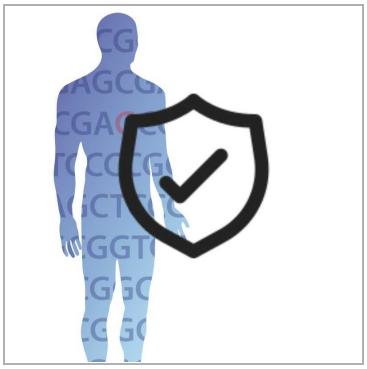

JUN 23, 2023

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.4r3l225nql1y/v1

Protocol Citation: Anagha S Setlur, Chandrashekar K, Vartul Panhalkar, Sonia Sharma, Manas Sarkar, Vidya Niranjan 2023. In-silico-based toxicity investigation of natural repellent molecules against the human proteome: A safety profile design. **protocols.io** https://dx.doi.org/10.17504/protocols.io.4r3l225nql1y/v1

Protocol status: Working
We use this protocol and it's working

Created: Jun 21, 2023

Last Modified: Jun 23, 2023

PROTOCOL integer ID: 83767

In-silico-based toxicity investigation of natural repellent molecules against the human proteome: A safety profile design

Anagha S Setlur[1], Chandrashekar K[1], Vartul Panhalkar[2], Sonia Sharma[2], Manas Sarkar[2], Vidya Niranjan[1]

[1]Department of Biotechnology, RV College of Engineering, Bangalore-560059;
[2]Research and Development, Reckitt Benckiser India Pvt. Ltd., Gurgaon, Haryana- 122001

Vidya Niranjan

DISCLAIMER

1. This protocol was designed with the intent of assessing the safety risks of natural small molecule repellents. However, it may also be used for generally gauging the safety challenges of any small molecule, as preliminary studies to in-vitro and in-vivo.

2. UniProt has the updated total number of proteins in the human proteome. This protocol has the numbers that were, at the time the total number of proteins that were available.

Keywords: Human safety evaluation, proteome, natural small molecules, repellents, off-targets and toxicity, molecular docking and simulations

ABSTRACT

Comprehending the toxicity and other adverse effects of any given set of potential natural repellent molecules against various proteins in humans is essential to determine their safety prior to their use in formulations. An *in-silico* computational protocol provided in this article can be used to determine the toxicity of any given molecule against the entire human proteome and thus build a safety profile for the same. Preliminary toxicity predictions for oral, dermal and inhalation routes were made for the natural small molecules using computational tools such as Protox-II, SwissADME and T.E.S.T. The screening of the whole proteome to determine the off-targets was then conducted for a given set of natural small molecules with similar mode of action. Tools such as LigTMap, PharmMapper, Swiss Target Predictions and SuperPRED were used for off-target determination. Selection criteria were employed to shortlist the most interacting and probable off-targets to the ligands. These shortlisted protein targets were then manually docked using Schrodinger and the best and poor docked complexes were simulated at 100ns in the Desmond suite of Schrodinger to assess the stability of binding. The final safety profile was then designed after collating all the results obtained from the above-mentioned analyses, considering all parameters. The recommended set of natural small molecules can be visualized for better understandability using a bar chart from the safety profile created for the given set of molecules. This protocol can thus be employed for testing the toxicity and safety of any ligand computationally, so these predictions maybe used prior to any *in-vitro* studies, to understand the preliminary safety challenges.

Keywords: Human safety evaluation, proteome, natural small molecules, repellents, off-targets and toxicity, molecular docking and simulations

GUIDELINES

The small molecules that must be assessed for their safety risks, should be available in SDF and SMILES formats.

MATERIALS

All tools and web-servers used in this protocol are mentioned within text.

SAFETY WARNINGS

NA.

ETHICS STATEMENT

None.

BEFORE START INSTRUCTIONS

Appropriate computational resources must be available to run the tools and web-servers mentioned in the protocol.

TOXICITY PREDICTIONS

1 The toxicity predictions for the given set of potential molecules with mosquito repellent activities were first tested via several tools. The detailed description for each is provided below.

