

6-Ethoxy-4-N-(2-Morpholin-4-Ylethyl)-2-N-Propan-2-yl-1,3,5-Triazine-2,4-Diamine Endows Herbicidal Activity Against *Phalaris Minor* a Weed of Wheat Crop Field: An *in-Silico* and Experimental Approaches of Herbicide Discovery

Nandan Kumar

Central University of South Bihar

Priyanka Rani

Central University of South Bihar

Shikha Agarwal

Central University of South Bihar

Durg Vijay Singh (✉ dvsbiotech@gmail.com)

Central University of Bihar <https://orcid.org/0000-0001-7441-7704>

Research Article

Keywords: Phalaris minor, Herbicide resistance, Photosystem-II D1 Protein, De novo Design, MD Simulation, MM/PBSA, whole plant assay

Posted Date: July 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-631634/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

P. minor is a major weed of wheat crop which has evolved resistance against herbicides. Isoproturon is the most accepted herbicide developed in 1992. Later, introduced herbicides also developed resistance and cross-resistance to their respective binding sites. Isoproturon binds at the Q_B binding site of the D1 protein of photosystem-II (PS-II) which blocks the electron transfer in photosynthesis. In this work, a series of computational studies have been implemented to discover herbicides against the D1 protein of *P. minor*. Through computational study, twenty-four lead molecules are reported which have shown a higher binding affinity and inhibition constant compared to the reference ligand. The conformational stability of docked complexes was evaluated by molecular dynamics simulation and MM/PBSA method. Ala225, Ser226, Phe227, and Asn229 present in the binding site of protein play an important role in the stability of the protein-lead complex *via* hydrogen bond interaction and pi-pi interaction. Moreover, interactions and binding free energy between protein-lead complexes have been calculated using MM/PBSA methods. It has revealed that binding was energetically favourable and driven by electrostatic interactions. Among 24 leads, computational results have uncovered eight promising compounds as potential herbicides which have shown comparable physiochemical profile, better docking scores, system stability, H-bond occupancy, and binding free energy than terbutryn a reference molecule. Prioritized molecules were custom synthesized and evaluated for their herbicidal activity and specificity through whole plant assay under laboratory-controlled conditions. The prioritized lead molecule ELC5 (6-ethoxy-4-*N*-(2-morpholin-4-ylethyl)-2-*N*-propan-2-yl-1,3,5-triazine-2,4-diamine) has shown comparable activity to the reference isoproturon herbicide against *P. minor*.

1. Introduction

P. minor is the most serious grass weed of wheat; widely distributed over the rice-wheat cropping belt of north-west India [1–2]. The most threatened states of India are Punjab, Haryana, Himachal Pradesh and also a serious cause of concern in Uttar Pradesh, Uttaranchal, Madhya Pradesh and some parts of Bihar. Isoproturon has worked as the most effective herbicide to control *P. minor* growth, but it has developed resistance and thus withdrawn in 1992. Henceforth, some new alternate herbicides viz. clodinafop, fenoxaprop, sulfosulfuron, and tralkoxydim were introduced in the market in 1997-98. Application of these alternate herbicides brought the *P. minor* infestation under control for some period and helped to resume wheat production. However, cross-resistance to these four alternate herbicides had been also reported on a few biotypes [3–4]. Singh *et al.* has reported difference in RAPAD profiles of susceptible and resistant biotypes of *P. minor* [5]. Tripathi *et al.* has reported four point mutations in the D1 protein of *P. minor* which indicate target site alteration [6].

Isoproturon herbicide binds at the Q_B site of the D1 protein of PS-II located in the thylakoid membrane of the chloroplast. PS-II reaction center is a heterodimer of D1 and D2 protein encoded by the *psbA* and *psbD* gene [7]. PS-II complex binds with the reaction center pigment P680, the primary acceptor pheophytin as well as the secondary acceptors Q_A and Q_B as shown in Fig. 1. The D1 and D2 proteins of the reaction center, together bind with two pheophytin molecules and four chlorophyll-*a* molecules. The

herbicide and Q_B binding niche are prone to mutation lies in the transmembrane helix IV, and V. Horovitz et al. has reported that due to mutation of serine at 264th position is replaced by glycine in the mutant D1 protein [8]. Because of this, the herbicide molecule is deprived of one hydrogen bond as it cannot form H-bonds with glycine. This results in a considerable decrease in affinity with herbicide molecules towards the D1 niche and the normal PQ_B molecule easily replaces them from the niche thus, the normal electron transport continues in the mutant even in the presence of herbicide [9–10]. In our earlier *in silico* studies, few analogs against ACCase protein of *P. minor* were proposed [11] and few analogs were designed and tested against D1 protein wherein, triazol derivative of isoproturon was the most effective [12–13]. Kishore and Singh have reported that D1 protein of *P. minor* and wheat share high sequence similarity therefore; designing selective herbicides against this weed imposes a huge challenge [14]. Zhang et al. and Liu et al. have implemented several rational approaches for the identification of inhibitor/herbicides against the number of targets [15–16].

In this study, D1 protein has been modelled and then high throughput virtual screening, docking, *de novo* design of chemical molecules, molecular dynamic simulation, and binding free energy are carried out and compared with ref_lig to identify the potent herbicides against the D1 protein. To the best of our knowledge, we are the first to implement theoretical rational approaches like virtual screening, docking, MD simulation, and MM/PBSA calculation for D1 protein of *P. minor*.

By virtual screening, large datasets of retrieved small molecules have been screened to identify potential lead molecules and a *de novo* approach was followed up to design new analogs, against the D1 protein. The binding affinity of screened hit molecules has been tested by docking studies, and then the docked conformations were evaluated with MD simulations, and then further binding free energy of the selected molecules has been investigated. The implemented computational methods in the field of drug design provide the finest blueprint for *in vivo* experimental validation.

Later in this study, a whole plant assay [17] has been performed under laboratory-controlled conditions, to test and determine the biological activity, efficacy, specificity, and sensitivity of new molecules.

2. Materials And Methods

2.1 Template selection

Amino acid sequence of susceptible [gi|31506000|gb|AAP47827.1|psbA] and resistant D1 proteins [gi|30413115|gb|AAP33145.1|psbA] of *P. Minor* have been retrieved from the NCBI protein sequence database (<http://www.ncbi.nlm.nih.gov/protein/>). PDB BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was performed to identify the suitable template for modelling of D1 protein. The Cyanobacterial PS-II (pdb|3PRQ|A) complexes with terbutryn (triazine herbicide) at the Q_B binding site; resolution 3.20 Å; was selected as a template for building the model. The PDB ID: 3PRQ is now superseded with PDB ID: 4V82 (3PRQ file and 3PRR have been consolidated into a 4V82 for complete representation and none of the

coordinates have been changed; https://www.wwpdb.org/pdb?id=pdb_00003prq). The co-crystallized terbutryn molecule was taken as reference ligand (ref_lig) to design and discover new herbicides.

2.2 Homology modelling

The structure of the D1 protein was modelled by Modeller v9.12, a homology modelling program [18]. The modelling and validation approach is being described briefly as follows: (i) Target template alignment, (ii) Derivation of the model by MODELLER v9.12 (iii) Structure refinement, and (iv) Validation of the model. The spatial restraint method has been employed to construct the model which satisfies restraints derived from the template structure. The restraints, including bond lengths and bond angles, van der Waals contact distances, and dihedral angles, are then mapped on the target structure based on the alignment. Derived models were evaluated based on their discrete optimized protein energy (DOPE) scores and root mean square deviation (RMSDs). Energy minimization of modelled resistant D1 protein was performed for the 1000 steps of the steepest descent, then subsequently by 500 steps of the conjugate gradient method by Swiss PDB Viewer 4.10 [19]. Harmonic constraints were used for energy minimization in SPDBV. The Modelled D1 protein structure has been validated by using the structural analysis and verification server (<http://nihserver.mbi.ucla.edu/SAVES/>), which has inbuilt tools such as PROCHECK, WHAT_CHECK, WHATIF, PROSA, ERRAT, PROVE and VERIFY-3D and the superimposition between modelled D1 protein and template structure has been performed using TM-align server (<http://zhanglab.ccmb.med.umich.edu/TM-align/>) [20–22].

2.3 Lead identification

2.3.1 Preparation of virtual screening dataset

Herbicides, known for PS-II inhibition were retrieved from the international survey of herbicide-resistant weed database (<http://www.weedscience.com/summary/home.aspx>). The retrieved molecules were classified into three different classes, *i.e.* C1 (Phenyl-carbamates, Pyridazinones, Triazines, Triazinones, Triazolinone, and Uracils), C2 (ureas and amides), and C3 (nitriles and Phenyl-pyridazines). To maintain the possible diversity of the selected molecules of the aforementioned classes, all possible analogs having 70% 2D structural similarity or above were retrieved from ZINC [23] and PubChem [24] databases.

2.3.2 Virtual High Throughput screening (vHTS)

Virtual High Throughput Screening has been performed by Molegro [25], a method to identify new pharmacologically active compounds on massive trials. The vHTS has been employed to predict the putative binding affinities of small molecules with D1 protein of PS-II at the Q_B binding site. The bond length, bond order, explicit hydrogen, flexible torsion have been checked and assigned in both the protein and ligand data set. To screen the prepared ligand dataset, parameters were used are; the number of runs 50, population size 50, crossover rate 0.9, scaling factor 0.5, maximum iteration 5,000, and grid resolution 0.30 Å. The Prepared data set was taken for screening for both, rigid and flexible approaches at the center of ref_lig *i.e.* X: 20.63, Y: 65.50, and Z: 35.64 with 10Å diameter from the center of terbutryn binding. During flexible docking 43, amino acid residues were kept flexible, which lies under the 10Å diameter of

the grid. The flexible amino acids were Leu172, Phe173, Ser174, Ala175, Met176, HIS177, Ser179, Leu180, Val181, Ser183, Ser184, Tyr208, Asn209, Ile210, Val211, Ala212, Ala213, His214, Tyr216, Phe217, Arg219, Leu220, Ile221, Phe222, Gln223, Tyr224, Ala225, Ser226, Phe227, Asn228, Asn229, Ser230, Arg231, Leu233, His234, Phe235, Phe236 and Leu237. In both rigid and flexible approaches; molecules have been selected based on re-ranking, H-bond, and MolDock scores.

2.3.3 *De Novo* design of inhibitors

De novo design of new analogs has been performed by the LigBuilder v2 package [26], a tool for structure-based drug discovery. It has a CAVITY and BUILD module for pharmacophore modelling, *de novo* synthesis of ligands, and processing of synthesized molecules. The Ligand detection mode of the CAVITY module has been used to find the key interaction residues at the binding site with the minimal distance between any two pharmacophore features (not to be less than 3Å) and the binding site was selected based on ligandability less than or equal to 6.0Kd (equivalent to 1µM). The EXTRACTION submodule was used to design seed molecules to ensure the binding affinity that decides the manner in which ligand would grow and the growing mode of BUILD module was used to construct molecules in the binding pocket. The whole construction process is controlled by a genetic algorithm. Protein-ligand binding affinity is evaluated by using an empirical scoring function; chemscore [26]. Besides binding affinity, the synthesizability of the ligands is also taken into account by applying certain sets of chemical rules (LigBuilder, <http://www.ligbuilder.org/>).

2.3.4 Clustering of small molecule

Ligands obtained from vHTS and *de novo* design have been clustered separately in a hierarchical fashion by the ChemAxon's JKlustor Suite. The MCS tool of JKlustor has been used to search for the maximum common substructures (MCS) of a compound library hierarchically. Clustering in a hierarchical fashion identifies the largest substructure shared by several molecular structures (JKlustor<https://www.chemaxon.com/products/jklustor/>). All the vHTS and *de novo* synthesized ligands have been clustered into 13 and 11 groups respectively. These molecules were clustered to retain at least one typical molecule of each class of herbicides. The vHTS and *de novo* clusters are further named as existing ligands clusters (ELC) and *de novo* synthesized ligand clusters (DLC) respectively.

