Chapter 4

In-silico analysis of target prediction and gene interactions of agrochemicals

4.1 Introduction

The first step of research often begins in academia, where a hypothesis is generated; like the inhibition or induction of a protein or pathway as a curative outcome in a toxic condition. Indeed, a crucial point of the research process is the selection of a target, which can be a range of biological entities such as proteins, RNA, and genes that can be selected via bioinformatics analyses (Cava et al., 2016). An optimal target must be accessible to the assumed toxic molecule and the binding toxic–target complex should induce a biological response (Hughes et al., 2011), which can be quantified with *in vitro* models. The binding affinity between the toxic and the target can be calculated *In silico* with molecular docking. Thus, *In silico* and *in vitro* screenings help to quickly identify the toxicity of the tested molecules, so avoiding further steps such as in vivo studies(Cava & Castiglioni, 2020).

In silico approaches with docking studies require at least two elements: a protein database and a molecular docking algorithm. The rapidly increasing number of structures has created big data, which offer a wide range of biological and chemical information and are a recent opportunity to develop better knowledge of the relationships between toxic substances and usually targets proteins (Meng et al., 2011). Although the available data are often heterogeneous and incomplete, computational methods can exploit this knowledge to deepen these interactions (Gligorijevic & Przulj, 2016). Given the cost and time consumption of experimental methods, high-performing computational algorithms are needed. The computational technique known as "docking" predicts the binding of a toxic molecule–protein complexes, as well

as the conformation of the ligand upon binding to a protein target (Ferreira et al., 2015). The binding free energy of interactions establishes the affinity of an association and the conditions for forming a complex (Bolnykh et al., 2021). Ranked binding free energies are not always precise, but they can be used to understand and predict the molecular mechanism by experimentally testing by a virtual screening approach (Berry et al., 2015). In addition, molecular docking can be also used for predicting the effects of a toxic substance in identifying an undesired interaction between a compound and off-targets. To date, 57,000 abstracts/papers have been published on molecular docking, indicating the importance of this computational method (Cava & Castiglioni, 2020; Ferreira et al., 2015).

Despite encouraging results, the real condition of the cellular environment, such as the pH and temperature, cannot be fully replicated in a docking study. Each docking algorithm has its limitations and advantages. Therefore, it has been reported that binding free energy that integrates the results from different docking algorithms can lead to higher performance in a virtual screening process (Pinzi & Rastelli, 2019). Moreover, molecular docking, being a structure-based method, is limited to receptors and ligands with a known stable structure. Thus, the integration of *in vitro* and *in vivo* studies as a validation step of *In silico* methods is an indispensable part of xenobiotic interaction (Meng et al., 2011).

Namely two major classes of methods for computational gene prediction are being found. One is the sequence similarity searches based on the gene sequence, and the other is gene structure and signal-based searches, which is also referred to as ab initio gene finding. The prediction of biological targets of any xenobiotic having materials that are machine-related can be exclusively performed by computational target prediction tools (Zhuo Wang et al., 2004). Lab-based work and computational approaches are used to envisage biological targets that interact with a toxicant. It has been agreed upon by

various research groups that experimental-based methods are usually costlier and slower than computational approaches, as computational approaches make predictions based on set models and algorithms with several approximations (Schomburg et al., 2014).

In general, the data on computational target prediction approaches fall into two major categories of target-based methods (also called structure-based or receptor-based) and ligand-based methods (Batool et al., 2019). Ligand-based methods incorporate chemical structures to predict targets (Schenone et al., 2013). Hence, the chemical similarity criteria for bioactive molecules play key roles in ligand-based modelling (Wang et al., 2016) . Target-based methods rely on three-dimensional (3D) receptor structures to predict receptor–toxicant interactions (Forouzesh et al., 2019; Haupt & Schroeder, 2011). With regards to the speediness, ligand-based methods tend to be faster, while target-based methods take substantially more computational resources for a docking run against hundreds, or even thousands, of targets still not achieving reliable results (Koutsoukas et al., 2011).

The prediction of biological targets of molecules like toxicants, xenobiotic or bioactive compounds from machine-readable materials can be routinely performed by computational target prediction tools (CTPTs). However, the prediction of biological targets of toxicants from non-digital materials (e.g., printed or handwritten documents) has not been possible due to the complex nature of the compound which leads to unapproachable computations. However, with many advances over the last decades, computational target prediction remains a very challenging task, as reflected by the low experimental target validation success rate (Thafar et al., 2019). The removal of false positives reduces the risk of yielding predictions that could incorrectly affect the downstream experiments for drug and pesticide discovery (Forouzesh et al., 2019; Zhonghua Wang et al., 2016).

Many attempt(s) has been made to improve the target prediction success leading to the creation of an innovative method based on chemical similarity. The recently developed methods of prediction have five distinctive advantages: First, no statistical method is used in the prediction. Second, accuracy is high. Third, it can be used appropriately without similarity calculations in non-digital materials and with similarity calculations (perfect similarity) in machine-readable materials. Fourth, deeper insights into understanding the interactions happening between a molecule of interest and the target can be extrapolated. Fifth, it requires basic knowledge for high-performance computing techniques or algorithms that do not stop its implementation part (Thafar et al., 2019; Mathai & Kirchmair, 2020).

With the great evolution of genetic/molecular biology, several high-throughput profiling technologies such as genomics, transcriptomics, proteomics, and metabolomics have been developed, among which toxicogenomics combines toxicology with these approaches to analyze the gene expression profile of several thousand genes aiming to identify changes associated with xenobiotics-induced toxicities (Cui & Paules, 2010). In the last two decades, epigenetic alterations have been shown to play a role in the transcriptional processes that regulate gene expression. Therefore, the field of toxicogenomics coupled with *in-silico*, which studies the relationship between epigenetic modifications and disease status in response to exposure to environmental contaminants and toxic agents, is now at the forefront of environmental health science.

