

The Design of New and Environmentally Safe Herbicides Using AI and Molecular Modeling

Micah Shaver and Jerry A Darsey*

Center for Molecular Design and Development, University of Arkansas at Little Rock, United States of America

ISSN: 2637-8078



***Corresponding author:** Jerry A Darsey, Center for Molecular Design and Development, University of Arkansas at Little Rock, 2801 S University Ave, Little Rock, AR, USA

Submission:  January 23, 2022

Published:  February 08, 2023

Volume 5 - Issue 5

How to cite this article: Micah Shaver, Jerry A Darsey*. The Design of New and Environmentally Safe Herbicides Using AI and Molecular Modeling. Significances of Bioengineering & Biosciences. 5(5). SBB. 000624. 2023.

DOI: [10.31031/SBB.2023.05.000624](https://doi.org/10.31031/SBB.2023.05.000624)

Copyright@ Jerry A Darsey, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

Herbicides are biocidal chemical compounds used to kill weeds primarily by inhibiting their growth [1]. Weeds can cause serious damage in agriculture, resulting in critical losses of yield, quality, and profit. Herbicide residues have been found on food for human consumption, and some are dangerous to human health. For example, paraquat has been banned or severely restricted in many countries due to concerns over its effect on the environment and on people's health. The discovery of less toxic compounds that prevent the growth of weeds could greatly impact food sustainability as well as allow for a safer environment for humans and crops alike. The goal of this research is to take known herbicides and, using molecular modeling, modify these molecules in order to identify herbicides with more potent weed killing abilities but with reduced environmental impact. This is done by training an Artificial Neural Network (ANN) software to predict the IC_{50} values of the newly designed herbicides. The half-maximal inhibitory concentration (IC_{50}) of an herbicide is an informative measure of an herbicide's effectiveness [2]. A lower IC_{50} value means that the herbicide is effective at low concentration. In this research we will model approximately 200+ modifications to 16 known herbicides. This should provide several molecules that will have superior weed-killing properties but with much less environmental impact.

Keywords: IC_{50} ; Artificial neural network; Artificial intelligence; Gaussian molecular modeling software

Abbreviations: ANN: Artificial Neural Network; AI: Artificial Intelligence; IC_{50} : Inhibition Concentration at 50%; LD_{50} : Lethal Dose at 50%; LUMO: Lowest Unoccupied Molecular Orbital; HOMO: Highest Occupied Molecule Orbital