2.4 Prioritization of leads

Docking has been performed for both ELC and DLC lead molecules at the binding site (terbutryn binding site) of D1 protein using AutoDock [27]. AutoDockTools (ADT) was used to select the parameters and to prepare the ligand and protein files. Rigid docking was followed by flexible docking, over screened molecules (screened through vHTS) of each cluster of the ELC and DLC. The grid parameter file was prepared by setting the dimensions of the grids for resistant D1 proteins, 40:40:40 Å, and spacing 0.375 Å with the center coordinates (20.63, 65.50, and 35.64). Grid point spacing was calculated for each of the atom types of protein and ligand (protein: A (aromatic C), C, HD, N, NA, OA, and SA; ligand: C, A, HD, N, NA, e (electrostatic), and d (desolvation)). The AutoDock run parameters were set as maximum iterations of energy evaluations, iterations of genetic algorithm, and number of GA per run were set to 2,500,000; 27,

000, and 100 respectively. The Lamarckian Genetic Algorithm (LGA) was selected for the ligand conformational search. Moreover, the best molecule of each cluster has been selected based on binding energy, ligand efficiency, and inhibition constant from each ELC and DLC clusters. Finally, twenty-four leads were selected (13 from ELC and 11 from DLC).

2.5 Lead validation through Molecular Dynamics Simulation

Molecular Dynamics (MD) simulation has been performed using Gromacs4.6.5 [28] to validate the 24 prioritize leads. The stability of the D1 protein; complexed with ref_lig (docked complex) was evaluated in the lipid bilayer. The 24 leads were evaluated in two sets. In the first set, 13 ELC and in the second set, 11 DLC ligands were docked with D1 protein and evaluated. The topology files for the ELCs and DLCs ligands have been generated using the PRODRG server [29] in the framework of GROMOS 53A6 force-field. In order to impersonate the real environment, simulation of protein-lead complexes has been carried out in DPPC (dipalmitoylphosphatidylcholine) lipid bilayer environment [30]. The desired orientation of the D1 protein-ligand complex in lipid bilayer was taken from the Orientations of Proteins in Membranes (OPM) database (<http://opm.phar.umich.edu>). The bilayer was then expanded by translating lipid molecules and scaling the lateral dimensions of the box size using a scaling factor 4 and a shrinking step was performed to scale down the lipids by a factor of 0.95 to avoid the overlapping of the atoms in the system. During the scaling of the lipid bilayer, there were 22–23 steps of shrinking and minimization was performed to reach an area per lipid of $\sim 71 \text{ \AA}^2$, (Close to the experimental value of $\sim 62 \text{ \AA}^2$). Lipid parameters and topology files were obtained from Tieleman's website (<http://moose.bio.ucalgary.ca>) [31]. Afterward, the DPPC head groups were hydrated by adding sufficient solvent molecules on either side of the lipids (unit cell dimension $8.30626 \text{ \AA} \times 8.33877 \text{ \AA} \times 6.59650 \text{ \AA}$). The resulting simulation system contained 125 DPPC molecules (63 molecules in the upper leaflet and 62 in the lower leaflet) solvated on both sides with SPC water. The complete system contains ~ 21000 atoms. A periodic boundary condition was employed to avoid edge effects. At physiological pH, the protein was having a negative charge, thus, to make the system electrically neutral, 10 Na^+ ions were added in the simulation box. The system was relaxed by energy minimization (10,000 steps of the steepest descent and 10,000 steps of the conjugated gradient) to remove bad van der Waals contacts while all protein and ligands were harmonically restrained to their original position with an isotropic force constant of $1000 \text{ kJ.mol}^{-1}.\text{nm}^{-2}$. The above restraints allow for further equilibration of the bilayer and solvent while keeping the conformation of protein and ligand unchanged. The bilayer, solvent, and protein-ligand were separately coupled to the temperature and pressure bath. Temperature and pressure were maintained constant at 300 K and 100 kPa using weak coupling technique with coupling and relaxation parameters of 0.1 and the 1.0 Ps respectively [32]. The resulted system was subjected to production MD simulation for 240 ns (10 ns for each complex) using GROMOS 53A6 force-field [33] and the linear constraint solver (LINCS) algorithm [34] was employed for covalent bond constraints. Results were analyzed by extracting information from trajectory files obtained during the simulation. The root mean square deviation (RMSD), root mean square fluctuation (RMSF), and hydrogen bonds were analyzed and graphs and images produced using Xmgrace and Visual Molecular Dynamics (VMD) [35].

2.6 Binding free energy calculation

Binding energy (ΔG_{bind}) of D1 protein complexes of ELCs and DLCs leads was estimated from the trajectory file of each complex obtained in MD simulation using the molecular mechanics-Possion-Boltzmann surface area (MM-PBSA) method [36]. The polar solvation energy values were calculated using the implicit solvation Poisson–Boltzmann model (PB) and nonpolar solvation energy values were calculated based on the solvent accessible surface area (SASA). Finally, all these were combined to obtain absolute, binding free energy value. The binding free energy was calculated by Eq. 1.

$$\Delta G_{\text{bind}} = \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} + \Delta G_{\text{PB}} + \Delta G_{\text{sur}} + -T\Delta S \quad (1)$$

Where, ΔE_{vdw} and ΔE_{elec} represent van der Waals interactions and coulombic interactions respectively. ΔG_{PB} and ΔG_{sur} represent polar and nonpolar solvation energies, respectively. The entropic term $T\Delta S$ was computed in the normal mode approximation [37].

2.7 Theoretical study of physiochemical properties of known herbicides of similar group:

During the experimental testing, most of the molecules fail due to their poor physiochemical properties. Therefore, properties of existing herbicides have been investigated using Drug Likeness Tool, DruLiTo, (http://www.niper.gov.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html).

2.8. Whole plant bioassay for the herbicidal activity of test molecules

2.8.a. Testing the bioactivity of test molecules

The proposed prioritized molecules have been synthesized and characterized by Alba Nova (Alba Nova Life Science Private Limited, Hyderabad). The molecules were high in purity which was directly employed for testing. Some of the prioritized molecules were excluded due to the high-cost index of synthesis. The efficacy and specificity of custom synthesized molecules have been evaluated through whole plant assay. The experiment was designed in triplicates under laboratory-controlled conditions. The test has been conducted in plastic petri-plates (9cm diameter). Ten seeds of *P. minor* as well as wheat have been arranged on the moist double layered Whatman filter paper in each plate. The prepared plates have been placed in the plant growth chamber for germination of seeds, at 17°C in dark conditions, and relative humidity of 45% for six days. An aliquot of distilled water has been continuously supplied to petri plates to keep the germinating seeds hydrated. The test herbicides have been evaluated on *P. minor* at a concentration of 600 μM , along with control, isoproturon treated, and untreated control in triplicates. Isoproturon herbicide is taken as control/reference for bioactivity testing since terbutryn has an adverse effect (non-selective) over wheat plantlets; a non-selective herbicide. After six days (post germination of seeds) the seedlings in petri plates have been subjected to test as accordingly mentioned factors. The seedlings have been provided with added 2mL of respective solutions (distilled water in control,

isoproturon in treated control, and test molecules in test treated population) at an interval of two days, till a total 10 mL of the solution has been supplied to each plate, followed by adding an aliquot amount of distilled water for hydration. Petri plates have again been incubated in plant growth chambers under the same optimal conditions but with altering light conditions (8/16h day/night cycle) for the next fifteen/sixteen days. Afterward, observations regarding the effect of the compound upon plantlet growth, biomass accumulation, and other physical changes that occurred in plantlets due to treatment, have been noted. Further data of wet and dry weight of plantlets have also been tabulated to compare the effect of compounds through biomass accumulation. Later, the molecule with the most effective efficacy and specificity towards the said weed has been determined by comparing the results with the reference molecule isoproturon.

2.8.b. Testing the efficacy and selectivity of test molecule in laboratory controlled condition

The different subjects considered in this test were the same as earlier mentioned (untreated control population, treated control population with isoproturon as a reference, and treated populations with synthesized lead molecules). Further, the effective molecule was tested in triplicate to determine the concentration at which the growth rate of weed plantlets gets reduced by 50% (GR₅₀). The testing has been performed on a concentration of 0-600 µM at a regular interval of 40 µM in order to determine meticulous GR₅₀ dose of the new effective herbicide molecule.

3. Results And Discussion

3.1 Homology modelling

Resistance D1 protein (gi|30413115|gb|AAP33145.1) was taken as a query sequence to find out the most suitable template and it has shown significant alignment with chain-A of the photosynthetic oxygen evolving center (1S5L) having a score: 503, Query coverage: 99%, identity: 90% and E value: 1e-178 and resolution 3.5Å. Cyanobacterial D1 protein (pdb|3PRQ) complex with terbutryn at the QB binding site has also shown significant alignment; *i.e.* query coverage of 100%, identity 90%, similarity 95.2%, and E-value 1e-178 without any gap in the alignment and resolution 3.2 Å (Fig. 2). The crystal structure of Cyanobacterial PS-II consists of 20 polypeptides along with 88 ligands and 10 cofactors [38] and it was selected as a template protein for modelling. As a result, 10 models were built and selected based on their discrete optimized protein energy (DOPE) scores and root mean square deviation (RMSDs). A short 1000 and 500 step of the steepest descent and conjugate gradient minimization have been performed respectively to minimize the strain of modelled D1 protein Fig. 2.

The modelled D1 protein has shown phi and psi torsion angle of 93.6% amino acid residues in core regions and 6.4% amino acid residues in the allowed region, whereas the template structure showed 82.3% amino acid residues in the core region and 17.0% amino acid residues in the allowed region (Table S1 and Figure S1). The quality of the model was also validated by the ProSA server, a web server for protein structural analysis. Moreover, The ProSA Z-score was - 4.0 which indicates the overall quality and

measures the deviation from the total energy of the structure with respect to an energy distribution derived from random conformations. A negative PROSA score indicates the correctness of the modelled protein. Verify 3D was used for validation of the modelled protein and it showed an average 3D-1D profile score - 0.09 and 0.52. TM-align server was used to further validate the modelled protein which showed TM score 0.99125 ($0.5 < \text{TM Score} > 1$, indicates about the same fold), suggesting the modelled protein to be of good quality, as shown in Fig. 3. Thus, results suggest that the comparative modelled structure of D1 protein is of good quality, and could be used for further studies. The homology modeling was performed to get the blueprint of the binding mode of herbicide terbutryn complexed with D1 protein (PDB ID3PRQ/4V82) which is precisely located at the QB site of D1 protein for the design and development of herbicides. The target template identity is ~ 90% without any gap so the model structure could be considered as equivalent to crystallographic structure except few side chains [39].

In a previous study, our group has built a chimeric homology model of the PS-II reaction center of *P. minor*. The D1 protein is truncated at both N and C terminal, so both missing ends were concatenated with ref sequence of *T. aestivum*'s D1 protein. Sequence for D2 protein was taken from its nearest homolog *i.e.* *T. aestivum* (from Ref Seq Database). The build reaction center was having one β -carotene, one bicarbonate ion, five chlorophyll-a, two pheophytin-a, two plastoquinone) and cofactors (iron and OEC) along with D1 and D2 protein. To check the integrity and stability of the build chimeric model it was subjected for MD simulation, and we reported that the complete reaction center deviates by 1.7175 Å RMSD from its equivalent bacterial reaction center *i.e.* template; PDB ID 2AXT [13].

3.2 Virtual High Throughput screening (vHTS)

The stringent criteria were applied in vHTS to narrow down the number of molecules; the vHTS procedure is shown in Fig. 4. vHTS for all retrieved 33016 ligands (Table S2) have been performed at the ref_lig binding site of modelled D1 protein and 7250 ligands were selected based on rerank, H-bond and MolDock score either more or equal to ref_lig molecule score Table S4. To enrich the screening, selected ligands were passed through three times flexible docking and at least one molecule from each class and subclass of C1, C2, and C3 group has been retained. Best screened ligands were selected based on the rerank score (best 5% ligands), H-bond and MolDock score either more or equal to ref_lig score.