As we wanted to understand the different possible targets apart from the candidate gene profiles performed in previous chapters. To cover the lacunae and unravel the other possible targets, this chapter aims to account for in-silico prediction of test agrochemicals and different gene interactions that will help understand the overall effect of agrochemicals in the system and will also be helpful for the prediction of its future use.

4.2 Materials and methodology

Database Construction and Target Prediction:

Chemical SMILES structures of ingredients were found from the online software PubChem. These formulae were imported into the online Swiss Target Prediction network database (http://www.swisstargetprediction.ch/) (Gan et al., 2019) to identify possible target proteins in *Homo sapiens* species. Results were received in the form of binding Probability for Imidacloprid, pyrazosulfuron ethyl, cymoxanil, and Mancozeb compounds. SMILES formula of each molecule was used to run target prediction (Table 4.1)

Sr No	Compound Name	SMILES formula
1	Imidacloprid	[O-][N+](=O)N=C1NCCN1Cc1ccc(Cl)nc1
2	Cymoxanil	CCNC(=O)NC(=O)C(=N/OC)\C#N
3	Mancozeb (zineb) (maneb)	C(CNC(=S) [S-]) NC(=S) [S-].[Zn+2] C(CNC(=S)[S-])NC(=S)[S-].[Mn+2]
4	Pyrazosulfuron ethyl	CCOC(=O)C1=C(N(N=C1)C)S(=O)(=O)NC(=O)NC2=NC(=CC(=N2)OC)OC

Table 4.1: Agrochemicals and SMILES formula

Gene Interactions:

Association of different genes with the selected candidate markers, bioinformatics software was used for the representation of pathways and to investigate which pathways were being affected by the tested agrochemicals. The pathway maps were generated for the candidate genes listed in the Table 4.1 with their upstream and downstream association of possible genes, which may get alter under the influence of the agrochemicals. Genes were selected

by searching the homology using NCBI Blast which was also found in a teleost, hence they were considered for analysis, other genes were rejected which are yet to be discovered in teleost. The analysis was carried out using pathway commons (https://www.pathwaycommons.org), and Wiki pathways (https://www.wikipathways.org/index.php/WikiPathways) and was visualized in Cytoscape bioinformatic software.

4.3 Result

Target Prediction:

Target prediction using online software Swiss Target Prediction delivered the for Imidacloprid, pyrazosulfuron ethyl, Cymoxanil, and Mancozeb have been done. Results were obtained in form of Pie charts, indicating categorization of target protein, as well as target protein with binding probability. Figure 4.1 to 4.4 indicate the structure of the respective agrochemical and its pie charts. Each Pie chart explains % of the Protein class of total target proteins belonging to a particular protein class. Out of all the molecules, Imidacloprid exhibited binding in diversified protein classes including nuclear receptor, cytochromes, enzymes, proteases, Kinases, GPCRs, and transporters. Whereas rest compounds like pyrazosulfuron ethyl, Cymoxanil, and Mancozeb exhibited very little binding probability with proteins. Elaborating, Imidacloprid showed 26.7 % of binding with ligand-gated channel and Family a G protein-coupled receptor class of proteins and 20 % of binding with Enzymes Figure 4.1: Imidacloprid structure and Pie chart indicating Target proteins class distribution

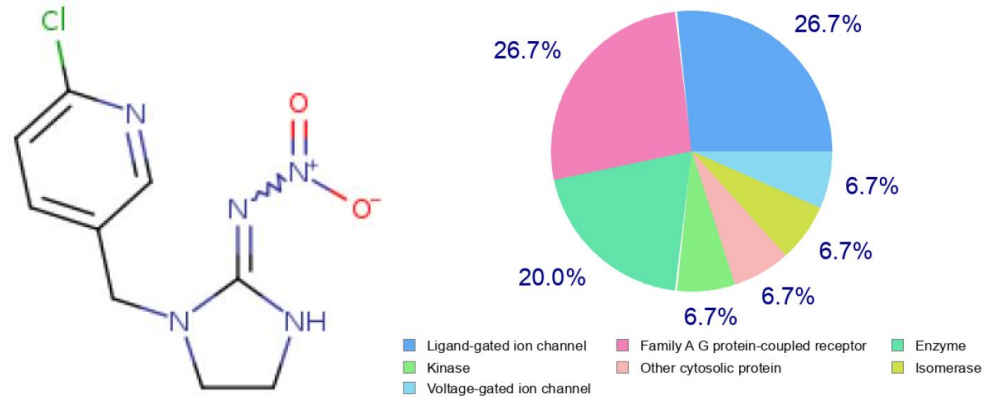

Figure 4.1: Imidacloprid structure and Pie chart indicating Target proteins class distribution

Sr no	Compound	Target	Target Class	Probability	UniProt ID
1	Imidacloprid	Neuronal acetylcholine receptor; alpha4/beta2	Ligand gated ion channel	1.0	P17787
2		Neuronal acetylcholine receptor protein alpha-4 subunit (by homology)	Ligand gated ion channel	1.0	P43681
3		Bax	Ligand Binding Site	0.45	Q5EAR7
4		Cyclin A	Cyclin	0.41	A0A3P9B0M9
5		Cytochrome P450	Substrate binding Site	0.41	P79739
6		Cyclin E	Cyclin	0.35	A0A3P9AY88
7		Neuronal acetylcholine receptor protein alpha-7 subunit	Ligand gated ion channel	0.048	P36544
8		Neuronal acetylcholine receptor subunit alpha-3	Ligand-gated ion channel	0.048	P32297
9		Trace amine associated receptor 1 (by homology)	Family A G protein coupled receptor	0.048	Q96RJ0