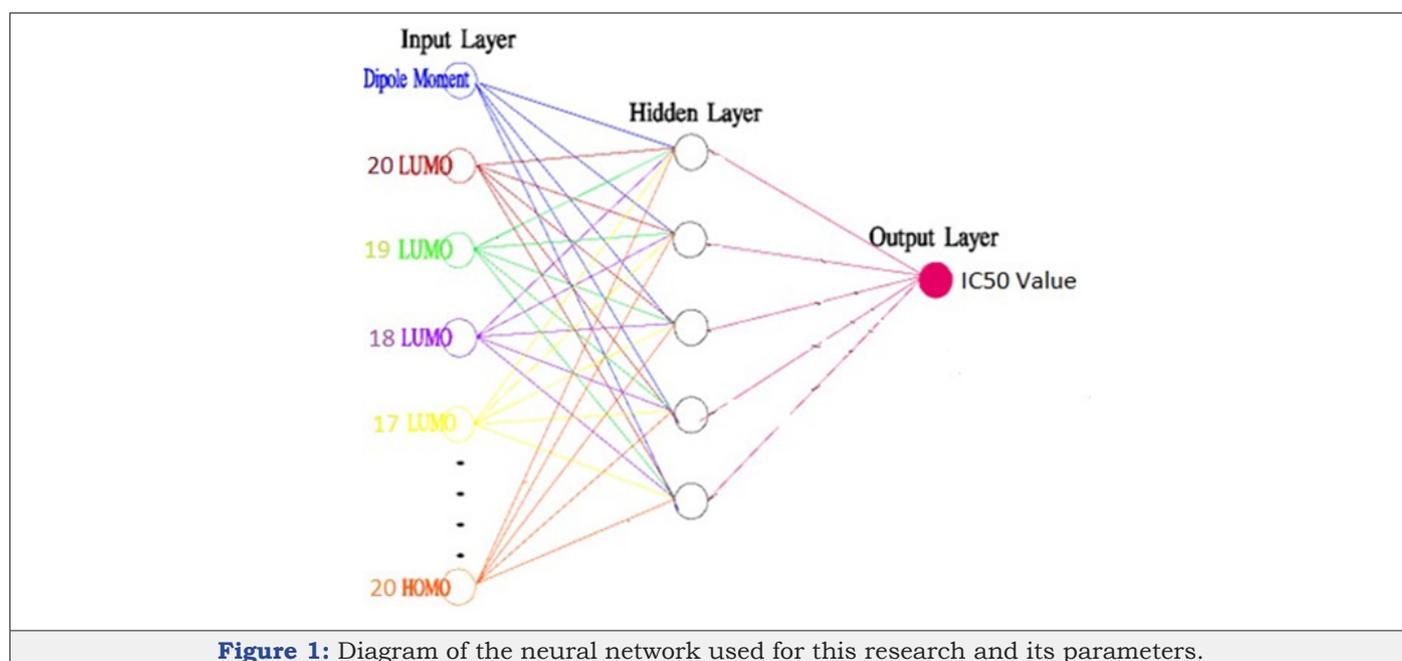
Introduction

Herbicides are biocidal chemical compounds used to maximize crop productivity by eliminating weed growth. This is important because weeds can minimize vegetation and decrease crop productivity. Weeds compete with other crops for sunlight, water, nutrients, and space which can lead to a lower yield of plant crops. Crops play a huge role in the economy, and they also provide a massive food source for humans. Arkansas, where this research is conducted, is a major producer of a variety of crops such as rice, corn, wheat, sorghum, soybeans, and more. Herbicides are essential for getting high yields of crops and decreasing the amount of weed infestations. While they are necessary, they can come with a variety of drawbacks. Like any chemical, all herbicides are potentially hazardous to animal and human health. Some herbicides are no more toxic than table salt, while others are considered toxic and carcinogenic. Other herbicides have been linked to air, water, and soil pollution [3]. Many herbicides have been banned, restricted, or only allowed to be used under certain conditions. The goal of this research is to model and design new herbicides that have less of these disadvantages. The following 16 herbicides were chosen for study: By making modifications to each of these herbicides, there is potential to find more potent and environmentally safer herbicides. The IC_{50} value is the half-maximal inhibitory concentration. In other words, it is a measure of the effectiveness of a compound in inhibiting some biological function. It is correlated to the potency of a herbicide; a lower IC_{50} value is better: this would mean it takes less concentration of a herbicide to inhibit a function by 50%. Having a lower IC_{50}

and subsequently higher potency is also worth studying because it would mean less amount of a chemical is going out into the environment. Although the IC_{50} does not have any correlation to the toxicity of a herbicide, finding a lower IC_{50} is desirable. Future research will be conducted to address the toxicity of the studied herbicides. Gaussian 09 is a general-purpose ab initio computational quantum mechanical software package that uses fundamental laws of quantum mechanics to calculate energies, molecular structures, spectroscopic data (NMR, IR, UV, etc.) and many more advanced properties [4].

For this research, only the optimization feature was used which optimizes a molecule's geometry and generates a file that contains data about the molecule. Some of the data includes the Lowest Unoccupied Molecular Orbitals (LUMOS) and the Highest Occupied Molecular Orbitals (HOMOS) as well as the total dipole moment. These are the data primarily used in this research. This program is typically utilized by chemists, biochemists, and physicists. Gaussian 09 was the primary quantum mechanical simulation software used in this project. The other software used in this research is the Artificial Intelligence (AI) or Artificial Neural Network (ANN) software, "NETS." The primary function of NETS is to provide a flexible system

for manipulating a variety of neural network configurations using generalized back propagation learning method [5,6]. An artificial neural network is a technique for building a computer program that learns from data. They are loosely based off of the way our brains are believed to work and also contain mathematical "neurons." First, a connection (or neurons) is created and then the network is asked to solve a problem which it attempts to do over and over again while diminishing the connections that lead to a failure and strengthening the connections that lead to a success (Figure 1). Basically, a neural network takes in data, trains itself to recognize patterns in that data, and then predicts the output(s) for a new set of similar data. One example of a commonly used neural network is facial recognition. Smart phones were trained to determine what is a face and what is not a face and were then trained to only unlock the phone for specific faces. Self-driving cars are another example of neural networks being employed-in this case, the networks train themselves to distinguish between road, stop signs, etc. Neural networks are becoming increasingly more common in everyday life, and they can recognize complex patterns that normal programs or sometimes even humans can't understand. It is very useful in solving for non-linear data as it is multi-dimensional.