3.3 Designing of new analogs at the binding site of D1 protein

The *de novo* drug design program Ligbuilder v2 was used to design the new analogs of selected ligands iteratively using an inbuilt library of organic fragments. Overall 28 ligands, *i.e.* 16, 7, and 5 ligands from class C1, C2, and C3 were set for the *de novo* synthesis (Table S3). Ligand detection mode of the CAVITY module was used to detect the binding site of the protein and the best binding site was decided based on ligandability (Fig. 5). These were compared with ref_lig binding sites in D1 protein. The automatic extractor module was used which has extracted 71, 29, and 19 seeds from C1, C2 and C3 class respectively (Table S3). Extracted seed molecules were initially used for growing to design new ligands molecules by the BUILD module. BUILD module uses a pocket and grid file generated by the CAVITY

module which acts as pharmacophore for *de novo* synthesis. Here, molecules were synthesized considering the Lipinski rule of five, chemscore, binding affinity, synthesizability, and toxicity parameter for all molecules as well as the molecules were represented with the route of synthesis with cost index. As a result, 282 ligands were synthesized which belongs to C1 (169), C2 (72), and C3 (41) classes.

3.4 Clustering

The hierarchical clustering of both vHTS ligands (230) and *de novo* synthesized ligands (282) were performed to get 13 and 11 clusters respectively (Table S3). The vHTS and *de novo* clusters are named as existing ligands clusters (ELC) and *de novo* synthesized ligand clusters (DLC) respectively. Screened leads from individual clusters of ELC and DLC (Table S3) were selected for the docking study.

3.5 Docking

The ELC and DLC hits were docked at the center coordinates (X: 20.63, Y: 65.50, and Z: 35.64) of the binding site of modelled D1 protein using AutoDock4.2 as well as the Molegro tool. The criteria chosen to select the best pose of each hit were AutoDock term (Binding energy, Ligand efficiency, and inhibition constant) and Molegro term (MolDock score, H-bond score and rerank score). One hundred poses per ligand were generated and clustered with an RMSD tolerance limit of 1.5 Å. Binding poses of all hits of the ELC and DLC were selected from the most populated cluster and a summary of these hits are tabulated in Table S5 and Table S6. Thirteen hits from ELC and eleven hits from DLC were selected from each cluster of the ELC and DLC respectively after employing the docking score filter. Three amino acids, *i.e.* His177, Phe227, and Phe236 in binding sites play a crucial role in interaction with hits. Docking results have shown that ref_lig has two H-bonds and one pi-pi interaction in the binding site of protein with ligand efficiency, -0.25, and inhibition constant 1.08 mM (Table S4 & Figure S2).

The ref_lig docking scores were set as criteria to select the best hits from each cluster, as a result, thirteen and eleven hits selected from ELC and DLC clusters. In addition to ref_lig based criteria, two or more H-bonds and one pi-pi interaction with better binding energy, ligand efficiency, and inhibition constant were employed as selection criteria to select the molecule (Table S5-S6 & Figure S3-S6).

These lead molecules were selected through the rigorous procedure and several attributes were analyzed. It is found that among the extracted leads, most of them have one ring common, having at least one heteroatom with molecular weight, donor site, and acceptor site in the range of 140 to 315, 1 to 4, and 4 to 8 respectively.

3.6 Molecular Dynamics Simulation

To find the binding affinity and binding mode of ligand, the rigid docking followed by flexible docking was applied, but it does not properly sample the conformational space that could lead to a variety of different binding modes of the ligands [40]. The goals of these simulations were to observe the structural behaviour of protein (flexibility of the D1 protein with ref_lig, ELC, and DLC leads) and ligands (the behaviour of leads inside the binding site of D1 protein). MD simulation was performed for each lead protein complex for 10ns using Gromacs4.6.5 in an environment of a lipid bilayer embedded in a periodic

box solvated with water shown in Fig. 6. PS-II complex has 20 peptides and 95 Co-factors (PDB ID 4V82). D1 protein complexes with at least one β -carotene, one pheophytin, two α -chlorophylls, two plastoquinone, at least two Fe²⁺ ions, and one bicarbonate. The PS-II reaction center proteins and associated molecules of the complex in a tight-knit providing stability to each other, and having interlinked functions. In our previous attempt, a chimeric D1/D2 protein was constructed from its nearest homolog and subjected to MD simulation. The chimeric D1/D2 protein has shown RMSD of 1.7175 Å from its templates protein (Singh et al 2012). Therefore, we have refrained from doing any long simulation with the D1 subunit in this study. RMSD curves of the protein backbone; energy profile and H-bond profile of each protein-lead complexes were calculated and compared with the ref_lig protein complex. Variation in the total energy plot of the ref_lig protein complex has shown that the overall energy tends to stabilize after 3ns of simulation (Fig. 6B), which confers the average energy remains constant, thus suggesting structural stabilization. The total energy patterns were similar for all ELC and DLC lead protein complexes, (data are not shown).

Throughout the simulation period, no significant fluctuations were observed in the backbone of protein with ref_lig. It implies that the binding of ref_lig at the binding site of protein is not only stable and strong but also does not perturb the protein backbone stability. The binding stability of the ELC leads (ELC2, ELC3, ELC4, ELC5, ELC6, ELC10, ELC11, and ELC12) and DLC leads (DLC4, DLC6, DLC7, and DLC8) in the binding site of the protein which, have shown comparable fluctuation with ref_lig protein complex throughout the simulation in the size of the noticeable window of 0.30–0.40 nm Fig. 7. It ensures and suggests that these leads are well accommodated inside the binding site throughout the simulation. H-bonds at the binding site of the ref_lig protein complex were determined which revealed that it encompassed four H-bonds throughout the simulation period. The amino acid residues involved in the key H-bond formation at the binding site were Ala225, Ser226, and Phe227. The succession of frames for ELC and DLC leads inside the binding site of the protein have shown that ELC leads ELC2, ELC3, ELC5, ELC7, ELC8, ELC11, and ELC13 and DLC leads DLC3, DLC4, DLC6, DLC8, and DLC11 comprises three to four H-bonds throughout the simulation as shown in Fig. 8. The amino acid residues involved in the key H-bonds in the binding site were His177, Ala225, Ser226, Phe227, Asn229, and Ser232 throughout the simulation. It seems that these leads are also interacting in a similar fashion as ref_lig; which confirms the accommodation of these leads in the protein binding site because ref_lig fitted well into the binding site of protein over the simulated time.

3.6.1 Amino acid fluctuation analysis

The root mean square fluctuation (RMSF) analysis revealed the possible movements of amino acid residues Phe201, Phe222, Tyr224, and Phe227 (inside the binding pocket) and Arg26, Gln127, Arg187 (out the binding site) in the ref_lig ligand-protein complex (Fig. 9). In the case of selected ELC and DLC lead complexes, RMSF appears to be almost similar to the ref_lig except for the other small fluctuations. This signified that protein-ELC and protein-DLC lead complexes are also stable like the ref_lig-protein complex. The amino acid residues inside the binding sites have shown slightly less fluctuation because of the formation of a *pi-pi* interaction between the leads and the aromatic ring of Tyr224 and Phe227. It might

be fluctuating to stabilize the protein lead interaction. Moreover, residues of the binding site showed lower fluctuation, revealing to be a more stable protein-ligand system.

3.6.2 Analysis of hydrogen bond occupancy

The MD trajectories were analyzed for the H-bonds distance and the H-bond occupancy. H-Bond distance is determined by the bond length ($< 3.5 \text{ \AA}$) and the H-bond occupancy is defined as the time percent existence of a specific H-bond during the simulation. The H-bonds predicted for the ref_lig, ELC, and DLC lead and amino acid residues of the binding site of protein were observed consistently for the major part of trajectory with high occupancy for all the proposed lead molecules. H-bond interactions were categorized as strong ($> 60\%$), medium ($40-60\%$), and weak ($< 40\%$) based on H-bond occupancy with ref_lig and the binding site of the protein. Four H-bond were found between ref_lig and protein, *i.e.* O@Ala225...HAG@ref_lig, HG@Ser226...NAL@ref_lig, O@Phe227...HAG@ref_lig, and H@Phe227...NAL@ref_lig with occupancy 63.0%, 59.9%, 55.2%, and 60.3%, respectively, and the distance of H-bonds was fluctuating in windows of 1 \AA to 3.5 \AA . Occupancies of the H-bonds between protein, ELC, and DLC lead molecules showed a diverse range of fluctuation which is comparable to ref_lig. ELC leads ELC2, ELC3, and ELC8 have shown four H-bonds, where two H-bonds are having occupancy $> 70\%$ and another two H-bonds having occupancy 70 to 30%. ELC4 and ELC7 showed three stable H-bonds with significant occupancy $> 40\%$. DLC leads DLC3, DLC6 and DLC9 showed four H-bonds occupancy $> 40\%$. DLC4 and DLC8 ligands also have shown two stable H-bonds with significant occupancy. Other ELC and DLC lead molecules (except above described ELC and DLC lead molecules) have not shown significant stable H-bond as compared to ref_lig. The occupancies of the H-bonds analysis also showed that amino acid residues His177, Ala225, Ser226, Phe227, and Asn229 were involved in the H-bonding with ref_lig, ELC and DLC lead molecules. However, the residues Ser226 and Phe227 were obtained as key residues to hold the molecules at the binding site as these residues showed higher occupancy in most of the ELC and DLC lead molecules.

3.6.3 Free energy calculations

Binding free energy of D1 protein complexed with ref_lig, ELC, and DLC molecules was determined by the MM/PBSA method. In the case of ref_lig, van der Waals interactions (ΔG_{vdw}), showed the most significant contributions among the van der Waals interactions (ΔG_{vdw}), the electrostatic energy (ΔG_{ele}), and the nonpolar solvation free energy ($\Delta G_{\text{nonpol, soil}}$). The results reveal that leads from ELC class (ELC1, ELC2, ELC4, ELC5, ELC8, ELC11, ELC12, and ELC13) and lead from the DLC class (DLC1, DLC3, DLC4, and DLC6) displayed higher binding free energy compared to ref_lig binding free energy ($-104.627 \pm 10.792 \text{ kJ/Mol}$). The order of binding free energy of ELC's and DLC's is ELC4 $>$ ELC 11 $>$ ELC 8 $>$ ELC 13 $>$ ELC 5 $>$ ELC 12 $>$ ELC1 $>$ ELC2 and DLC6 $>$ DLC4 $>$ DLC1 $>$ DLC3 respectively. Binding energies of the ELC and DLC lead molecules are shown in Table 1, which reveals that ΔG_{vdw} has the most significant contribution to the binding energy.

Table 1

The Table shows the van der Waals interaction energy (ΔE_{vdW}), electrostatic energy (ΔE_{elec}), polar solvation energy (ΔG_{polar}), nonpolar solvation energy ($\Delta G_{\text{nonpolar}}$), and binding energy ($\Delta G_{\text{binding}}$) of ref_lig molecule, ELC lead, and DLC lead protein complexes. The energy values are average values in KJ/Mol which were determined by the MM/PBSA method.