1.1 Protox-II

This tool predicts the primary toxicity properties of all natural small molecules to help generate a safety profile via oral routes. It also provides the acute toxicity values. The analysis of toxicity of each small molecule revealed its probable levels of target organ toxicity towards humans, if and when these were consumed accidentally. Protox-II (https://tox-new.charite.de/protox_II/)employs machine learning (ML) models to predict endpoints of toxicities such as acute toxicity, hepato-toxicity, cyto-toxicity, mutagenicity, immunotoxicity, carcinogenicity and other adverse outcomes via Tox-21 pathways.

Protox-II also provides information regarding the probability of organs/targets in the human body that the repellent molecules are active or inactive to, providing an idea on the oral toxicity predictions.

Analysis of Protox II results

The Protox-II outcomes revealed specific threshold values to interpret the toxicity predictions. The toxic doses are provided as LD50 values in mg/kg body weight. LD50 values are the lethal dose at which 50% of the test animals die upon exposure to the compound. Thus, higher the LD50 value, lower the toxicity.

The toxicity classes are also provided in Protox-II as an outcome and are described as follows according to the GHS Classification criteria (https://www.era-environmental.com/blog/ghs-hazard-classification).

Class I: fatal ($LD50 \leq 5$)
Class II: fatal ($5 < LD50 \leq 50$)
Class III: average toxicity ($50 < LD50 \leq 300$)
Class IV: non-toxic, safe to use ($300 < LD50 \leq 2000$)
Class V: non-toxic ($2000 < LD50 \leq 5000$)
Class VI: non-toxic ($LD50 > 5000$)

Acute toxicity refers to the adverse effects following an oral or dermal administration of a substance. Thus, to estimate and predict acute toxicity, Protox-II and SwissADME (explained in the next section) were used.

Table 1: Acute toxicity hazard categories (LD50/LC50) and values defining the respective categories: Threshold values for toxicity class predictions (1-5) (according to the GHS classification criteria)

	A	B	C	D	E	F
	Exposure route	**Category 1**	**Category 2**	**Category 3**	**Category 4**	**Category 5**
	Oral (mg/kg)	5	50	300	2000	5000
	Dermal (mg/kg)	50	200	1000	2000	
	Gases (ppm)	100	500	2500	5000	
	Vapours (mg/l)	0.5	2.0	10	20	
	Dusts and Mists (mg/l)	0.05	0.5	1.0	5	

1.2 SwissADME

It is important to understand dermal toxicity predictions prior to using a given set of naturals in repellent formulations. This was understood using SwissADME (http://www.swissadme.ch/) that predicts the skin permeation along with other ADME and pharmacokinetic parameters.

SwissADME computes the ADME properties (Absorption, distribution, metabolism, excretion) of the naturals, and other pharmacokinetic attributes. It is essential to understand the various pharmacokinetic properties of potential natural repellent molecules since when used in formulations, they may find their way to human skin/body through multiple routes of exposures such as dermal, oral or aspiration. Thus, an idea on its pharmacokinetic safety is crucial. Thus, all the naturals that were run on Protox II were also subjected to SwissADME studies.

The properties studied included:

1. Human gastrointestinal absorption (HIA): estimation of the permeability into the human GI system.
2. Interaction with cytochrome molecules (CYP450): for comprehending adverse effects of naturals.
3. Blood-brain barrier (BBB) permeation: This barrier protects the central nervous system

from exposure to any molecules that have a toxic effect. Thus, it is pivotal to understand if a compound is a BBB permeant.
4. Skin permeation (Kp): Prediction of Kp values determine the transport of the molecule through the mammalian epidermis. More negative the log Kp value is, lesser the skin permeation.

The route of exposure: dermal

1.3 T.E.S.T Tool

T.E.S.T (https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test)predicts the general small molecule properties that can help in comprehension of the natural small molecules. But, aside that, this tool predicts two types of toxicities:

1. Developmental toxicity (DevTox): estimates if the molecule is involved in causing any developmental toxicity to humans.
The route of exposure studied in this tool: Oral and inhalation
2. AMES mutagenicity: this tool detects if a specific molecule can cause mutations in the DNA of the test organism, in this case, humans.
The route of exposure studied in this tool: oral, inhalation

Since classifying them either route wise or toxicity wise will lead to repetition of tools and content, a better way of expressing this is provided in the table below.