Materials and Methods

The first step of this research was to pick and study different herbicides. The 16 listed in (Table 1) were researched and their IC_{50} values were determined from the literature. All the IC_{50} values were converted to units of ng/ml as that happened to be the unit most commonly used in the literature. When the herbicides were chosen and their IC_{50} values were recorded, the 16 herbicides were modeled using the Gaussian 09 software. The optimization feature was used to optimize the molecule's geometry and provide data on each molecule. As mentioned before, the optimization feature also calculates the LUMOS, HOMOS, total dipole moment of a molecule,

and much more. These data were calculated for each herbicide and to be compatible with the ANN. The first 20 LUMOS, the first 20 HOMOS, and the total dipole moment made up the input layer of the neural network. The number of LUMOS and HOMOS was chosen to maximize the correlation between the LUMOS, HOMOS, and dipole moment with the IC_{50} s. The number 20 was chosen based on past research experience. Each molecule has a different number of LUMOS and HOMOS, so it was necessary to choose the same number (in our case, 20) for each molecule to maintain consistency for the neural network. In the neural network created, the input layer consisted of a total of 41 input nodes: the dipole moment, 20

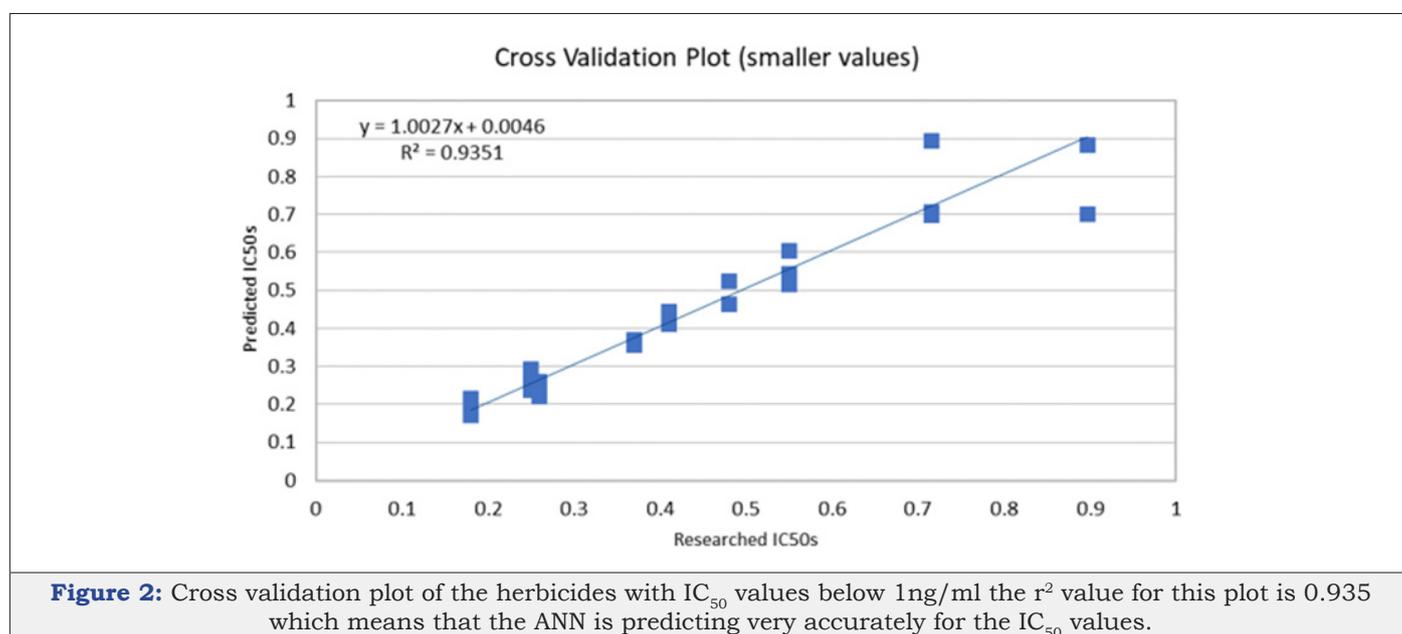
LUMOS, and 20 HOMOS. The “hidden layer” in a neural network is a layer of mathematical functions where the training occurs [7]. In this research, the number of nodes in the hidden layer was 10. This number was chosen based on the research mentor’s previous experience with artificial intelligence—once again, this parameter is more of an art than a science.

Table 1: The 16 studied herbicides.

