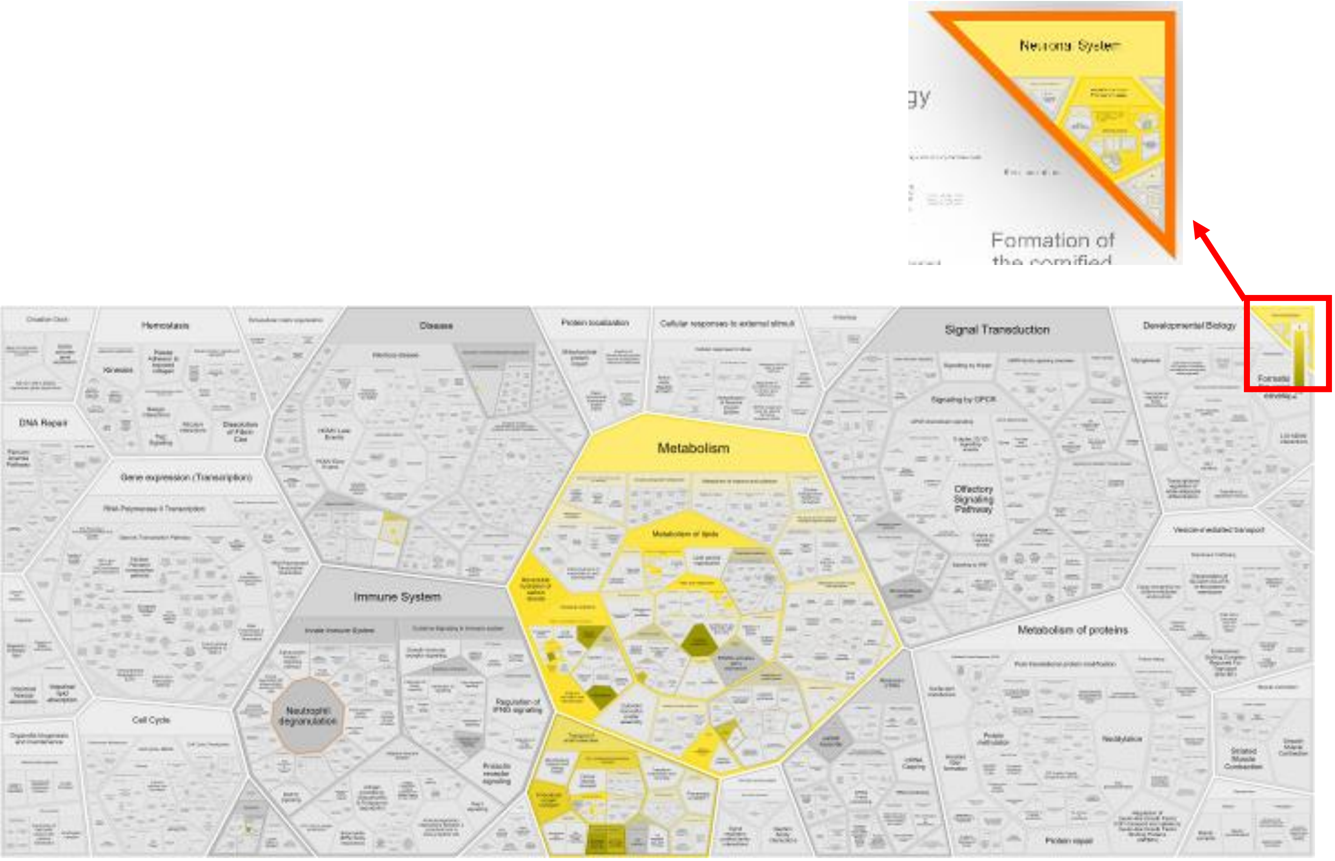


FIGURE 16 *PATHWAY OVERVIEW (ACCESSED THROUGH REACTOME) – NEURONAL SYSTEM IS FRAMED IN RED*

5.2.2. Interactome-pathway fingerprint

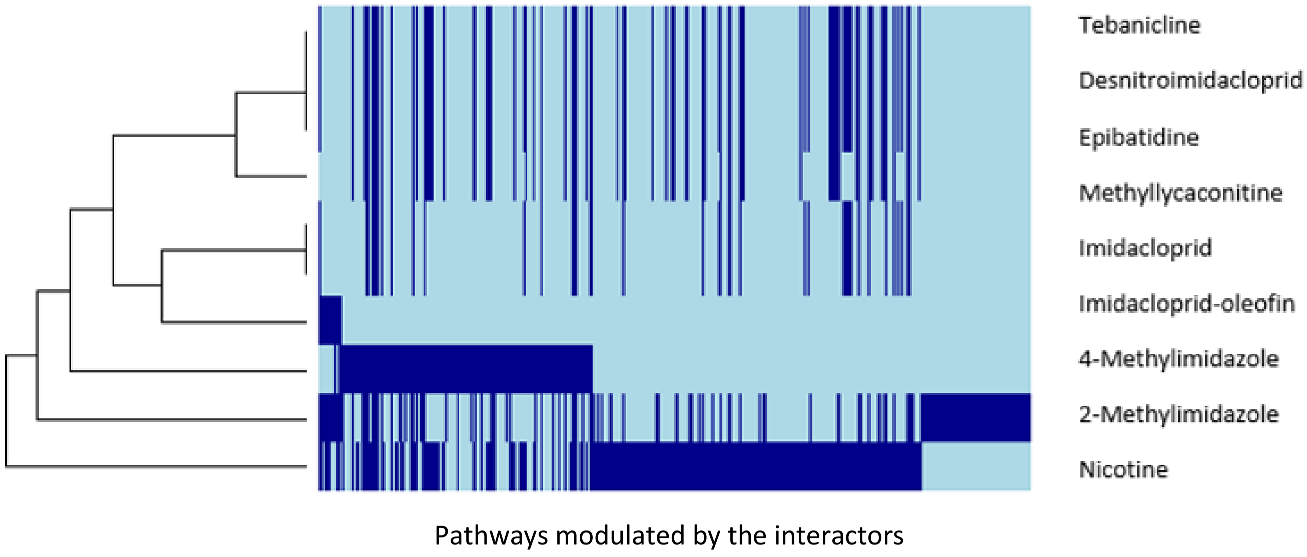


FIGURE 17 INTERACTOME-PATHWAY FINGERPRINT

The heatmap shown in Figure 17 represents the target protein-protein interactions and which pathways the interactors follow in the human body. On the x-axis the pathways that are modulated by the interactors of the target proteins are shown, and on the y-axis are the names of the examined compounds.

For only 10 compounds, out of the 29 of the CS-14, valid interactors were found, but nevertheless 624 different pathways that are connected to the interactors were obtained. The heatmap shows that Tebanicline, Desnitroimidacloprid and Epibatidine build a cluster due to the same pathways that seem to be followed by their interactors. Regardless, the received pathways still scatter over various different systems of the body including metabolism of RNA, signal transduction, gene expression, apoptosis, neuronal system and more. Thus, the results are hard to analyze and leads to blurry assumptions. The visualization below (Figure 18) represents the large number of different pathways that have been obtained.