System	$^a\Delta E_{\text{vdW}}$	$^a\Delta E_{\text{elec}}$	$^a\Delta G_{\text{polar}}$	$^a\Delta G_{\text{nonpolar}}$	$^a\Delta G_{\text{binding}}$
Ref_lig					
Ref_lig	-136.282 +/- 10.140	-8.184 +/- 3.246	53.919 +/- 6.532	-14.080 +/- 0.933	-104.627 +/- 10.792
ELC lead					
ELC1	-136.168 +/- 18.364	-10.486 +/- 10.291	52.763 +/- 14.004	-12.773 +/- 1.491	-106.664 +/- 16.450
ELC2	-133.086 +/- 10.140	-63.940 +/- 7.822	105.192 +/- 6.745	-13.402 +/- 0.798	-105.236 +/- 8.860
ELC3	-135.888 +/- 11.678	-23.117 +/- 10.308	79.837 +/- 10.864	-13.025 +/- 1.037	-92.192 +/- 9.491
ELC4	-226.920 +/- 13.782	-30.143 +/- 10.856	92.789 +/- 10.406	-20.306 +/- 1.053	-184.580 +/- 13.217
ELC5	-152.496 +/- 9.783	-8.146 +/- 2.641	53.635 +/- 5.890	-16.407 +/- 0.979	-123.415 +/- 9.441
ELC6	-79.543 +/- 9.136	-11.230 +/- 3.373	36.082 +/- 5.416	-7.967 +/- 1.059	-62.658 +/- 8.925
ELC7	-90.852 +/- 12.170	-30.849 +/- 6.863	77.723 +/- 8.950	-11.725 +/- 1.164	-55.703 +/- 10.491
ELC8	-172.742 +/- 13.993	-43.575 +/- 7.894	90.215 +/- 8.879	-15.768 +/- 1.331	-141.870 +/- 13.809
ELC9	-127.335 +/- 14.787	-27.564 +/- 8.311	66.847 +/- 11.559	-14.443 +/- 1.138	-102.495 +/- 11.241
ELC10	-108.888 +/- 22.820	-3.367 +/- 3.659	28.827 +/- 10.771	-12.242 +/- 2.191	-95.670 +/- 16.687
ELC11	-193.160 +/- 25.566	-1.984 +/- 1.230	40.587 +/- 8.122	-17.862 +/- 1.535	-172.420 +/- 21.020
ELC12	-140.765 +/- 9.307	-5.079 +/- 2.853	48.274 +/- 5.784	-14.356 +/- 0.834	-111.926 +/- 8.968
ELC13	-156.923 +/- 15.543	-34.140 +/- 6.858	75.406 +/- 8.382	-15.46 +/- 0.866	-131.119 +/- 13.870
DLC lead					

System	$^a\Delta E_{vdW}$	$^a\Delta E_{elec}$	$^a\Delta G_{polar}$	$^a\Delta G_{nonpolar}$	$^a\Delta G_{binding}$
DLC1	-135.061 +/- 13.621	6.168 +/- 6.265	31.185 +/- 9.587	-13.956 +/- 1.233	-111.665 +/- 12.895
DLC2	-107.463 +/- 11.009	-12.515 +/- 5.824	47.658 +/- 6.065	-11.334 +/- 0.920	-83.655 +/- 10.542
DLC3	-154.59 +/- 12.638	-28.974 +/- 7.750	87.996 +/- 7.825	-13.844 +/- 0.978	-109.416 +/- 11.518
DLC4	-162.907 +/- 12.942	-21.123 +/- 13.218	73.483 +/- 10.909	-16.567 +/- 1.199	-127.114 +/- 13.063
DLC5	-119.804 +/- 15.573	-6.646 +/- 3.462	39.220 +/- 8.657	-13.798 +/- 1.281	-101.029 +/- 12.107
DLC6	-192.460 +/- 15.952	-37.893 +/- 7.908	97.723 +/- 11.232	-17.244 +/- 0.945	-149.874 +/- 13.212
DLC7	-122.524 +/- 8.754	-16.470 +/- 8.328	75.856 +/- 9.176	-14.270 +/- 1.021	-77.407 +/- 9.632
DLC8	-90.167 +/- 14.000	-25.311 +/- 8.323	59.914 +/- 12.091	-10.921 +/- 1.050	-66.486 +/- 10.326
DLC9	-127.765 +/- 16.923	-25.236 +/- 19.172	71.181 +/- 16.018	-13.815 +/- 1.181	-95.635 +/- 12.928
DLC10	-110.780 +/- 8.874	-27.165 +/- 11.678	78.280 +/- 18.602	-11.361 +/- 0.621	-71.026 +/- 10.219
DLC11	-108.314 +/- 9.016	-12.881 +/- 8.995	53.002 +/- 6.521	-10.721 +/- 0.591	-78.915 +/- 8.438

From Table 1, it can be seen that ELC1, ELC13, DLC1, and DLC3 showed better binding affinity than ref_lig. However, these lead molecules are discarded in the broad-spectrum purpose as either the number of H-bonds was decreased from the initial structure or the H-bond occupancy was obtained lower than the other lead molecules. Therefore, the lead molecules ELC2, ELC4, ELC5, ELC8, ELC11, ELC12, DLC4, and DLC6 have been selected. These selected ELC and DLC lead molecules and the other known herbicides for D1 protein of PS-II protein were undertaken to predict the physiochemical properties (molecular weight, logP, AlogP, number of hydrogen donors and acceptors, total polar surface area (TPSA), number of rotatable bonds, number of acidic groups, number of rigid bonds, number of aromatic rings and number of hydrogen bonds) (Fig. 11, Table S7).

From the data presented in Fig. 11 Table S7, it is remarkable that the properties of all the proposed molecules are comparable to known PS-II herbicides and having no toxic functional groups. Therefore, proposed molecules (ELC2, ELC4, ELC5, ELC8, ELC11, ELC12, DLC4, and DLC6) considered as potential lead molecules for the experimental testing and further authentication (structures are shown in Fig. 12); ZINC, PubChem ID, and IUPAC name are shown in Table S8. However, only four molecules (ELC8, ELC5, ELC11, and ELC12) were synthesized among the proposed eight molecules. Other molecules were not

synthesizable due to the high polarity of intermediates, thereby increasing the cost of synthesis which would not be viable as an herbicide.

Table 2
Synthesized molecules for laboratory testing.

Sl. no.	Prioritized small molecules	Code	PubChem/ZINC ID.
1.	4-chloro-2,6-bis[(5-chloro-2-hydroxyphenyl)methyl]phenol	ELC11	PubChem ID.- 62616 ZINC ID.- 1679491
2.	1-benzyl-3-(5-tert-butyl-1,2-oxazol-3-yl)-1-methylurea	ELC8	PubChem ID.- 20330566 ZINC ID.- 59961084
3.	6-ethoxy-4-N-(2-morpholin-4-ylethyl)-2-N-propan-2-yl-1,3,5-triazine-2,4-diamine	ELC5	PubChem ID.- 767097 ZINC ID.- 19520234
4.	2-[[4-chloro-6-(2-hydroxyethylamino)-1,3,5-triazin-2-yl]amino]-2-methylpropan-1-ol	ELC12	PubChem ID.- 22333149 ZINC ID.- 6721886

3.7. Whole-plant bioassay

3.7.a. Testing the bioactivity of synthesized molecules

The tabulated data of untreated control, treated control with isoproturon (ref, and treated with test molecules (Table 2) were compared to interpret the effect of test molecules against *P. minor* and wheat with respect to the solvent (2% DMSO) and distilled water. Non-treated seedlings have a healthy growth rate as compared to treated seedlings. Besides late germination in treated seeds, yellowing of leaflets of treated plantlets has been observed after one week of treatment. Progression in the yellowing of leaflets has been observed in the treated plantlets, reflecting the visible chlorosis in treated plantlets which could be an indirect reflection of hindrance in photosynthesis thereby loss in biomass accumulation. At the end of the experiment, biomass accumulation in plantlets has been tabulated. Biomass accumulation has been compared among the test molecules and isoproturon (ref), to determine the most efficient molecule among the four test molecules graph has been plotted (Fig. 13, Table S9). From the graph Fig. 13, it has been recorded that there is no effect of either control isoproturon or test molecules upon wheat plantlets. However, it has been observed that there is a selective effect of test molecules upon *P. minor*, with different activity range *i.e.* ELC5 > ELC12 = ELC8 > ELC11. The efficacy and selectivity of ELC5 (6-ethoxy-4-

N-(2-morpholin-4-ylethyl)-2-N-propan-2-yl-1,3,5-triazine-2,4-diamine) is comparable as isoproturon. Thus, it has been further tested and evaluated to determine the GR₅₀ of ELC5 molecule.

3.7.b. Testing the efficacy and sensitivity of ELC5 molecule in laboratory-controlled condition

The ELC5 has been further tested to evaluate its activity at different concentrations (0-600 µM at a regular interval of 40 µM), to determine its GR₅₀. The plot (Table S10) decipher that there is no effect of either isoproturon or test molecule ELC5 upon wheat at any concentration, thus promising the selective activity of ELC5 against *P. minor* only. The tabulated data of biomass has been statistically analyzed using a descriptive statistics (Confidence level: 95%) approach to verify the truthfulness of the results. Upon further evaluation of tabulated data (Table S10), it has been interpreted that upon the treatment of plantlets with ELC5 there is ~ 50% reduction in biomass of *P. minor* plantlets at the concentration of ~ 40 µM (Fig. 14).

ELC5 (6-ethoxy-4-N-(2-morpholin-4-ylethyl)-2-N-propan-2-yl-1,3,5-triazine-2,4-diamine) molecule has shown comparable activity to isoproturon a reference molecule against *P. minor* under laboratory-controlled conditions. It could be further taken for pot assay and field assay (pilot-scale synthesis and testing for various parameters) to test its effectiveness, efficacy, selectivity, and sensitivity in a natural uncontrolled environment. The outcome of the assays could suggest whether the molecule is ready to use or need some chemical modification and optimization.

4. Conclusion

Herbicide resistance to *P. minor* is one of the major causes of reduced wheat production, in the wheat cropping belts in India as well as in other countries. Due to the development of resistance to existing herbicides, it is of utmost importance to find a novel herbicide to restore wheat production. *In silico* studies, *i.e.* molecular modelling, virtual screening, *de novo* synthesis, docking as well as molecular dynamics simulation and MM/PBSA calculation were performed to find molecules with a high binding affinity at the Q_B site of D1 protein and compared to ref_lig (terbutryn; complexed with D1 protein PDB:3PRQ). The screened small molecules from large datasets and *de novo* synthesized molecules were taken for docking studies, molecular dynamics simulation, and MM/PBSA calculation. From the result obtained, ELC2, ELC4, ELC5, ELC8, ELC11, ELC12, DLC4, and DLC6 have shown better features in all aspects in comparison to ref_lig. These molecules were selected based on protein backbone stability with lead molecules, the number of H-bond and their occupancy, non-bonded interaction, and binding energy. These theoretical results suggest that the selected molecules are accommodated well inside the binding site throughout the MD simulation, which reveals the stability of the system, later confirmed by the RMSD analysis. H-bond occupancy and the strength of binding were evaluated by binding energy and physiochemical profiles are comparable to the other PS-II herbicides. The prioritized molecules were synthesized and tested in whole plant assay in laboratory-controlled conditions. The whole plant assay has shown ELC5 (6-ethoxy-4-N-(2-morpholin-4-ylethyl)-2-N-propan-2-yl-1,3,5-triazine-2,4-diamine) is equally effective as isoproturon (GR₅₀ at 40 µM) and has shown no adverse effect over wheat plantlets.

The newly discovered lead molecule mentioned formerly, is proposed for pilot scale testing in pot and crop fields.

Declarations

Funding: The authors acknowledge the funding supported by the Science and Engineering Research Board, India file No.YSS/2015/001662.

Conflicts of interest/Competing interests: The authors of this manuscript declare that there is no conflict of interest:

Availability of data and material: Supplementary information is available for this paper.

Code availability: Not Applicable

Authors' contributions: D. V. Singh conceived and supervised the study. N. K. and P. R. performed the calculations and experiments. N. K. and S. A. interpreted the results and writing of the manuscript is done by all the authors.

Acknowledgement

The authors of this manuscript also acknowledge to Prof. Krishna Mishra, Indian Institute of Information Technology Allahabad for her support in the completion of this research work.