Table 4.2: Results of In-silico analysis of Imidacloprid

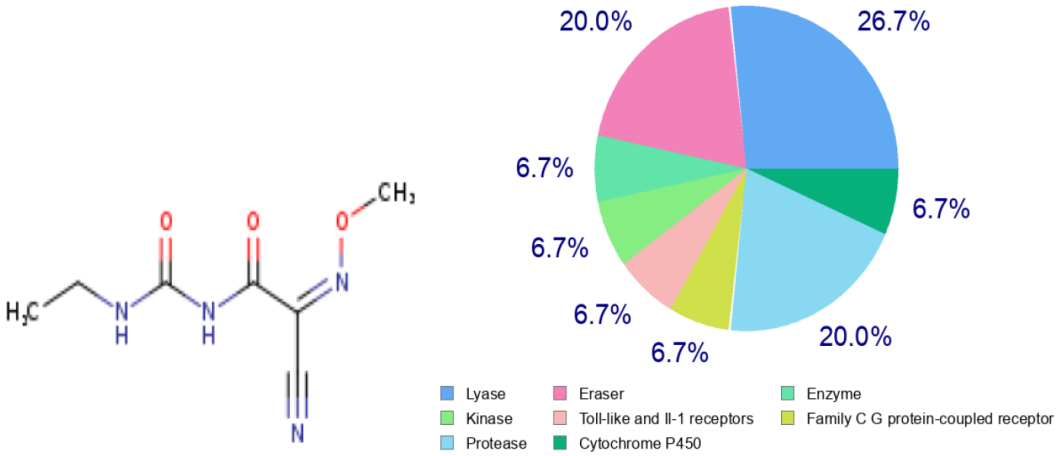

Figure 4.2: Cymoxanil structure and Pie chart indicating Target proteins class distribution

Sr no	Compound	Target	Target Class	Probability	UniProt ID
1	Cymoxanil	Carbonic anhydrase II	Lyase	0.031	P00918
2		Carbonic anhydrase XII	Lyase	0.031	O43570
3		Carbonic anhydrase IX	Lyase	0.031	Q16790
4		Lysine-specific demethylase 5C	Eraser	0.031	P41229
5		Lysine-specific demethylase 4B	Eraser	0.031	O94953

Table 4.3: Results of *In-silico* analysis of Cymoxanil

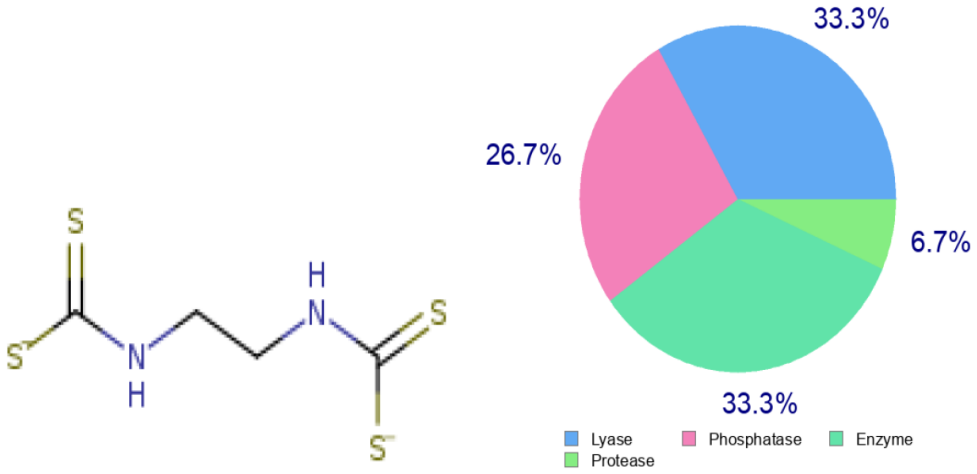

Figure 4.3: Mancozeb structure and Pie chart indicating Target proteins class distribution

Sr no	Compound	Target	Target Class	Probability	UniProt ID
1	Mancozeb	Carbonic anhydrase II	Lyase	0.105	P00918
2		Carbonic anhydrase I	Lyase	0.084	P00915
3		Carbonic anhydrase XII	Lyase	0.084	O43570
4		Carbonic anhydrase IX	Lyase	0.084	Q16790
5		Carbonic anhydrase IV	Lyase	0.023	P22748

Table 4.4: Results of *In-silico* analysis of Mancozeb

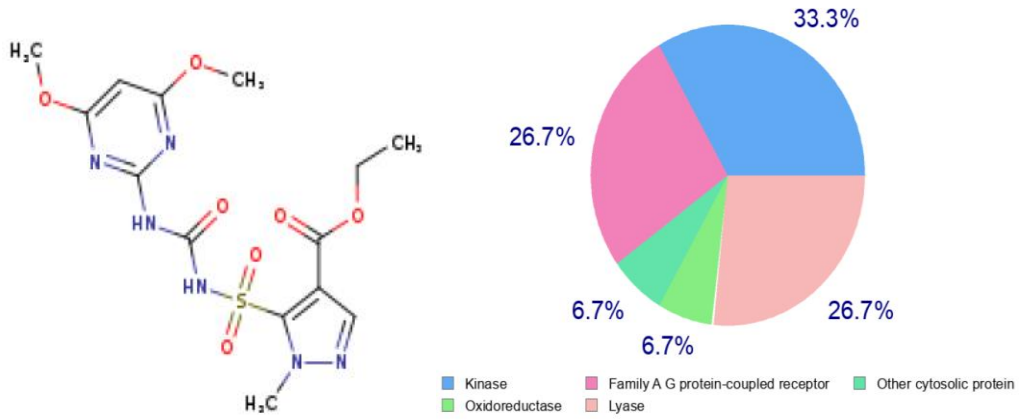

Figure 4.4 Pyrazosulfuron ethyl structure and Pie chart indicating Target proteins class distribution

Sr no	Compound	Target	Target Class	Probability	UniProt ID
1	Pyrazosulfuron ethyl	c-Jun N-terminal kinase 1	Kinase	0.115	P45983
2		c-Jun N-terminal kinase 3	Kinase	0.115	P53779
3		c-Jun N-terminal kinase 2	Kinase	0.115	P45984
4		Adenosine A1 Receptor	Family A G protein coupled receptor	0.115	P30542
5		Adenosine A2a receptor	Family A G protein coupled receptor	0.115	P29274

Table 4.5: Results of in-silico analysis of– Pyrazosulfuron ethyl

Besides Pie charts, the software provided the binding probability of each target protein for that agrochemical along with target protein name, class, and UniProt ID. Out of all the compounds, Imidacloprid showed a high (1.0) binding probability. The rest of all the other agrochemicals showed very little binding probability and were in the range of 0 – 0.15. All the results are mentioned in Table: 4.5.