2,4-D	Acetochlor	Ametryn	Atrazine
Bromacil	Chlorsulfuron	Cyanazine	Desmedipham
Paraquat	Prometon	Prometryn	Propanil
Propazine	Simazine	Terbacil	Triclopyr

The output layer consisted of one output node containing the IC_{50} value. Figure 1 shows a diagram of a typical neural network’s architecture. First, the ANN was trained with data from 14 of the 16 herbicides. After training the input and output for the 14 herbicides, the network was asked to predict the IC_{50} values of the remaining 2 herbicides that weren’t included in the training. The neural network was only given values for the 14 herbicides and was asked to predict the IC_{50} for the two left-out herbicides. This process is called cross validation and was repeated until each herbicide had their IC_{50} value predicted at least twice. The purpose of this process was to determine how well the NETS program was predicting an herbicide’s IC_{50} values and to determine if there is a correlation between the input layer and output layer. This process also

allowed for a cross validation plot to be created. A cross validation plot is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model, and the other used to validate the model. This can be plotted and the r^2 value of the plot can be determined. The r^2 value is an estimate of the relationship between movements of a dependent variable based on an independent variable’s movement. The closer the r^2 is to 1, the better the model fits the data. Figure 2 shows the cross-validation plot of the IC_{50} values below 1ng/ml. The experimental or “literature” IC_{50} values are on the x-axis and the neural network predicted IC_{50} values are on the y-axis. After training the neural network, the 16 herbicides were arbitrarily modified. Each herbicide in this research has an aromatic ring, and for the first modification, a bromine replaced either hydrogen or some another element. This was done using the Gaussian 09 software and the same optimization feature was used to get the 20 LUMOS, 20 HOMOS, and total dipole moment of these new molecules. These modified molecules’ data were used as the input data of the trained NETS program. After the neural network had been trained to predict IC_{50} values, it was then used to predict the newly modified molecules’ IC_{50} values. Each herbicide has been modified 5 times giving a total of 80 modifications. The research is ongoing, and the goal is to find new molecules that have a lower IC_{50} value by a factor of at least 10; this would mean the new molecule is ten times more potent in killing weeds than the original herbicide.



Results and Discussion

The modified herbicide showed enhanced results for two modifications. One example is for the herbicide Propanil. The figure below shows the structure of propanil. The green atoms are chlorines, the blue atom is nitrogen, the red atom is oxygen, the dark gray atoms are carbons, and the small light gray atoms are hydrogens. Figure 3 shows the original chemical structure of

propanil (chemical formula: $C_9H_9Cl_2NO$) without any modifications. Figure 4 shows the modification made to propanil. Using the modeling software, one of the hydrogens on the benzene ring was replaced by a bromine (dark red). Propanil’s original IC_{50} value was 2.7ng/ml. After replacing the hydrogen with a bromine atom, the neural network predicted that this modified herbicide would have an IC_{50} value of .08ng/ml (Table 2). This is the best result thus far and is very promising as the IC_{50} value could potentially be reduced

by almost 2 orders of magnitude. This research will be continued until more modifications of more herbicides show greater potency

(lower IC_{50} s). The following (Table 3) shows the ANN predicted IC_{50} values for the first modifications of each herbicide.

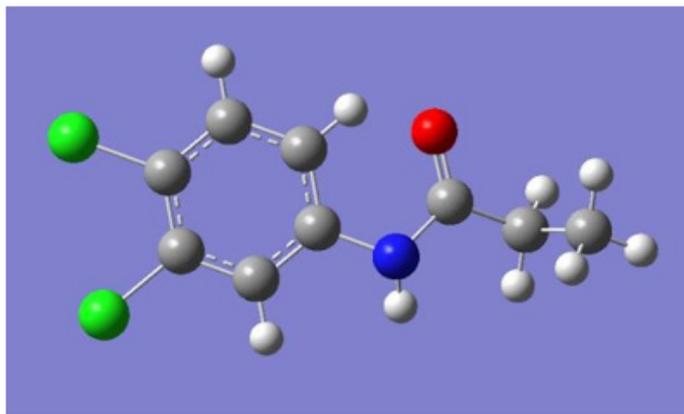


Figure 3: Chemical structure of propanil ($C_9H_9Cl_2NO$) modeled using Gaussian 09.

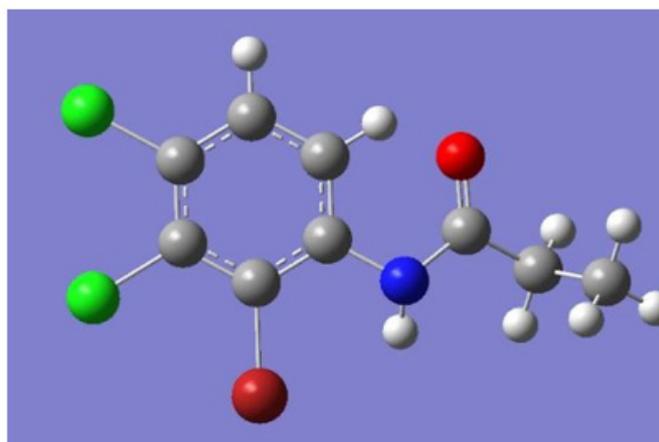


Figure 4: First modification made to propanil using the Gaussian 09.