References

1. Chhokar RS, Sharma RK (2008) Multiple herbicide resistant in little seed canary grass (*Phalaris minor*): A threat production in India. *Weed Biology Management* 8:112–123
2. Chhokar RS, Sharma RK, Sharma I (2012) Weed management strategies in wheat-A review. *Journal of Wheat Research* 4:1–21
3. Yadav A, Malik RK (2005) Botanical Characteristics, Biology and Distribution in Herbicide Resistant *Phalaris minor* in Wheat - A Sustainability Issue. Hisar, India Department of Agronomy and Directorate of Extension Education. CCS Haryana Agricultural University, pp 5–15
4. Yadav A, Malik RK (2012) Alternate Herbicides against Isoproturon Resistant *Phalaris minor* in Herbicide Resistant *Phalaris minor* in Wheat- a Sustainability Issue. Department of Agronomy and Directorate of Extension Education CCS Haryana Agricultural University, Hisar, pp 89–98
5. Singh DV, Gaur AK, Mishra DP (2004) Biochemical and Molecular Mechanisms of Resistance Against Isoproturon in *Phalaris minor*: Variations in Protein and RAPD Profiles of Isoproturon Resistant and Sensitive *Phalaris minor* Biotypes. *Indian Journal of Weed Science* 36:256–259
6. Tripathi MK, Yadav MK, Gaur AK, Mishra DP (2005) PCR based isolation of *psbA* (herbicide binding protein encoding) gene using chloroplast and genomic DNA from *Phalaris minor* biotype(s). *Physiology Molecular Biology of Plants* 11:161–163

7. Marder J, Chapman DJ, Telfer A, Nixon PJ, Barber J (1987) Identification of psbA and psbD gene products, D1 and D2, as reaction centre proteins of photosystem 2. *Plant Mol Biol* 9:325–333
8. Horovitz A, Ohad N, Hirschberg J (1989) Predicted effects on herbicide binding of amino acid substitutions in the D1 protein of photosystem II. *FEBS Lett* 234:161–164
9. Hirschberg J, Bleeckes A, Kyle DJ, McIntosh L, Arntzen CJ (1984) The molecular basis of triazine resistance in higher plant chloroplasts. *Natur Forch* 39:412–419
10. Hirschberg J, Ben-Yehuda A, Pecker I, Ohad N (1987) Mutations resistant to photosystem II herbicides. *Plant Molecular Biology*, 357–366
11. Rani P, Kumari J, Agarwal S, Singh DV (2019) Binding mode of aryloxyphenoxypropionate (FOP) and cyclohexanedione (DIM) groups of herbicides at the carboxyl transferase (CT) domain of Acetyl-CoA carboxylase of *Phalaris minor* Network Modeling Analysis. *Health Informatics Bioinformatics* 8(1):10
12. Singh DV, Adeppa K, Misra K (2011) Mechanism of isoproturon resistance in *Phalaris minor*: in silico design, synthesis and testing of some novel herbicides for regaining sensitivity. *J Mol Model* 18:1431–1445
13. Singh DV, Agarwal S, Kesharwani RK, Misra K (2012) Molecular modelling and computational simulation of the photosystem-II reaction center to address isoproturon resistance in *Phalaris minor*. *J Mol Model* 18:3903–3913
14. Kishore G. Singh DV (2018) Isoproturon tolerance and resistance in *Phalaris minor*. Sequence and Structural similarity in *Phalaris minor* and wheat D1 protein. *Pesticide Research Journal* 30(2):246–250
15. Zhang H, Yao Y, Yang H, Wang X, Kang Z, Li Y, Li G, Wang Y (2012) Molecular dynamics and free energy studies on the carboxypeptidases complexed with peptide/small molecular inhibitor: Mechanism for drug resistance. *Insect Biochem Mol Biol* 42:583–595
16. Liu J, Liu M, Yao Y, Wang J, Li Y, Li G, Wang Y (2012) Identification of Novel Potential β -N-Acetyl-D-Hexosaminidase Inhibitors by Virtual Screening, Molecular Dynamics Simulation and MM-PBSA Calculations. *Int J Mol Sci* 13:4545–4563
17. Burgos NR (2015) Whole-plant and seed bioassays for resistance confirmation. *Weed Sci* 63(SP1):152–165
18. Fiser A, Sali A (2003) Modeller: generation and refinement of homology based protein structure models. *Methods Enzymol* 374:461–491
19. Guex N, Peitsch MC (1997) SWISS-MODEL and Swiss-Pdb Viewer: an environment for comparative protein modelling. *Electrophoresis* 18:2714–2723
20. Laskowski RA, MacArthur MW, Moss DS, Thornton J (1993) PROCHECK: A program to check the stereo chemical quality of protein structures. *J Appl Crystallogr* 26:283–291
21. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:407–410

22. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309
23. John J, Irwin, Brian K, Shoichet (2005) ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J Chem Inf Model* 45:177–182
24. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH (2015) PubChem Substance and Compound databases. *Nucleic Acids Res* 44:202–213
25. Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. *J Med Chem* 49:3315–3321
26. Yuan Y, Pei J, Lai L (2011) LigBuilder 2: A practical de Novo Drug Design Approach. *J Chem Inf Model* 51:1083–1091
27. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 16:2785–2791
28. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4:435–447
29. Schuttelkopf AW, Aalten MFV (2004) PRODRG: a tool for high throughput crystallography of protein-ligand complexes. *Acta Crystallogr* 60:1355–1363
30. Marrink SJ, Berger O, Tieleman DP, Jahnig F (1998) Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations. *Biophys J* 74:931–943
31. Tieleman DP, MacCallum JL, Ash WL, Kandt C, Xu Z, Monticelli L (2006) Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid–protein interactions, side chain transfer free energies and model proteins. *J Phys: Condens Matter* 18:1221
32. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* 81:3684–3690
33. Oostenbrink C, Villa A, Mark AE, Gunsteren WFV (2004) A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J Comput Chem* 25:1656 – 1676
34. Amiri S, Sansom MS, Biggin PC (2007) Molecular dynamics studies of AChBP with nicotine and carbamylcholine: the role of water in the binding pocket. *Protein Eng Des Sel* 20:353–359
35. Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Mol Graph* 14:33–38
36. Kumari R, Kumar R, Lynn A, g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of Chemical Information and Modeling*, 54, 1951–1962
37. McQuarrie DA (1973) *Statistical Thermodynamics*, 2nd ed., Harper and Row: New York
38. Broser M, Glockner C, Zouni A (2011) Structural basis of cyanobacterial photosystem II inhibition by the herbicides terbutryn. *The Journal of Biological Chemistry* 286:15964–15972

39. Sippl MJ (1993) Recognition of errors in three dimensional structures of proteins, 17. Proteins, pp 355–362
40. Lu Q, Wang DS, Chen CS, Hu YD, Chen CS (2005) Structure-Based Optimization of Phenylbutyrate-Derived Histone Deacetylase Inhibitors. J Med Chem 48:5530–5553

Figures

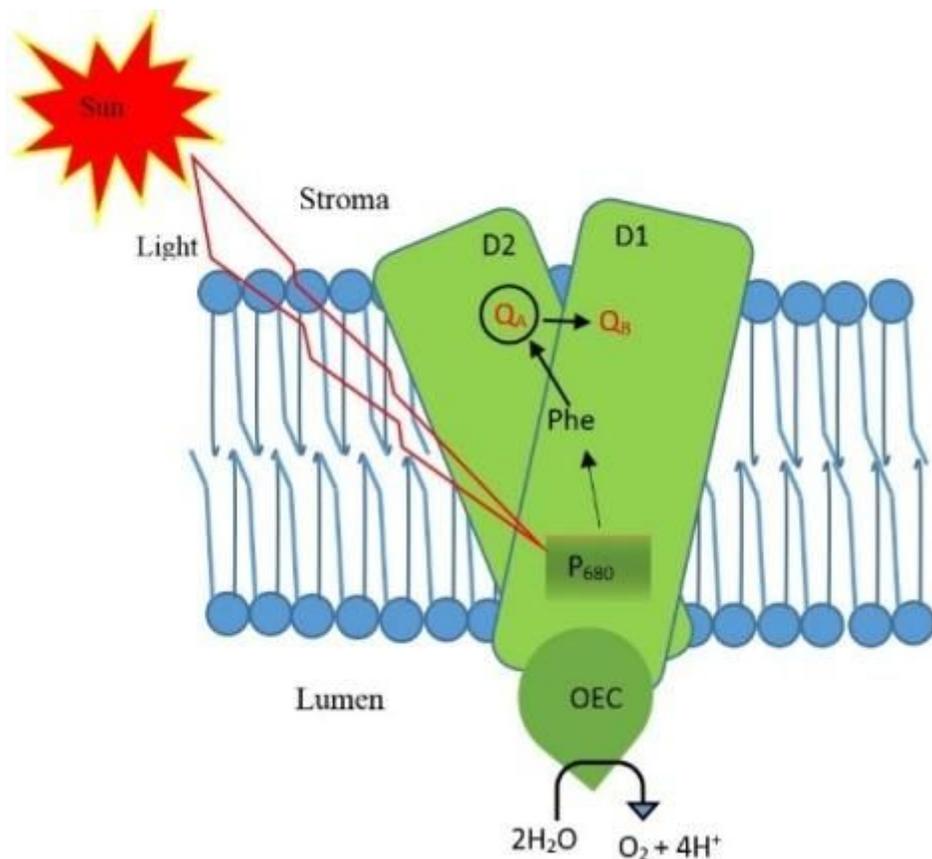


Figure 1

Architecture of D1 and D2 protein of Photosystem-II and QA site has shown by a circle on D2 protein.

A

Query	1	PTLLTATSVFIIAFIAAPPVDIDGIREPVS	60
		PTLL AT F+IAPFAAPPVDIDGIREPVS	
3PRQ_D1	30	PTLLAATICFVIAFIAAPPVDIDGIREPVS	89
Query	61	AASVDEWLYNGGPEYELIVLHFLLG	120
		AAS+DEWLYNGGPEY+LI+ HFLLG +	
3PRQ_D1	90	AASLDEWLYNGGPEYQLIIFHFLLG	149
Query	121	LIYPIGQGSFSDGMPLGISGTFNFMI	180
		LIYPIGQGSFSDGMPLGISGTFNFMI	
3PRQ_D1	150	LIYPIGQGSFSDGMPLGISGTFNFMI	209
Query	181	VTSSLIRETTENESANEGYKFGQEEET	240
		VTSSLIRETTE ESAN GYKFGQEEET	
3PRQ_D1	210	VTSSLIRETTETESANYGYKFGQEEET	269
Query	241	FVVGWFTALGISTMAFNNG	261
		FVVG+WFTALGISTMAFNNG	
3PRQ_D1	270	FVVGWFTALGISTMAFNNG	290

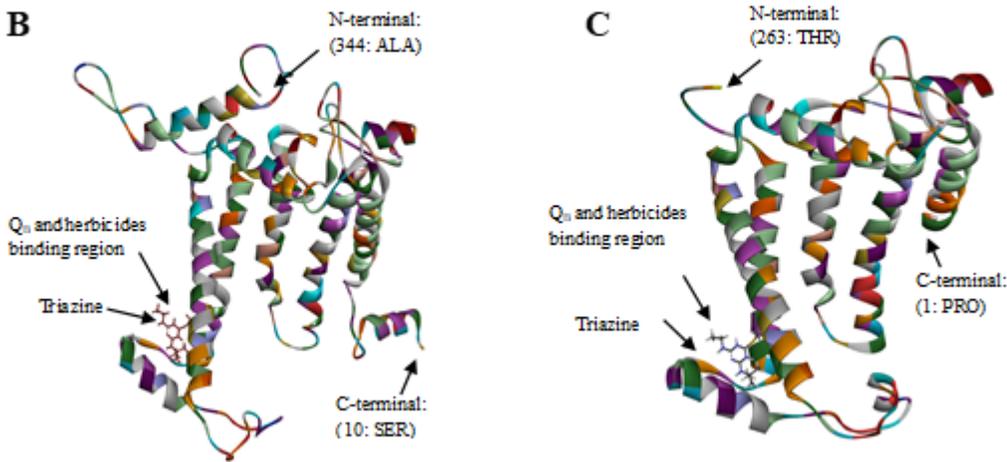


Figure 2

(A) Plot showing the alignment between query sequence (gi|30413115|gb|AAP33145.1| psbA) and template sequence (pdb|3PRQ|A) of D1 protein. (B) Template strand crystal structure of Cynobacterial PS-II; (PDB|3PRQ|A) bonded with herbicide molecule at QB binding site. (C) Modelled D1 protein of R-Biotype of *P. minor*. Protein structures are shown in ribbon drawing and ligands in stick from.