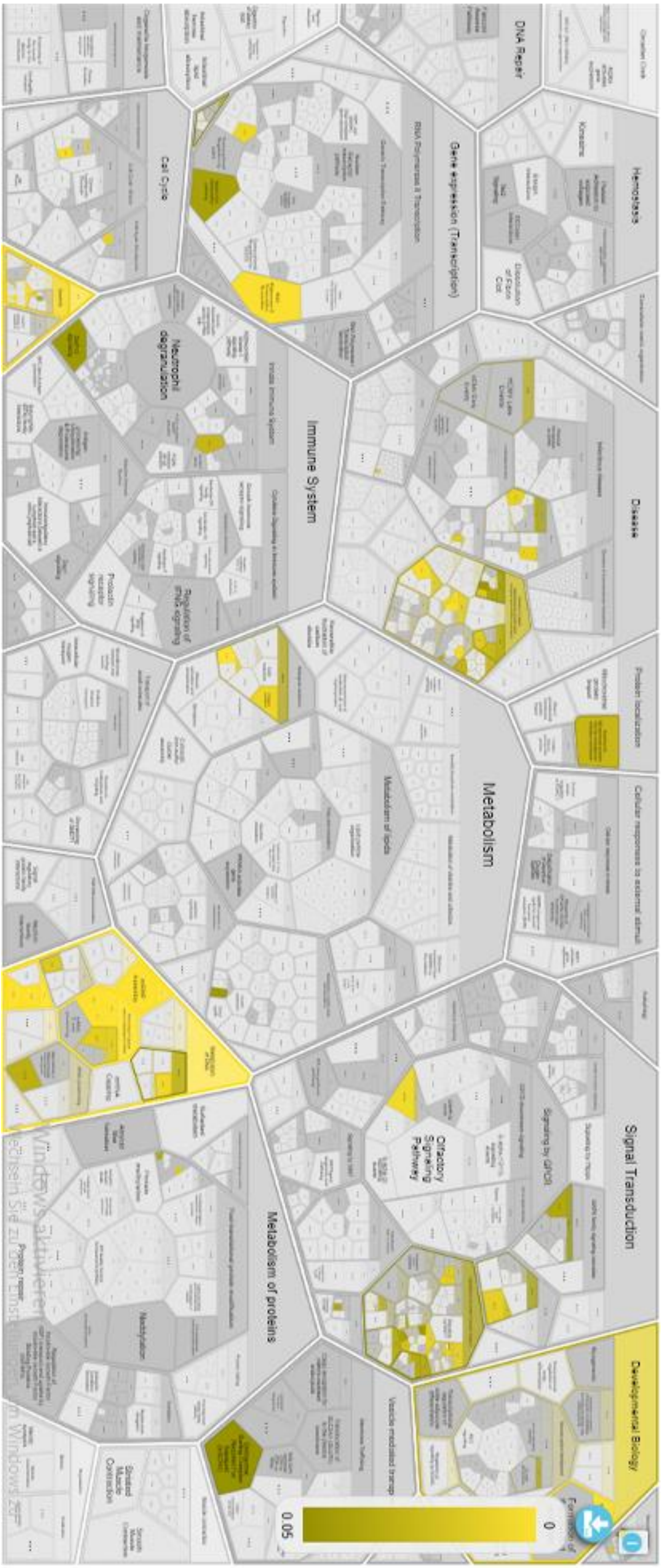


FIGURE 18 PATHWAY OVERVIEW (ACCESSED THROUGH REACTOME)

6. Compound-pathway interaction fingerprint (with the use of ToxPHACTS)

For this compound-pathway interaction fingerprint, the targets of the CS-14 compounds were obtained by using the software ToxPHACTS in order to illustrate if the results overlap with the previously created pathway-interaction fingerprint.

Two different approaches were considered while creating the compound-pathway interaction fingerprints with the use of ToxPHACTS. One fingerprint was created with the targets of the compounds and their similar compounds, while the other one was made by excluding the similar compounds, so that only the direct targets of the origin compounds of the CS-14 were analyzed. The p-ChEMBL value for the retrieved targets needed to be over 5 in order to be included in the analysis.

6.1. Data preparation

First step was to define every compound of the CS-14 through its SMILES code which reflects its chemical structure. The SMILES code was used as an input in order to obtain the structure in ToxPHACTS, so that the program can make a similarity research which offers all the compounds that are similar to the origin compound (similarity must have a threshold of 0.5). For the received similar compounds, information about the ChEMBL-ID, chemical structure of the compound, the mean similarity and if bioactivity has been found or not, are given.

In the last step, the pharmacology of the compounds is provided, including their hitted targets, their p-ChEMBL value and if the targets are associated with toxicity. These reports were then downloaded for every successfully found compound. After downloading the reports, the ChEMBL-IDs of the retrieved targets were translated into UniProt-IDs. Afterwards a table was manually created where the UniProt-IDs of the targets were connected to the ChEMBL-ID with its interacting compound, this last step is necessary in order to be able to analyze the compound-target pairs with the previous used ‘Pathway WF’ on KNIME and to create a fingerprint that is related to the compounds and not to the targets.