Table 2: Shows experimental IC_{50} value and ANN predicted IC_{50} value in ng/ml.

	Experimental IC_{50}	Predicted IC_{50}		Experimental IC_{50}	Predicted IC_{50}
2,4D	29.2	29.14	Desmedipham	130	129.82
		29.17			130.13
		29.16			41.24
Acetochlor	0.55	0.54	Paraquat	41.146	41.19
		0.6			0.26
		0.52			0.22
Ametryn	0.18	0.17	Prometryn	0.18	0.19
		0.17			0.19
		0.21			2.66
Atrazine	0.37	0.36	Propanil	2.7	2.63
		0.37			2.69
		0.36			0.52
Bromacil	0.25	0.25	Propazine	0.48	0.46
		0.26			0.24
		0.29			0.24
			Simazine	0.25	

Chlorsulfuron	0.7155	0.71	Terbacil	0.41	0.43
		0.89			0.44
		0.7			0.41
Cyanazine	59	58.84	Triclopyr	0.8976	1.04
		58.98			0.88
		58.98			0.7

Table 3: The ANN predicted IC_{50} values from the first modification. *This is a good result as the IC_{50} value of the modified herbicide was reduced by almost two orders of magnitude. **Some of the ANN IC_{50} value predictions gave negative numbers. We are unsure as to what it means.

Herbicide	Literature IC_{50} Value of Original Herbicide in ng/ml	ANN Predicted IC_{50} Value of Modified Herbicide
2,4-D	29.2	-0.343
Acetochlor	0.55	3.39
Ametryn	0.18	0.678
Atrazine	0.37	-0.872
Bromacil	0.25	-5.9
Chlorsulfuron	0.7155	1.03
Cyanazine	59	-1.52
Desmedipham	130	-0.79
Paraquat	41.146	-3.12
Prometon	0.26	-1.59
Prometryn	0.18	0.259
Propanil*	2.7	0.081
Propazine	0.48	-0.885
Simazine	0.25	-0.457
Terbacil	0.41	-0.258
Triclopyr	0.8976	-2.44

Conclusion

In conclusion, this research has shown that the use of artificial intelligence is a potentially powerful method when used to predict the IC_{50} values of herbicides. The cross-validation plots showed how well the neural

network was able to predict these values. This research has also led to some potentially more potent weed-killing molecules, such as the propanil modification shown in the results sections. In the future, more modifications will be made to find more molecules with better IC_{50} values. After this is complete, the research will be repeated-but instead of training the neural network to predict for IC_{50} values, it will learn to predict LD_{50} values. LD_{50} is the lethal dose that will kill 50% of test subjects (chemicals are usually tested on rats). In the case of the LD_{50} values, it would be beneficial to find molecules with a higher value. The higher the LD_{50} value, the less lethal it is. A high LD_{50} value means that it takes more concentration of a chemical to be lethal. So, if we can design an herbicide with a lower IC_{50} value and a higher LD_{50} value that would mean that the molecule is more potent and less lethal/toxic. If the research leads to a very good result (a molecule with a lower IC_{50} value and a higher LD_{50} value), then the next step will be to synthesize the

molecule. The molecule will be tested on the vegetable garden at the University of Arkansas at Little Rock. Lastly, we hope to test our new molecules on an actual farm.

Acknowledgement

We would like to thank the Collaborative Research Grant of the Donaghey College of Science, Technology, Engineering, and Mathematics of the University of Arkansas at Little Rock for financial support.

References

1. Todd, Brooke, Glenn Sutter Herbicides. EPA Environmental Protection Agency 2.
2. Aykul S, Martinez HE (2016) Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. *Anal Biochem* 508: 97-103.
3. Carbon ADE (2016) Pros and cons of herbicides. Carrhure, USA.
4. Software in chem-bio library. Gaussian 09. Research Guides.
5. Hsu Hansen (2020) How do neural network systems Work? CHM.
6. Ioannou Y (2018) Backpropagation derivation- delta rule. A shallow blog about deep learning.
7. Deep AI (2019) Hidden layer.