Gene association:

Gene association studies were carried out using two open-based software i.e. Pathway-common and Wiki Pathways for the candidate genes. The association was checked for the apoptotic markers, proliferation markers, and other toxicological markers. The interactions of pcna, ccne1 (Cyclin E), ccna2 (Cyclin A) was being found to be interacting with numerous genes. A total of 396 genes were found to be in close association with the candidate genes whose gene expression was studied. Out of which 18 were found to function as controlling state change of other genes, while seven were found to be involved in controlling the expression and the remaining 371 were designated as state change genes. Similarly, the genes of the apoptotic pathway were analyzed. Primarily, the interaction of bax and bcl2 was visualized. The interaction showed that the two were interacting controlling each other expression pattern. Furthermore, bcl2 also showed its controlling state for bak1 (bcl2 antagonist/killer 1) gene. While bax showed its controlling expression of agt (angiotensinogen). Both bax and bcl2 showed complex formation with VDAC1 (Voltage-Dependent Anion Channel 1).

Moreover, the interaction was elucidated by including casp3 with bax and bcl2. The interaction showed that casp3 had controlling state change with bcl2 and bax, while the expression pattern control was only found between bcl2 and casp3. The interaction also revealed that casp9, tnf (Tumor Necrosis Factor), fasl (Fas Ligand), ptk2 (protein tyrosine kinase 2) had a control state change. While there was no direct interaction between nfkb with casp3,9. Nevertheless, the addition of tnf with this interaction showed that it was controlling the expression of nfkb. The interaction further revealed that it showed the controlling the state change of hdac2 (histone deacetylase 2). Other genes and their interaction with the candidate genes are shown in the Table 4.6

Sr No.	Name of the genes	Associated Genes	Interaction types
1	ccna2	hdac1	Genera expression, in complex
		e2f1	in complex
		cdk2	In Complex
		atm	In complex
		nbn	In complex
		rad50	In Complex
		cdkn1a	Control state change, in complex
		e2f4	Control Expression
		rbl1	Control Expression
2.	ccne1	cdk2	In Complex
		cdkn1b	In Complex
		cdkn2b	In Complex
		rb1	In Complex
		skp2	In complex
3.	Pcna	ccne1	In complex
		cdk2	In complex
4.	nfkb1	rela	In complex
		bcl2l	Control gene expression
5.	casp3	sod2	Control state change
		map2k1	Control state change
		casp9	Control state change
		tnfsf10	Control state change
		ptk2	Control state change
		faslg	Control state change
6.	Bax	vdac1	In complex

		agt	Control expression
7.	Nfkb	ier3	In complex
		prkc1	Control state change, in complex
		prkc	Control state change, in complex
		ripk1	Control state change, in complex
		tradd	Control state change, in complex
		traf2	Control state change, in complex
		sostm1	Control state change, in complex
		prkcz	Control state change, in complex
		il6	Control gene expression, in complex
		cxcl8	Control gene expression, in complex
		bcl2l	Control gene expression, in complex
		cebpb	Control gene expression, in complex

Table 4.6 The candidate genes with associated genes and its interaction types. the interaction types were deduced from pathways commons.