6.2. Results – ToxPHACTS

6.2.1. First approach: Compound-pathway fingerprint (CS-14 compounds <u>and</u> their similar compounds)

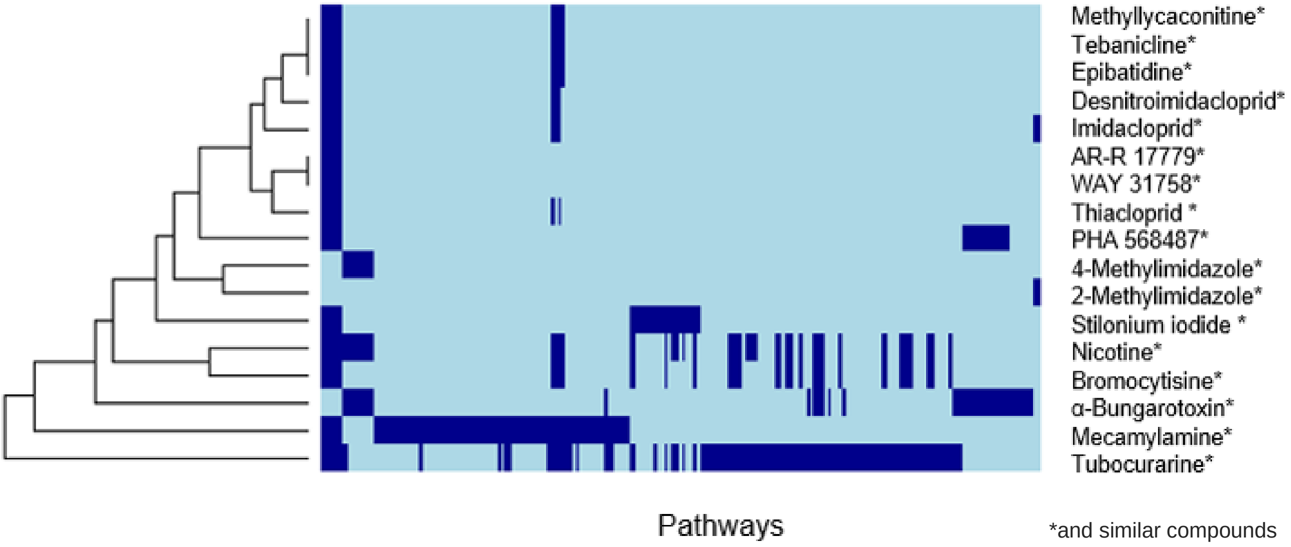


FIGURE 19 COMPOUND-PATHWAY INTERACTION FINGERPRINT WITH THE CS-14 COMPOUNDS AND THEIR SIMILAR COMPOUNDS

The generated fingerprint is based on the targets that interact with the origin compounds of the CS-14 and as well with their similar compounds. Due to the fact that the similar substances were also included in the fingerprint, it was possible to examine 17 compounds out of 29. A broad range of various pathways was the outcome of the analysis and no significant interaction patterns could be found among the analyzed compounds. The compound-target pairs that have been retrieved through ToxPHACTS led to 203 different pathways which are presented on the x-axis. On the y-axis are the origin CS-14 compounds listed. In order to keep an overview the similar compounds were connected to the ChEMBL-ID of the 'parent' compound, thus every row shows the pathways with which the compounds of the CS-14 and their similar compounds seem to interact with.

Metabolism, disease pathways, immune system, signal transduction, neuronal system, cell cycle, and immune system are all attack points of the analyzed compounds. Due to the large number of different pathways, it is difficult to analyze the interaction in detail. Figure 20 shows an overrepresentation analysis of the scattered pathways. The neuronal pathway is highlighted in red and zoomed in.

FIGURE 20 PATHWAY OVERVIEW (ACCESSED THROUGH REACTOME)

6.2.2. Second Approach: Compound-pathway fingerprint (only CS-14 compounds used, excluding similar compounds)

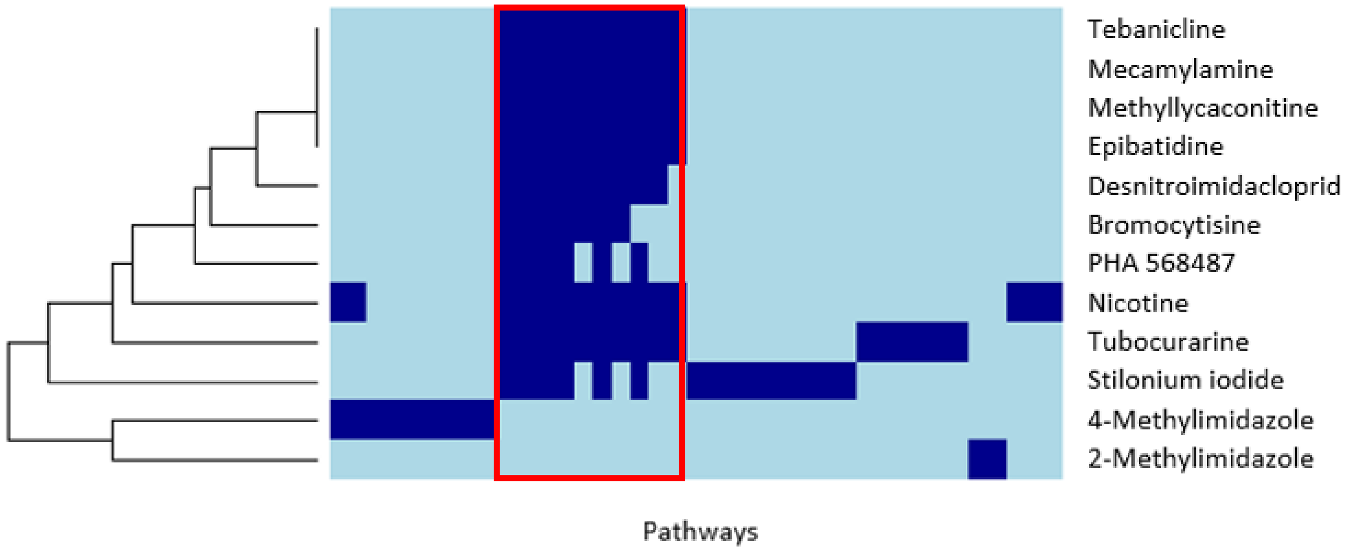


FIGURE 21 COMPOUND-PATHWAY FINGERPRINT (ONLY CS 14-COMPOUNDS) - NEURONAL WAYS ARE FRAMED IN RED

The fingerprint was created on the basis of the direct targets of the compounds of the CS-14 that were obtained with ToxPHACTS, but this time the similar compounds are excluded to get clearer results. The p-ChEMBL for the targets needed to be above 5 as well.