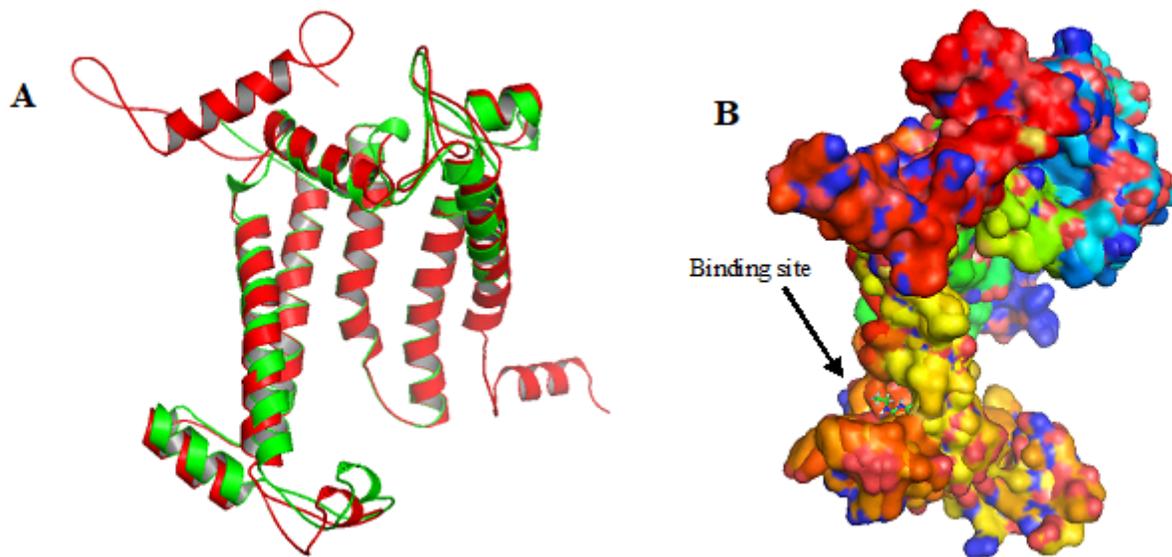


Figure 3

Superimposed Comparative ribbon model structure and the binding site of ligand. (A) Superimposed model of D1 protein (green) onto the template structure pdb|3PRQ|A (red) with RMSD and TM-score 1.3 Å and 0.99125 respectively. (B) The Connolly surface of modelled D1 protein and template structure pdb|3PRQ|A, showing ligand-binding site.

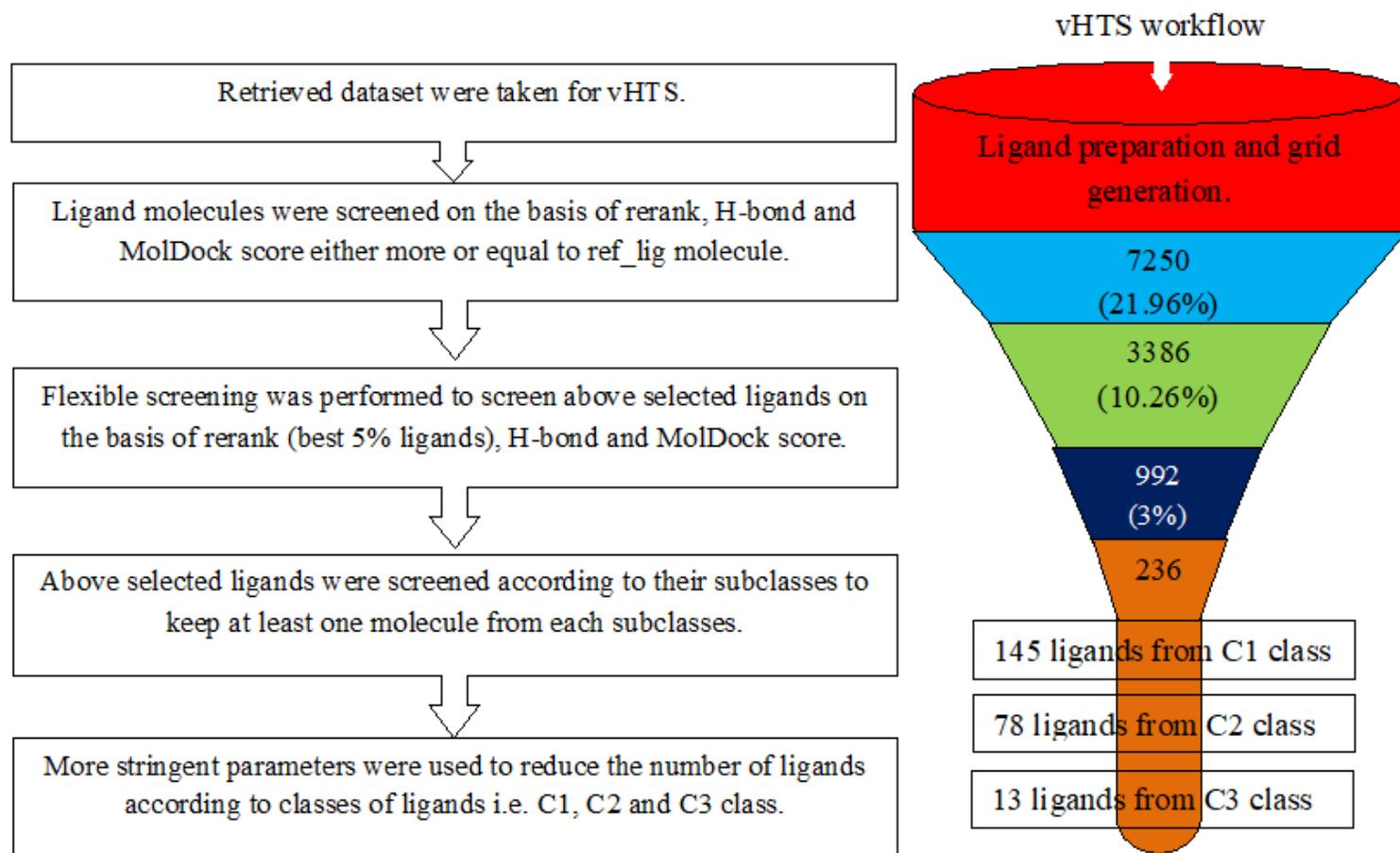


Figure 4

Flow chart is showing complete methodology of the vHTS and selection of ligands in each step of this study.

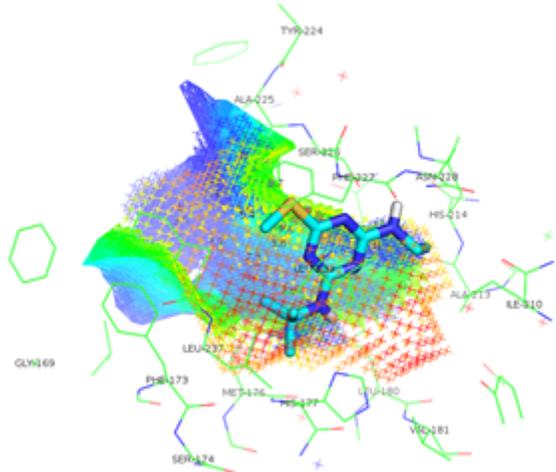
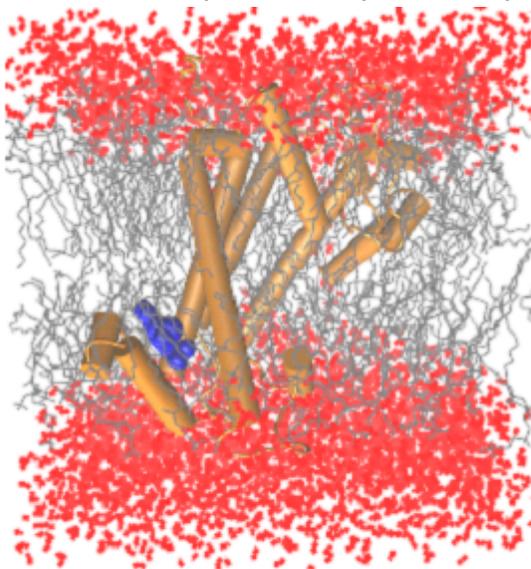
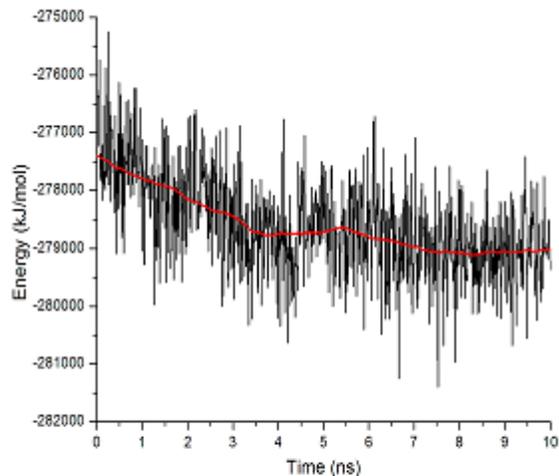


Figure 5

Isoelectronic surface of herbicide binding site of D1 protein (pdb|3PRQ); cavity and key sites of interactions and receptor based pharmacophore.



A



B

Figure 6

(A) Snapshot of the simulation system after solvation. Water molecules are shown in red; phospholipids groups are in shown in gray; protein is shown in orange and ligand is shown in blue. (B) Total energy variation for the ref_lig protein complex throughout the simulation.

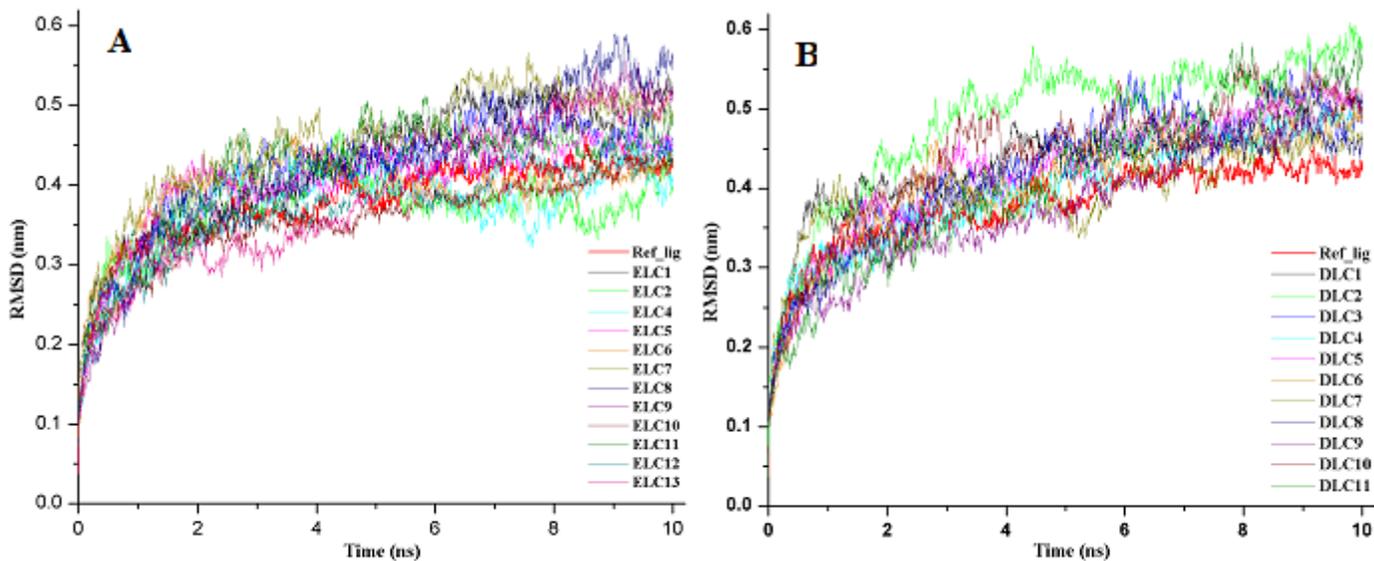


Figure 7

Time dependence root means square deviation (RMSD) of the backbone of protein throughout the simulation. (A) RMSD of protein with ELC leads and compared with RMSD of protein with ref_lig. (B) RMSD of protein with DLC leads is compared with an RMSD of protein with ref_lig.