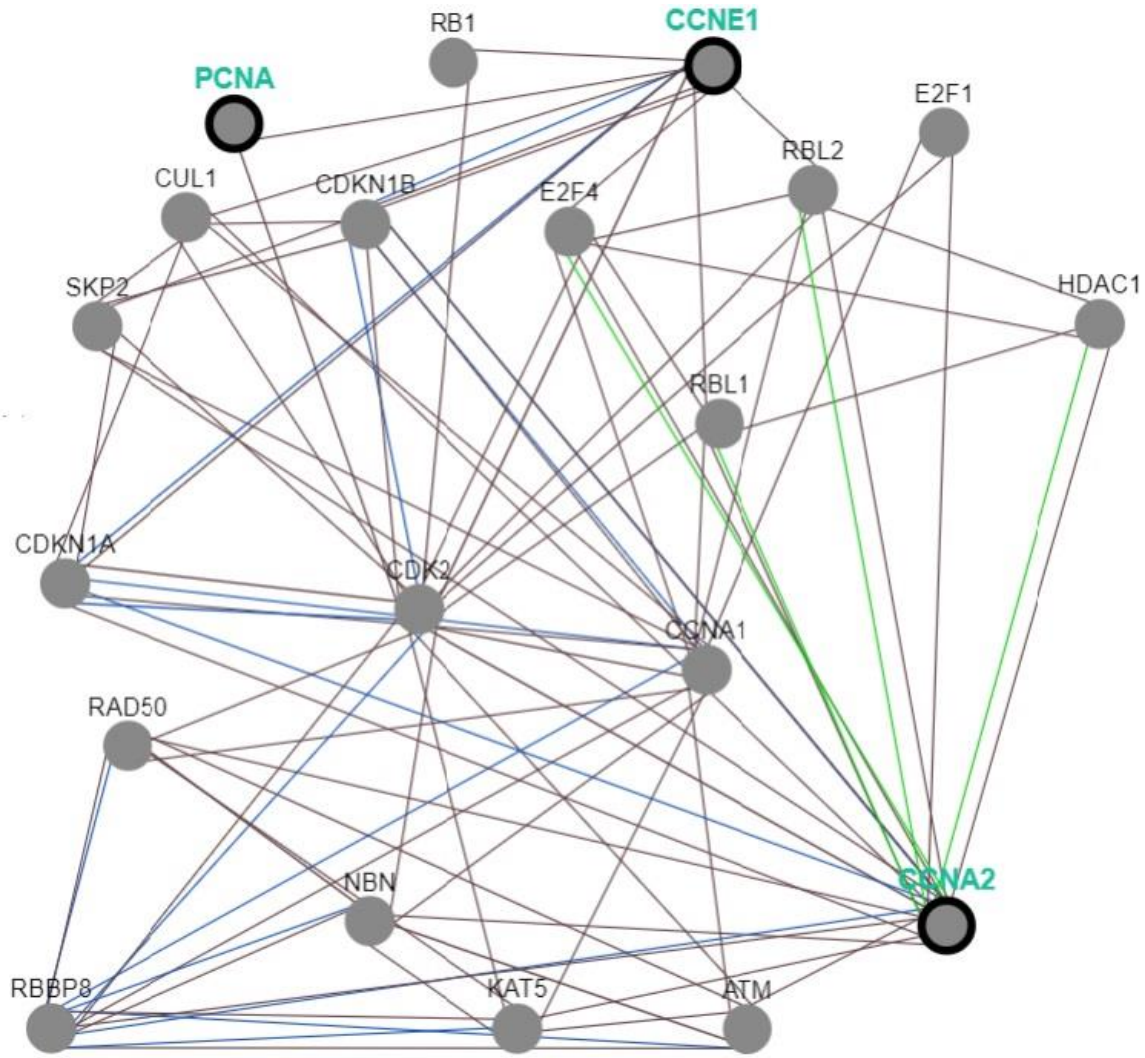

Figure 4.5 Gene interaction of PCNA with different genes, shows interaction with more than 10 genes (Green: represents in complex, Red: represents gene expression, Blue: controls stat change)

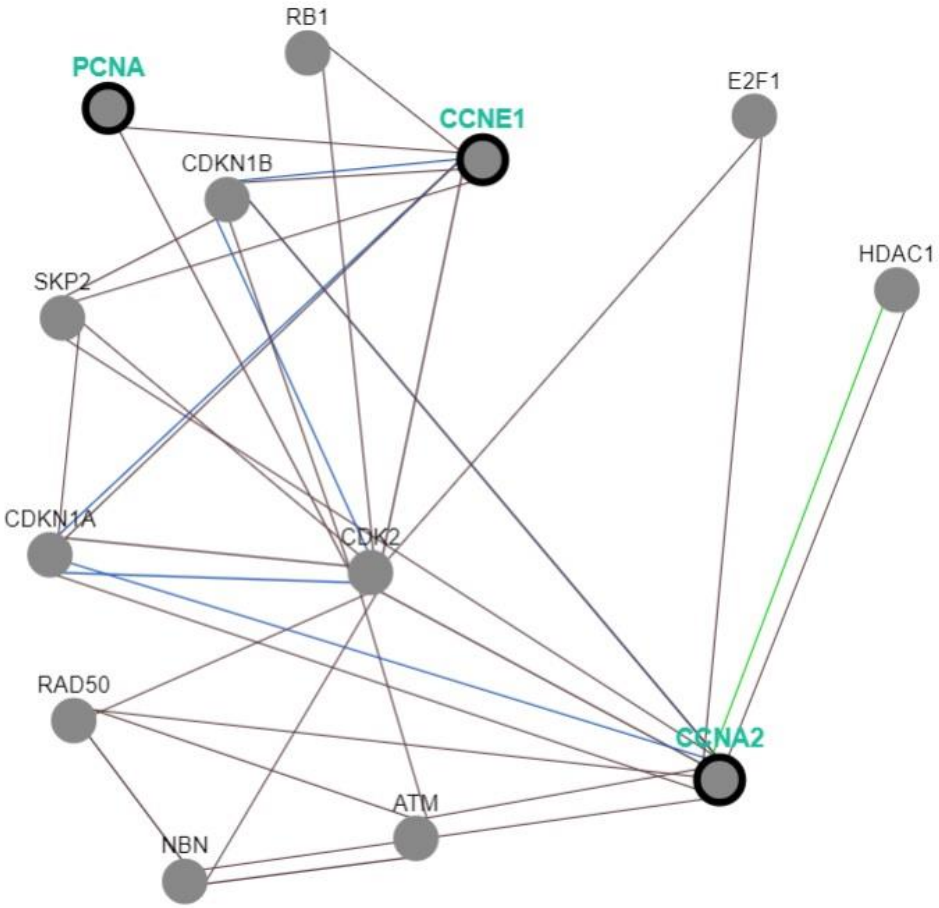

Figure 4.6: Gene interaction of CCNE1 with different genes. shows interaction with more than 10 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state)