The x-axis marks the pathways and the y-axis the compounds for which targets have been retrieved through ToxPHACTS. It was possible to receive targets with a pChEMBL value over 5 for 12 compounds out of 29. The analysis of the obtained compound-target pairs led to 39 different pathways and clustering can be seen among Tebanicline, Mecamylamine, Methyllycaconitine and Epibatidine which show the same interaction pattern.

The biggest focus lays on the pathways of the neuronal system (framed in red), with which most of the retrieved targets seem to interact with. Only two substances of the examined compounds are not involved in the neuronal pathways, these were 2-Methylimidazole and 4-Methylimidazole which mainly seem to affect the metabolism.

On the right column of Table 7 are the compounds listed that are interacting with the pathways mentioned on the left. Red colored fields indicate the neuronal pathways that are modulated by the nAChRs. Nicotine, Epibatidine, Tebanicline, Methyllycaconitine, Mecamylamine, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide and Tubocurarine follow the same neuronal attack points as known from the nAChRs.

Pathways	**Compounds**
Neuronal system	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Highly calcium permeable postsynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Tubocurarine
Acetylcholine binding and downstream events	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Postsynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Neurotransmitter receptors and postsynaptic signal transmission	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Transmission across Chemical Synapses	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Highly sodium permeable postsynaptic acetylcholine nicotinic receptors	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Tubocurarine
Highly calcium permeable nicotinic acetylcholine receptors	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine

	HCl, Desnitroimidacloprid, PHA 568487, Tubocurarine
Presynaptic nicotinic acetylcholine receptors	Nicotine, Epibatidine, Tebanicline, Methyllycaconitine citrate salt, Mecamylamine HCl, Bromocytisine, Desnitroimidacloprid, PHA 568487, Stilonium iodide, Tubocurarine
Metabolism	4-Methylimidazole, Nicotine
Metabolism of RNA pathways	Stilonium iodide, 2-Methylimidazole
Disease pathways	Stilonium iodide
Transport of small molecules	Tubocurarine
Immune system pathways	Desnitroimidacloprid, Nicotine, Tubocurarine, Methyllycaconitine citrate salt, Mecamylamine HCl, Epibatidine

TABLE 7 PATHWAYS CONNECTED TO THE COMPOUNDS (WORDS IN BOLD INDICATE MAIN PATHWAYS)

The overrepresentation analysis of the pathways in Figure 22 shows that mainly the neuronal system is attacked by the compounds, otherwise only a few pathways in the field of metabolism, disease pathways, signal transduction, and immune system light up.