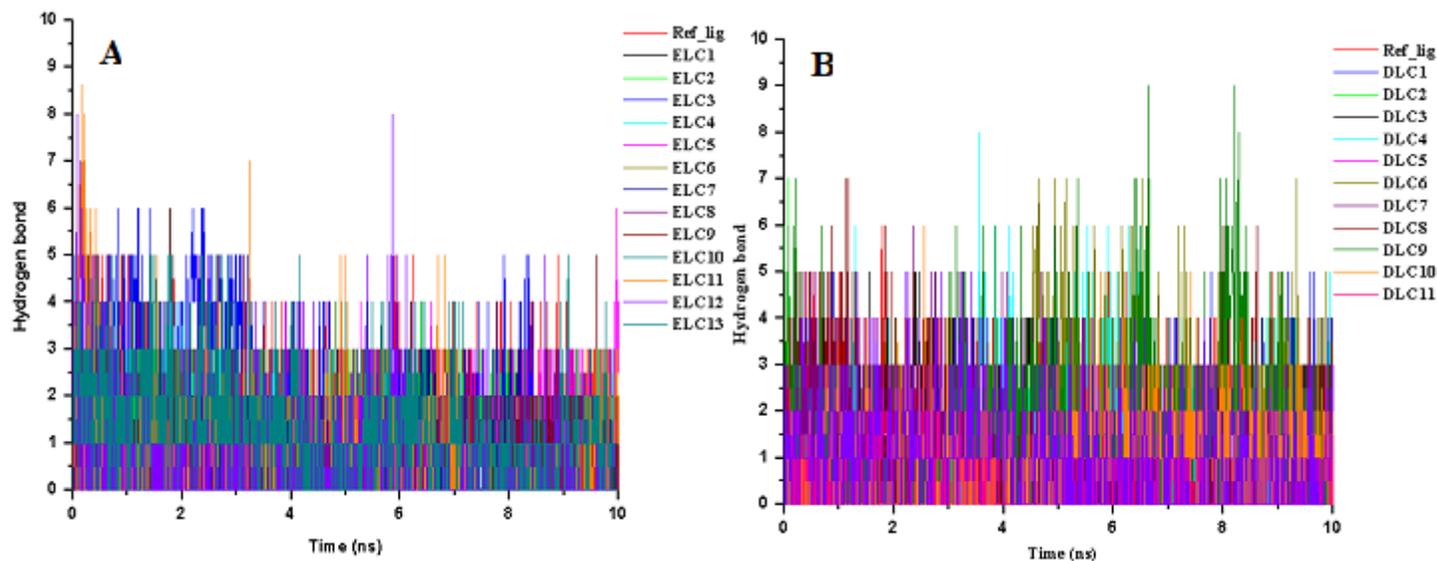


Figure 8

Plots are showing the number of inter-molecular H-bond interactions between proteins, ELC and DLC lead molecules throughout the simulation. (A) Number of H-bonds between protein and ELC lead molecules. (B) Number of H-bonds between protein and DLC lead molecules and compared with H-bond interaction between protein ref_lig.

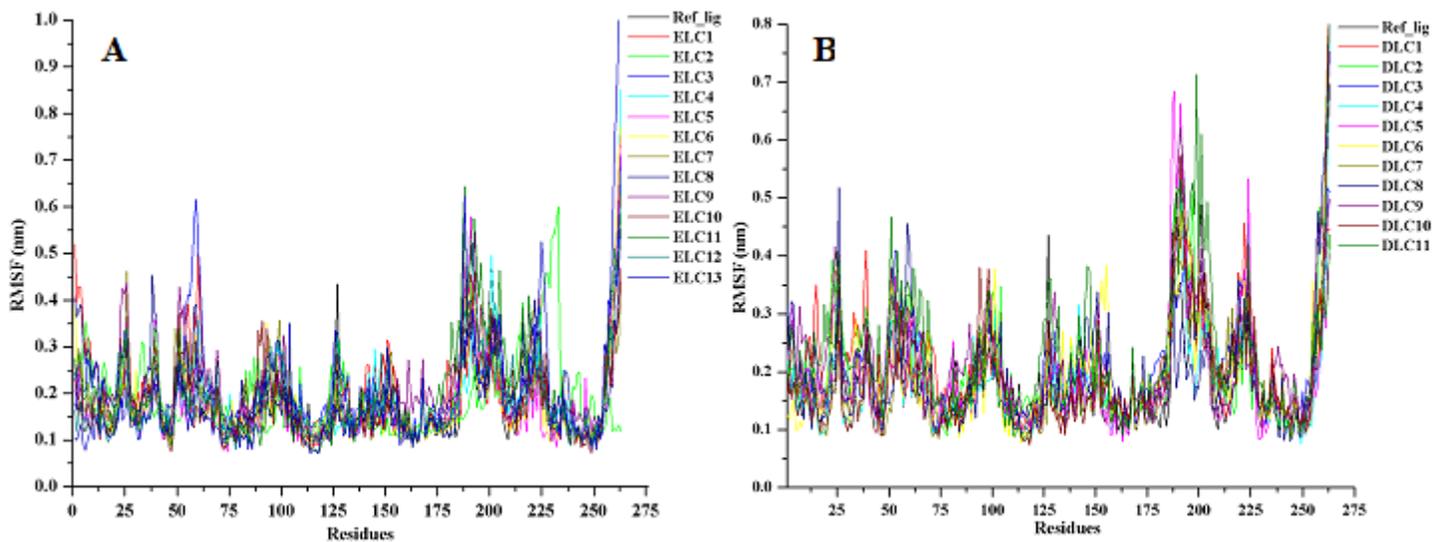


Figure 9

Root mean square fluctuation of modelled D1 protein. (A) RMSF of ELC lead-protein complexes. (B) RMSF of DLC leads-protein complexes.

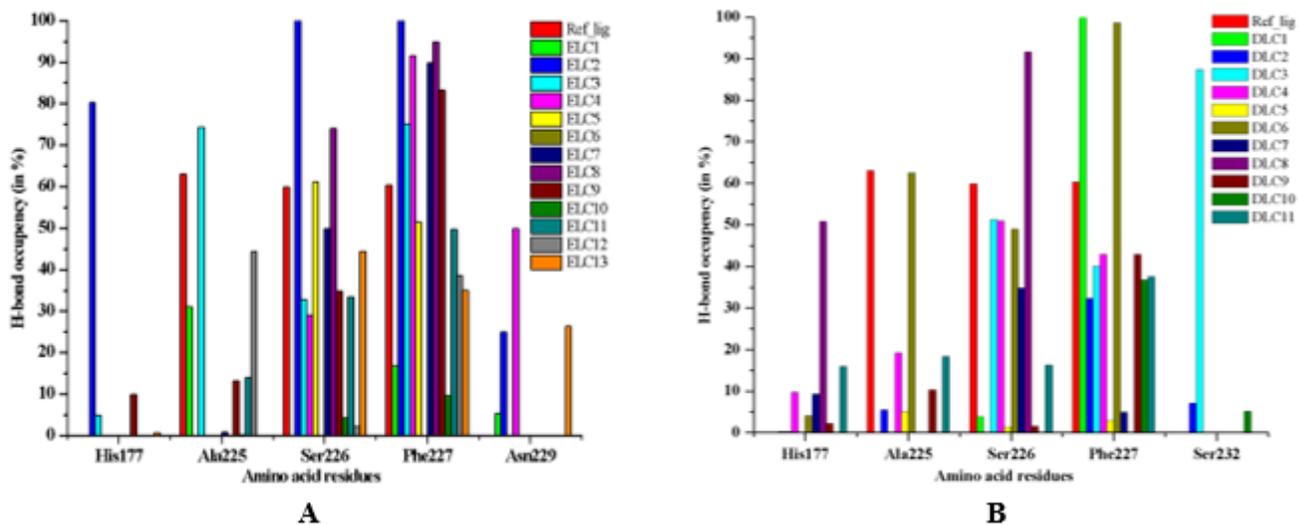


Figure 10

Plots are showing H-bond occupancy for (A) ELC lead molecules (B) DLC lead molecules in comparison to ref_lig.

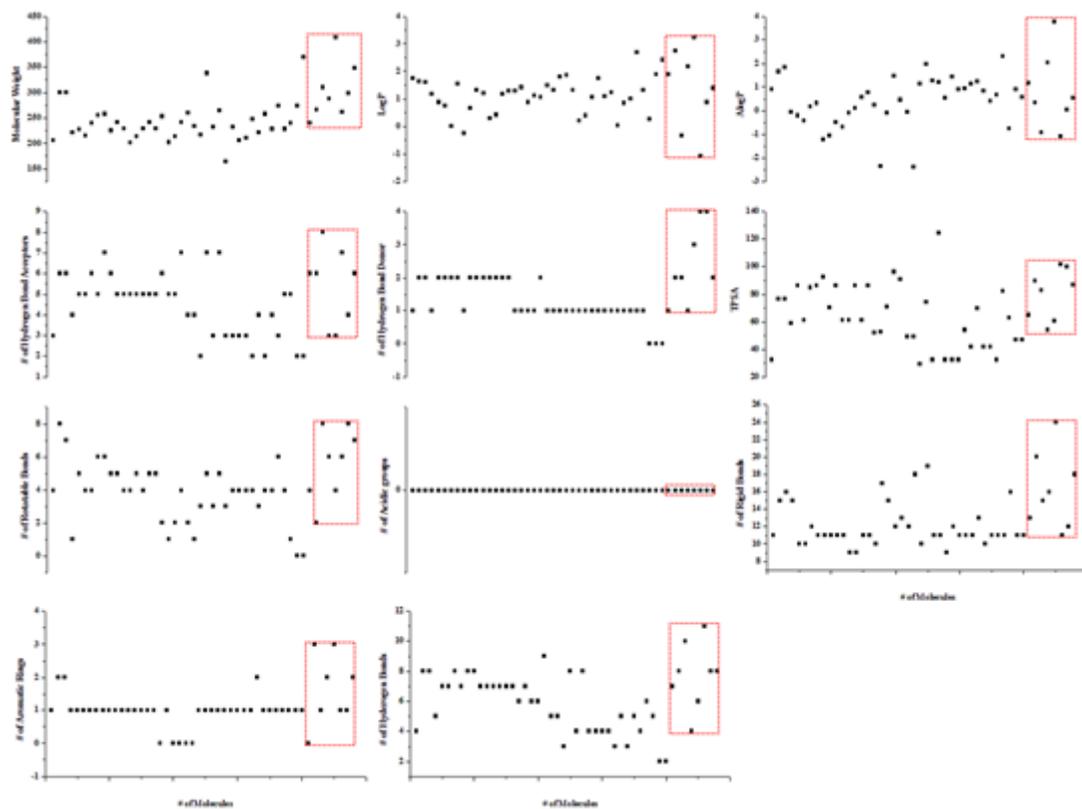
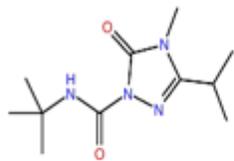
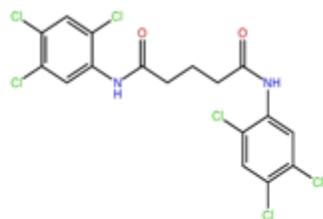


Figure 11

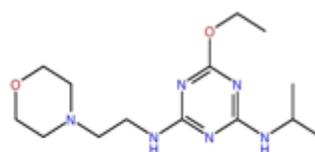
Dot plot is showing the physicochemical properties of six ELC, two DLC molecules in rectangle box and known PS-II herbicides.