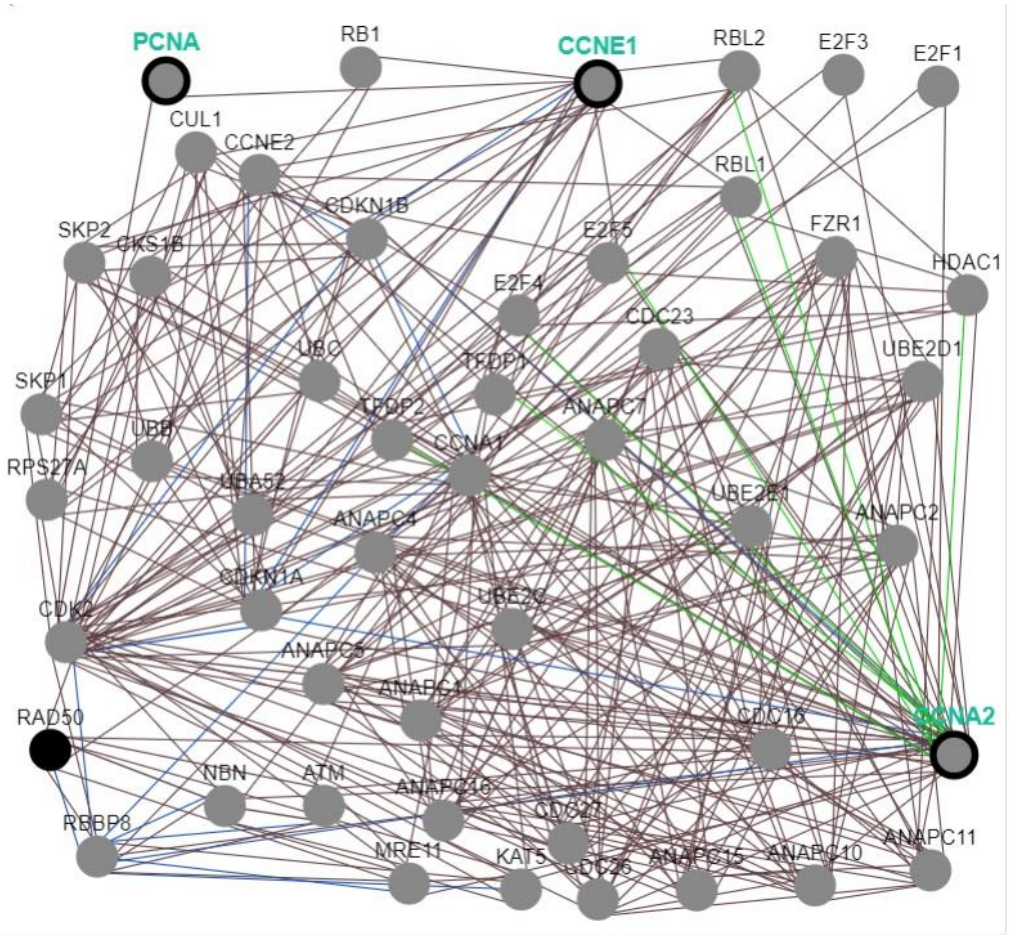

Figure 4.7: Gene interaction of CCNA2 with different genes, shows the overall interaction with 51 genes. (Green: represents in complex, Red: represents gene expression, Blue: control state change)

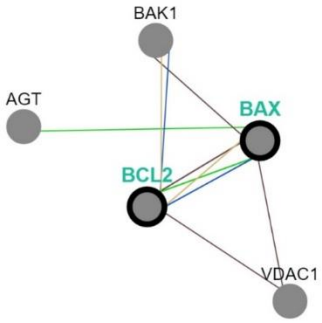

Figure 4.8: Gene interaction of bax, with different genes, shows interaction with more than 10 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state change)

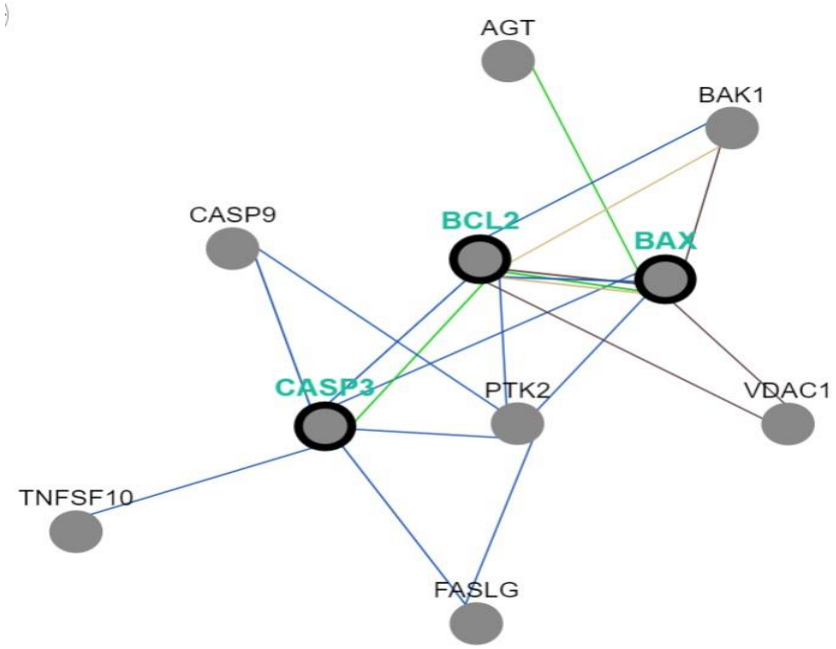

Figure 4.9 Gene interaction of bcl2, with different genes, shows interaction with more than 10 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state change)

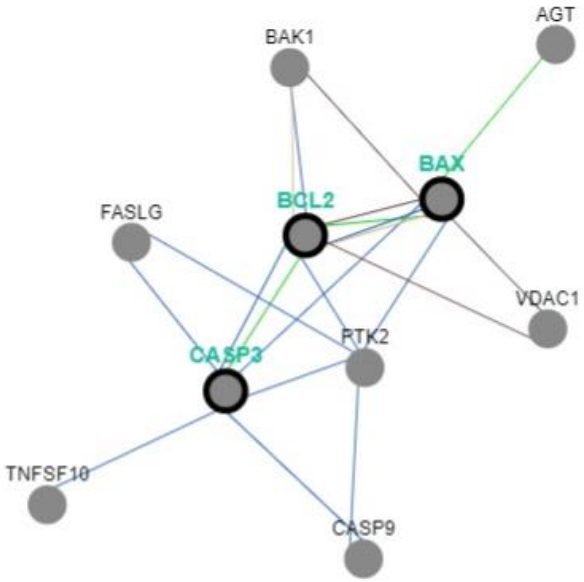

Figure 4.10: Gene interaction of casp3 with different genes, shows interaction with more than 10 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state change)