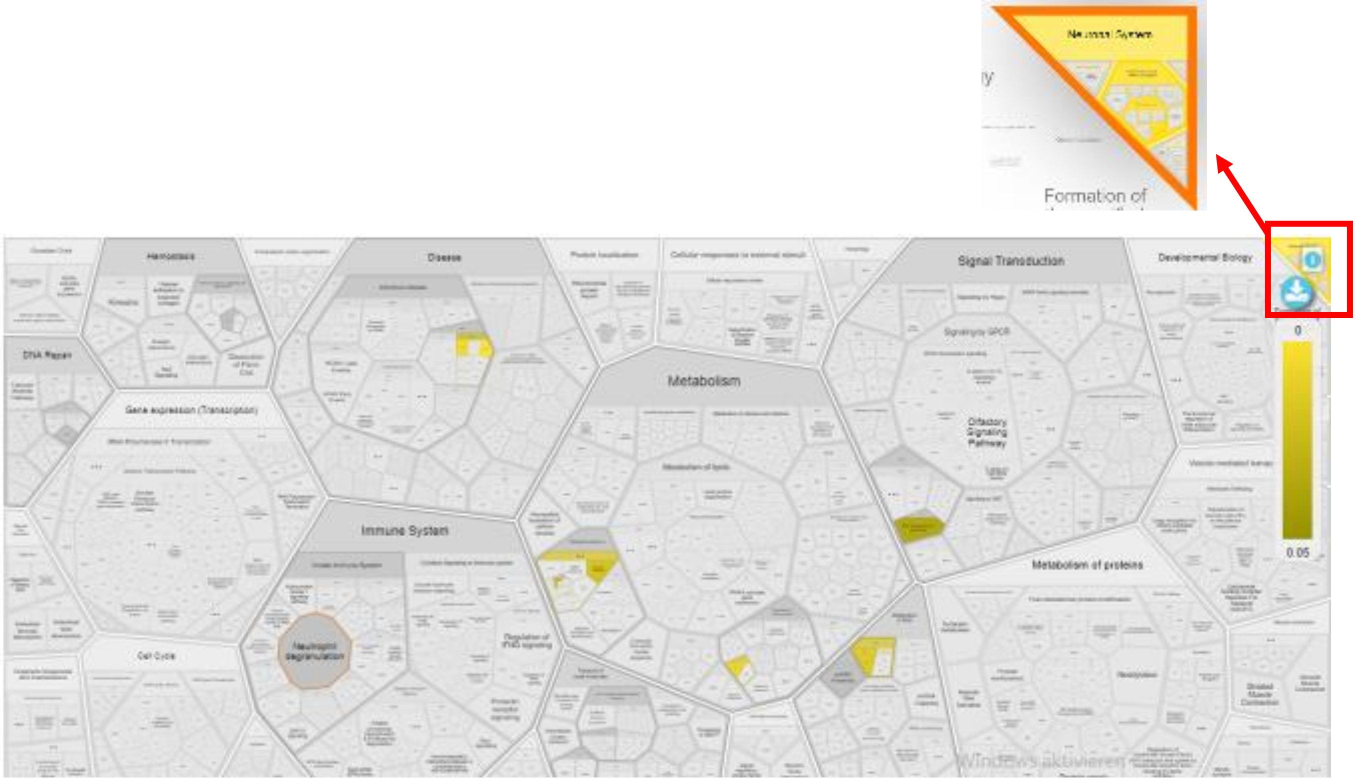


FIGURE 22 PATHWAY OVERVIEW (ACCESSED THROUGH REACTOME) – NEURONAL SYSTEM IS FRAMED IN RED

7. Discussion of the Results

First and foremost, this thesis based on the compounds of the CS-14 which are associated with potential neurotoxic effects and the aim was to illustrate if these adverse effects could be predicted in the human body by creating fingerprints with several approaches. Comparing and analyzing the different fingerprints shows that this computational approach has a huge potential for toxicity prediction, but in fact there are still some limitations in this area.

Considering the compound-pathway interaction fingerprints a valid ChEMBL-ID of the compounds needs to be available, so if there is no entry in ChEMBL, the substance cannot be analyzed. For example entries were missing for following CS-14 compounds: dibutyl-tin dichloride, butyl-tin trichloride and tributyl-tin chloride. Furthermore, not for every target protein a valid UniProt-ID is available which is a requirement for creating the compound-pathway fingerprint with the WF on KNIME. Due to this fact, not the full spectrum of target-interactions of a compound can be analyzed. Besides, there are no given information of the relationship of the compound and the target, so the mode of action (agonist or antagonist) stays unclear.

The approach where the interactome of the target proteins were included, led to inconclusive results, because the interactors trigger too many different pathways. Similar were the quality of the results regarding the compound-pathway fingerprint – created with the use of ToxPHACTS - where the targets of the CS-14 compounds and their similar compounds were included. Even though the pChEMBL value for the targets was set to > 5, there were still 203 different pathways found, including disease pathways, metabolism of proteins, immune system, gene expression, cell death, cell cycle, neuronal system, and more.

The other two compound-pathway fingerprints, where only the direct targets of the compounds of the CS-14 were examined – for one the targets were received with the KNIME WF and for the other the targets were obtained with ToxPHACTS – led to more significant results. The pathway predictions of the compounds were mostly overlapping in both fingerprints.

2-Methylimidazole and 4-Methylimidazole were in neither of any fingerprints involved in the neuronal system. Even though the only structural difference of the two compounds is that the methyl group is in a different position, they still seem to bind to different targets. 2-Methylimidazole attacks the survival motor neuron protein and 4-Methylimidazole the aldehyde dehydrogenase 1A1, which results in different pathways.

The CS-14 compounds Bromocytisine, Mecamylamine HCl, Tebanicline, Methyllycaconitine, Epibatidine, Desnitroimidacloprid and Nicotine, were examined in both approaches and exhibit a connection with the same pathways of the neuronal system that are, as well, modulated by the α7 and α3β4 nAChRs-subtypes.