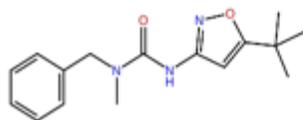
ELC2



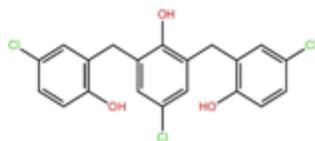
ELC4



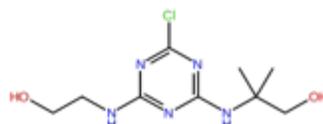
ELC5



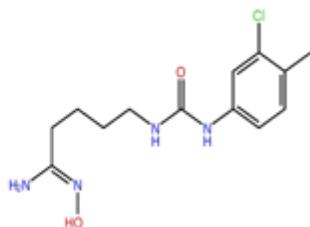
ELC8



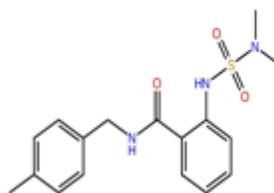
ELC11



ELC12



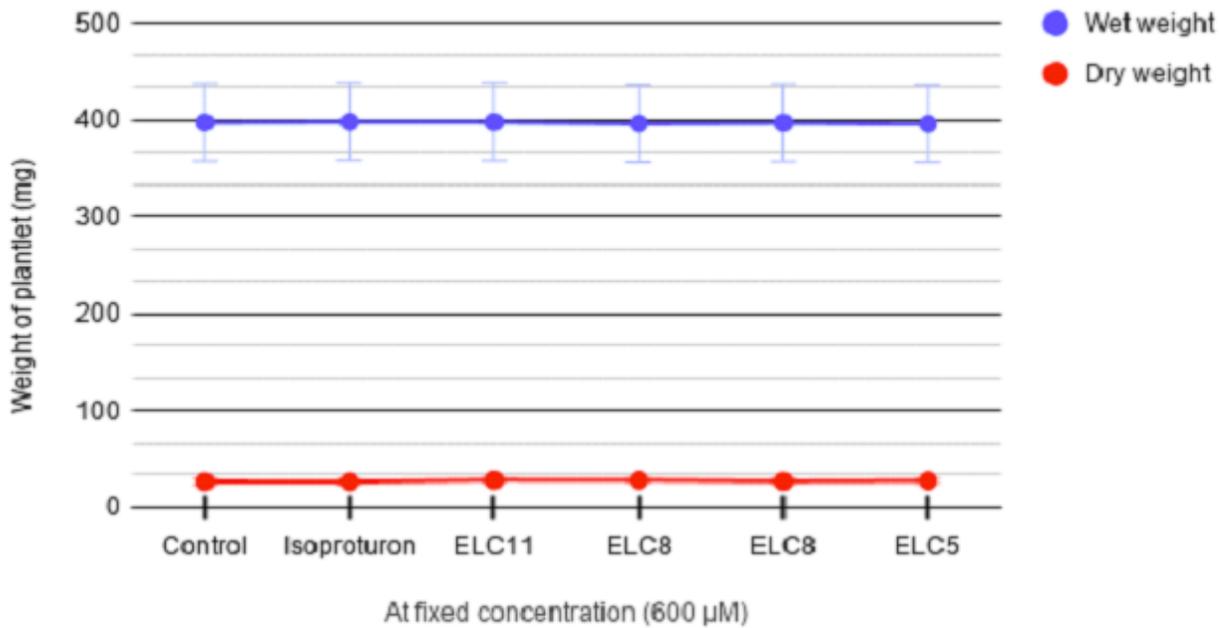
DLC4



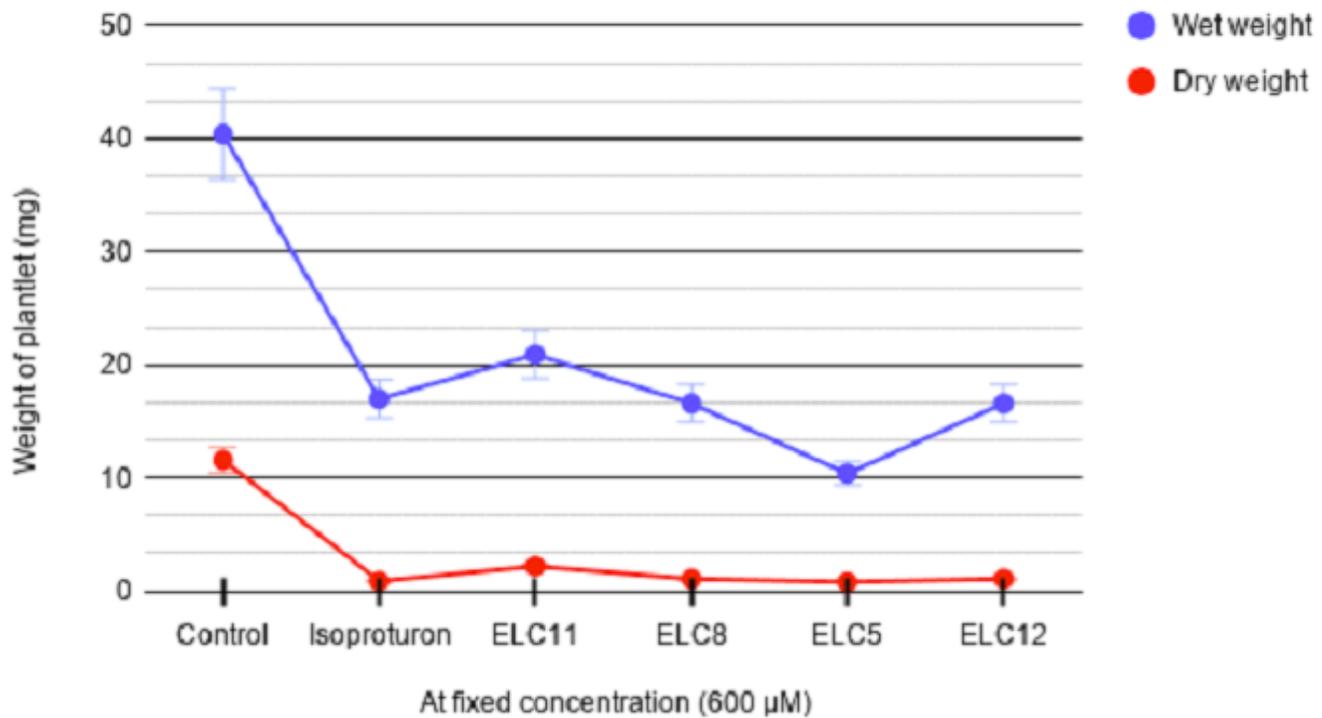
DLC6

Figure 12

2D Structure of proposed eight best lead molecules (Six from ELC and two from DLC) as inhibitors of D1 protein.



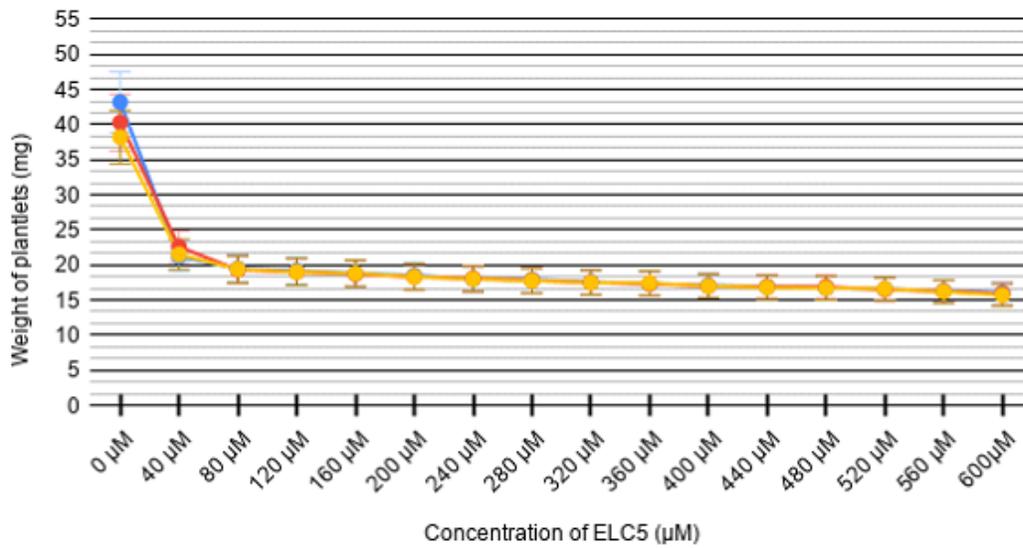
(A)



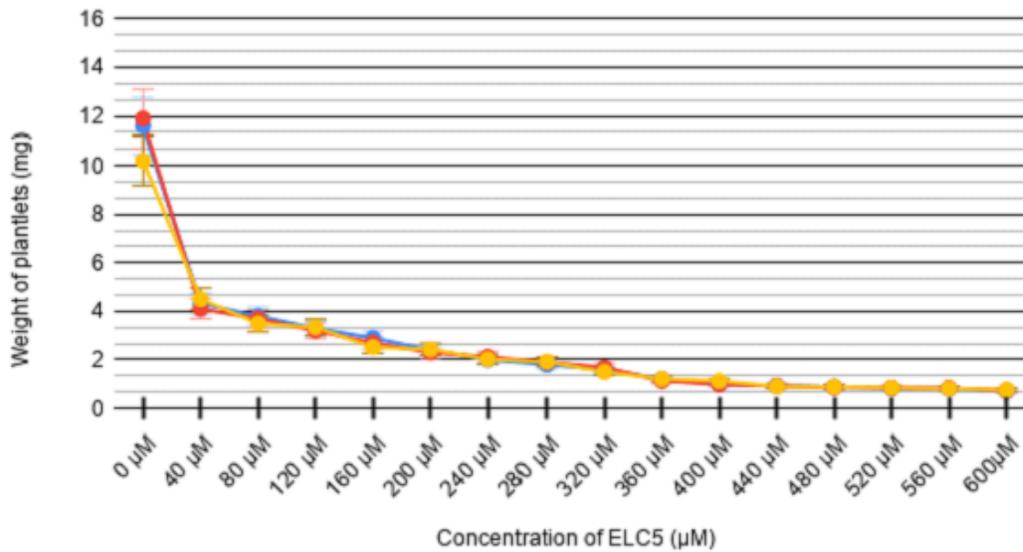
(B)

Figure 13

The effect on the biomass of plantlets upon treatment with new molecules on (a.) Wheat and (b.) *P. minor* plantlet under controlled laboratory conditions.



(A)



(B)

Figure 14

Effect and reduction in the biomass of plantlets upon treatment with new potential herbicide molecule ELC5 (a) wet weight of *P. minor* plantlet, and (b) dry weight of *P. minor* plantlet. The experiment was conducted in triplicate and the plotted lines in the respective color yellow, red and blue represents the same.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplimentray16June.docx](#)