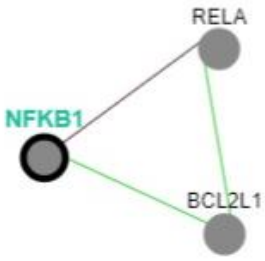

Figure 4.11: Gene interaction of nfkb with different genes, shows interaction with more than 10 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state change)

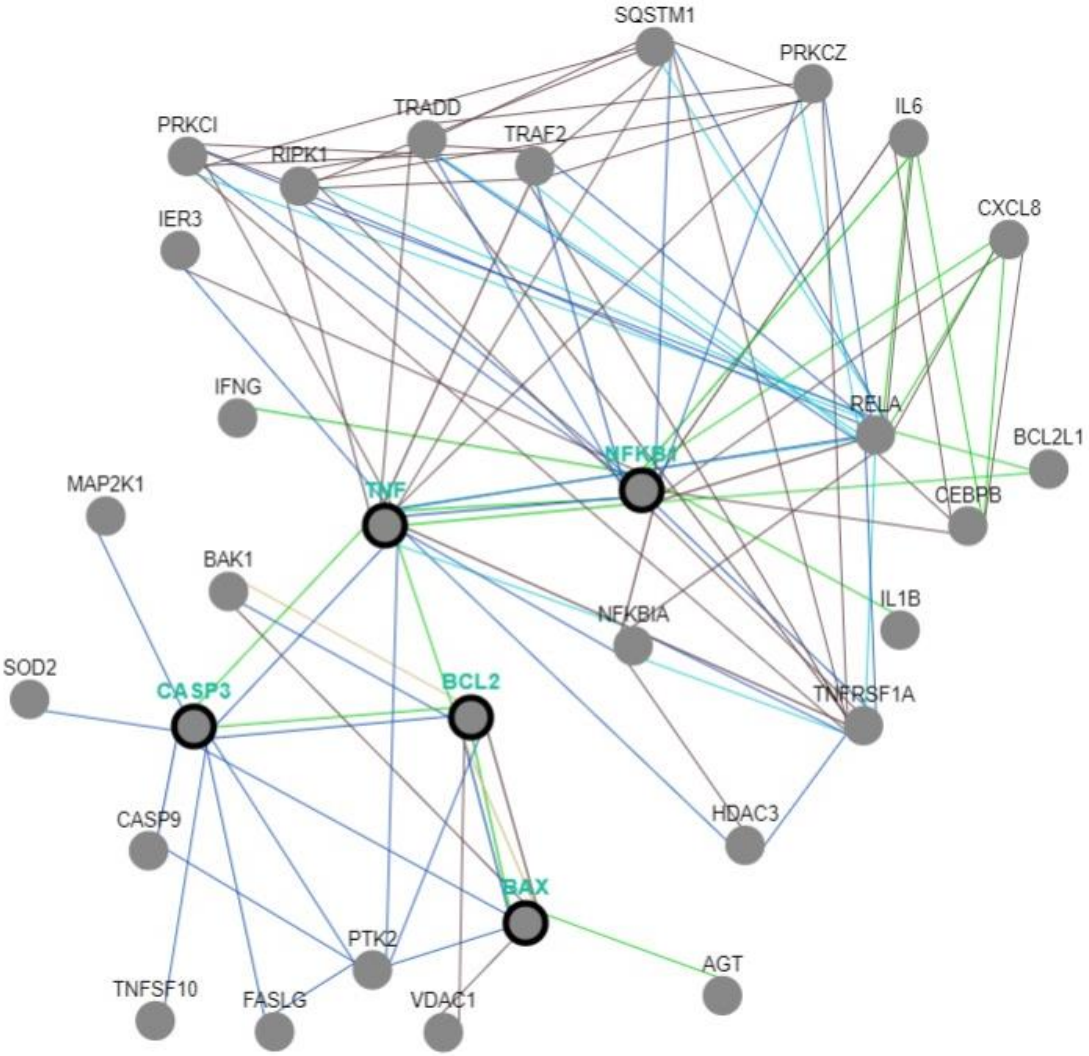

Figure 4.12: Gene interaction of tnfα with different genes, shows interaction with more than 30 genes. (Green: represents in complex, Red: represents gene expression, Blue: controls state change)

4.4 Discussion:

Since the last couple of decades, structure-based drug discovery has gained significant popularity. Advancements in information technology and the development of various bioinformatics software platforms have revolutionized the drug discovery approach (Kirchmair et al., 2008). *In-silico* study is the structure-based drug designing and target prediction which provides insights into the interaction of various chemical compounds with its specific target protein, yielding information about affinity and the specificity of the interaction. The drug design and target prediction-based approach aim to find out the identification of the compound as well as its molecular target. For instance, where the target protein or structure of the target protein is unknown, ligand-based target prediction approaches are employed to fulfill the desired information. Obtained structure or targets using such an in-silico method are often used to design further experiments.