The compounds Imidacloprid, Imidacloprid-olefin and 2-Mercaptobenzimidazole only could be investigated by the approach with the KNIME WF, because the WF not only searches through ChEMBL but also through other databases like IUPHAR and DrugBank. Whereas the compounds PHA 568487, Tubocurarine, and Stilonium iodide were only found through the targets retrieved with ToxPHACTS. Although ToxPHACTS also searches through ChEMBL, the difference is that the structure of the compounds can be used as an input which then derives the available ChEMBL-IDs for the compounds. It is also possible that for one substance more than one ChEMBL-ID is available, as for example nicotine has two different identifiers because of its stereoisomerism. So contrary to ToxPHACTS the KNIME WF already needs the ChEMBL-ID as an input. In order to receive the best results all available ChEMBL-IDs for one compound should be used.

Another approach for finding similar interaction patterns or cluster among the compounds was to create a compound-pharmacophore fingerprint. The advantage of this approach is that every compound of the CS-14 could be analyzed. But the results of this model were not meaningful, regarding the different amounts of pharmacophore models that were available for each protein. For example, for the protein CAH2 over 250 pharmacophore models were created whereas for the enzyme CP1A2 only one model was generated. Also keeping in mind, 39 target proteins were found to interact with the CS-14 compounds but only for 14 proteins an entry in the PDB was found.

Besides that, no clear interaction pattern or cluster have been found, so no predictions could be made.

Table 8 lists the compounds where a connection to the neuronal system is assumed or not based on the results of this thesis.

Compounds of the CS-14 which are **involved in the pathways of the neuronal system:**	• Nicotine • Epibatidine • Tebanicline • Methyllycaconitine citrate salt • Mecamylamine HCl • Bromocytisine • Desnitroimidacloprid • PHA 568487 • Stilonium iodide • Tubocurarine • Imidacloprid-oleofin • Imidacloprid • 2-Mercaptobenzimidazole
Compounds that were **not involved in any neuronal pathways:**	• 2-Methylimidazole • 4-Methylimidazole
Compounds that **could not be successfully included in the analysis:**	• AR-R 17779 HCl • Acetamiprid

	• 2-Imidazolidinthione • WAY 317538 • Dihydro-beta-erythroidine hydrobromide • Clothianidin • Dinotefuran • Thiamethoxam • Thiacloprid • α-Bungarotoxin • Tributyl-tin chloride • Butyl-tin trichloride • Dibutyl-tin dichloride • Butanone oxime

TABLE 8 COMPOUND PREDICTIONS

8. Summary and Outlook

Within this thesis, a fraction of potential neurotoxic compounds was analyzed through different pathway-interaction fingerprints and a pharmacophore-based approach. The results from the compound-pharmacophore fingerprint were not informative. One reason for this might be that for some target proteins no entry in the PDB was available. Also for some targets only one ligand-protein complex was found, whereas well examined proteins obtained more than 100 different complexes.

The outcome of the compound-pathway interaction fingerprint delivered clearer results provided the fingerprints were only created with the direct targets of the compounds. Including the interactors or the similar compounds led to blurry results.

Considering the compound-pathway interaction fingerprints there are still some limitations like:

- lack of data (missing entries in databases such as UniProt or ChEMBL)
- data situation (many data are scattered over several databases and websites which makes it difficult to get access to all the data)
- missing experimental values (such as bioactivity values)
- the exact knowledge of the pathways which lead to developmental neurotoxicity, for example a validated list of target proteins which are involved in this toxic endpoint could be helpful for predictions

If these limitations are overcome, the pathway interaction approach could prove to be a valid tool for investigating different toxic risks - such as developmental neurotoxicity – of new drugs.

List of Abbreviations

ADMET	Absorption, Distribution, Metabolism, Excretion, Toxicity
AOFB	Amine oxidase [flavin-containing] B
CAH2	Carbonic anhydrase 2
ChEMBL-ID	ChEMBL-Identifier
CLAT	Choline O-acetyltransferase
CP1A2	Cytochrome P450 1A2
CP2A6	Cytochrome P450 2A6
CP2AD	Cytochrome P450 2A13
CP2B6	Cytochrome P450 2B6
CS-14	Case Study-14
FDR	False Discovery Rate
nAChR	Nicotinic Acetylcholine Receptor
PDB	Protein Data Bank
UniProt-ID	UniProt-Identifier
WF	Workflow (in KNIME)

9. Appendix

9.1. FIGURES

9.2. TABLE

9.3. Bibliography

[1] D. Pietrzak, J. Kania, E. Kmiecik, G. Malina und K. Wator, „Fate of selected neonicotinoid insecticides in soilewater systems:Current state of the art and knowledge gaps," *Elsevier,* pp. 1-3, 2020.

[2] T. Marrs, „Mammalian Toxicology of Insecticides," pp. 8-9, 184-186, 2012.

[3] C. Bass und L. Field, „Neonicotinoids," pp. 772-773, 2018.

[4] Q. Zhang, Z. Li, C. Chang, J. Lou, M. Zhao und C. Lu, „Potential human exposures to neonicotinoid insecticides: A review," *Elsevier,* pp. 71-81, 2017.

[5] L. Smirnova, H. Hogberg, M. Leist und T. Hartung, „Developmental Neurotoxicity – Challenges in the 21st Century and In Vitro Opportunities," pp. 129-130, 2014.

[6] A. Nicke, S. Wonnacott und R. J. Lewis, „a-Conotoxins as tools for the elucidation of structure and functionof neuronal nicotinic acetylcholine receptor subtypes," pp. 2305-2319, 2004.

[7] E. P. o. P. P. P. a. t. R. (PPR), „Scientific Opinion on the developmental neurotoxicity potential of acetamiprid and imidacloprid.," *EFSA Journal,* pp. 12-14, 2013.

[8] A. Cimino, A. Boyles, K. Thayer und M. Perry, „Effects of Neonicotinoid Pesticide Exposure on Human Health: A Systematic Review," pp. 155-156, 2017.

[9] K.-K. Junko, K. Yukari , K. Yoichiro , H. Masaharu und K. Hitoshi , „Nicotine-Like Effects of the Neonicotinoid Insecticides Acetamiprid and Imidacloprid on Cerebellar Neurons from Neonatal Rats," *PlosOne,* pp. 1-5, 2012.

[10] „EU-TOXRISK," [Online]. Available: https://www.eu-toxrisk.eu/page/en/about-eu-toxrisk.php. [Zugriff am 11 Jänner 2021].

[11] „ToxPHACTS," Phenaris, 2020. [Online]. Available: https://toxphacts.univie.ac.at/. [Zugriff am 13 Jänner 2020].

[12] J. Cleve und U. Lämmel, „Data Mining," pp. 21-22, 2020.

[13] M. R. Berthold, N. Cebron, F. Dill, T. Gabriel und T. Kötter , „KNIME – The Konstanz Information Miner," pp. 26-31, 2007.

[14] A. Gaulton, L. Bellis, A. P. Bento, J. Chambers und M. Davies , „ChEMBL: a large-scale bioactivity database for drug discovery," pp. 1100-1107, 2011.

[15] A. Gaulton, A. Hersey und M. Nowotka , „The ChEMBL database in 2017," pp. 945-947, 2017.

[16] R. Hoffmann, A. Gohier und P. Pospisil, „Data Mining in Drug Discovery," pp. 31-33, 2013.

[17] G. Wolber und T. Langer, „LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters," pp. 160-169, 2005.

[18] „LigandScout User Manual,“ p. 143, 20 Jänner 2021.

[19] w. consortium, „Protein Data Bank: the single global archive for 3D macromolecular structure data,“ pp. 520-528, 2018.

[20] H. Berman, K. Henrick, H. Nakamura und J. Markley, „The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data,“ pp. 301-303, 2006.

[21] T. U. Consortium, „UniProt: a hub for protein information,“ pp. 204-212, 2015.

[22] M. L. Benson, R. D. Smith, N. A. Khazanov, B. Dimcheff, J. Beaver, P. Dresslar und H. A. Carlson, „Binding MOAD, a high-quality protein–ligand database,“ pp. 674-678, 2008.

[23] S. Kim, P. Thiessen, T. Cheng, J. Zhang, A. Gindulyte und E. Bolton, „PUG-View: programmatic access to chemical annotations integrated in PubChem,“ pp. 1-11, 2019.

[24] D. Croft, G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N. Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio und L. Stein, „Reactome: a database of reactions, pathways and biological processes,“ pp. 691-697, 2010.

[25] G. Alterovitz und M. Ramoni, „Knowledge based bioinformatics: from analysis to interpretation,“ pp. 293-296, 2010.

[26] R. Team, „RStudio, new open-source IDE for R,“ 2011. [Online]. Available: RStudio, new open-source IDE for R. [Zugriff am 17. Jänner 2021].

[27] N. Chadha und O. Silakari, „Active site fingerprinting and pharmacophore screening strategies for the identification of dual inhibitors of protein kinase C (PKCβ) and poly (ADP-ribose) polymerase-1 (PARP-1),“ pp. 747-761, 2016.

[28] H. Ye, K. Tang, L. Yang, Z. Cao und Y. Li, „Study of drug function based on similarity of pathway fingerprint,“ pp. 132-139, 2012.