Molecular docking is the preferred *in-silico* technique for the simulation of biomolecular interaction. Such methods hold the capability to give insights into the interaction at the molecular level, which provides an opportunity to study and characterize the binding and interacting site in target protein (Waghulde et al., 2018). With such bioinformatics tools, it could possible to explore the probable binding of agrochemicals with target proteins. Bioinformatics study in the present study has identified few target proteins with which study agrochemicals have a binding probability. These target proteins belong to various protein classes such as the Ligand-gated ion channel, Family A G protein-coupled receptor, Lyase, and Kinase. Out of four players, Imidacloprid has arisen to have strong binding probability with ligand-gated ion channels like Neuronal acetylcholine receptor, Cyclin A and E cytochrome p450, bax, and bcl2. This receptor works as a Sodium channel upon binding of acetylcholine. These results are supported by earlier published reports. It is established that Imidacloprid shows a high affinity toward

acetylcholine receptors and impairs the nervous system in insects, mites, termites, and ticks (Yadav et al., 2020). On reaching into the aquatic environment, it causes toxicity to fishes by affecting their physiology, behavior, hematology and biochemistry of fishes (Inyang, 2008; Qadir & Iqbal, 2016). Besides these, Pyrazosulfuron ethyl had shown binding probability with kinases and GPCRs like c-Jun N-terminal kinase 1/2/3 and adenosine receptors respectively. Correlating the target proteins of Imidacloprid and Pyrazosulfuron ethyl together, it can be anticipated that agrochemical might have a role in the impairment of GPCR, c-Jun terminal kinase, and Sodium ion-based pathways.

In-silico study of Cymoxanil and Mancozeb also revealed target proteins that belong to the category like Lyase and Eraser. Though binding affinity was less, both the agrochemicals show affinity toward Carbonic anhydrase enzymes as well as Lysine-specific demethylase. The gene interactions were studied with the help of pathway commons and were visualized in Cytoscape software. The overall aim was to understand that what are the genes that are getting altered with respect to the candidate genes. This was done not only to understand the downstream signaling but also to predict the genes getting altered in specific pathways. The genes which were found in the Zebrafish were only considered for the study and other genes related to humans were eliminated.

The cell cycle pathway is a cell division process that moves in one direction. It's a highly vital process that is critically required by the cell for its survival. It mainly comprises of 4 phases namely G1, S, G2 and M phase. These phases are regulated by two specific classes of proteins CDKs and Cyclins (Cava et al., 2016). These are also known to form complexes that will further downstream initiate the process via phosphorylation. The cell cycle pathway is a complex process that is highly regulated by three major checkpoints. The first checkpoint is the G1 checkpoint, which confirms that

whether or not a cell will enter into the cell division process. The second checkpoint, G2, ensures whether the cell will enter into mitosis or will remain in the G2 phase (Thakur & Chen, 2019). Both of the checkpoints are highly studied and are known to be affected by the presence or absence of various growth factors, DNA damage, or replicative senescence induced by the xenobiotics or a toxicant (Majeed et al., 2014). The ultimate checkpoint, metaphase, warrants proper chromosome packaging and alignment prior to cell division. The cell cycle pathway is intrinsically linked that maintains the homeostasis of cell survival and cell death. Malfunctioning and dysregulation of cell cycle checkpoints requirements will lead a cell to undergo apoptosis.

So to understand this relationship, pathway analysis for the cell cycle and associated genes were analyzed. A close association was studies where, 396 genes were found to interacting with the genes like pcna, ccne1 (cyclin E), and ccna2 (Cyclin A2). Among all the genes analyzed, 18 were found to function as controlling state change of other genes, while seven were found to be involved in controlling the expression and the remaining 371 were designated as state change genes. We predict that at the cells exposed to MD and HD of imidacloprid (insecticide) and Curzate (Fungicide), there may be a change in the expression of all the genes which are in close contact with the cell cycle regulators. Additionally, this cell will be having a highly dysregulated cell cycle and may eventually die due to excessive usage of ATP due to altered metabolism.

In this context, the apoptotic pathway was also validated. This pathway has been broadly classified into the intrinsic pathway and the extrinsic pathway of apoptosis. The classification is based on the origin of their occurrence for instance; extrinsic pathway begins outside the cells, when the extracellular conditions are not feasible for the cell survival, while the intrinsic pathway is initiated by an injury occurring within the cells resulting in the stress condition thus activating the apoptotic pathways. Both the pathways ultimately lead to the activation of a family of Cys (Cysteine) proteases that

are known as caspases which are proteases that functions in a cascade that leads to cell death (Parrish et al., 2013). So toxicants are also known to alter the core apoptotic pathway that leads to dysregulation of normal cellular homeostasis and cause pathogenic events.

To understand the process in more detail gene interaction studies were performed to get a deep insight into the regulation part. Genes like casp3, bcl2, bax, nfkb1 and tnfα were taken into consideration as we had already established its gene expression profile and fold change was noted. A multitrophic interaction system was obtained where casp3 was found to be controlling the expression pattern of bcl2 and bax. Similarly, it was also in a close connection with antioxidants genes like sod and cat. The pathway also revealed that numerous genes for instance of hdac4 for acetylation, il6 and il1B for inflammatory pathway, map2k1 for the cell cycle progression were found to be in close contact bcl2, casp3, nfkb. This close interaction and complex formation may get altered due to exposure of MD and HD of Imidacloprid and curzate. The exposure of these pesticides disrupts the ultimate homeostasis and leads the cells into the apoptotic pathway.

As the pesticides are acting on multiple pathways thus their mechanism of action can't be limited to one or two genes/proteins. It is therefore necessary to test the action via checking binding affinity and other Lipinski properties by performing individual docking studies. Additionally, the homology modeling, x-ray crystallographic studies will unravel the potential binding of the imidacloprid and curzate and will open up new avenues for further research.

4.5 Conclusion:

In a nutshell, the present study shows that

1. among all the agrochemicals validated, Imidacloprid has a very strong binding affinity with ligand-gated ion channel, cyclins, and bax, bcl2 for the apoptosis pathway. The present study also enlists the new possible targets of Imidacloprid, curzate which needs to be accounted in the in-silico databases.
2. Pathway analyses strongly suggest that alteration of candidate genes like ccne1, ccna4, pcna, bax, bcl2, casp3 tnf and nfkb will lead to change in the expression pattern of downstream genes and their protein products and ultimately leads the cell into apoptosis.
3. All the evidence i.e. from morphological examination, gene expression to the *In-silico* analysis suggests that cells are highly under stressed and the use of test pesticides i.e. imidacloprid, curzate should be controlled.