Table 2: The tools used, routes of exposure and toxicity types

	A	B	C	D
	Sl. No.	**Tool name**	**Route of exposure**	**Type of toxicity**
	1	Protox-II	Oral and dermal	Acute
	2	SwissADME	Dermal	Acute
	3	Toxicity Estimation Software Tool (T.E.S.T)	Oral, inhalation	Developmental, AMES mutagenicity

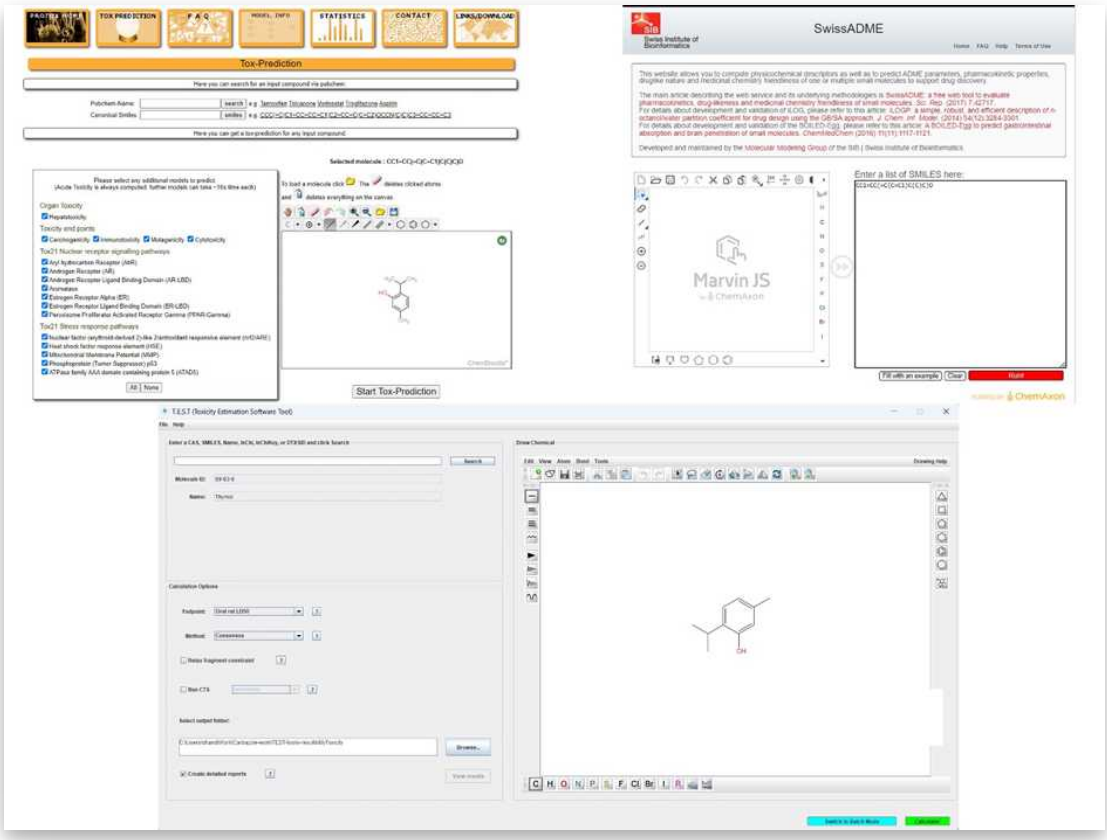

Figure 1: Protox-II, SwissADME and T.E.S.T tools used for obtaining a toxicity profile.

HUMAN PROTEOME SCREENING

2 Screening of the human proteome

The human proteome consists of approximately 81,837 proteins **(Proteome ID UP000005640).** It is thus essential to screen the entire proteome to identify if the potential insecticide components have any effect on the human proteins. For this purpose, a protocol was developed to screen the entire proteome and develop a safety profile for a given set of naturals to comprehend their toxicity/adverse effects on humans prior to being taken forth for repellent formulations.

The UniProt organism ID used for this study was 9606. The screening of the entire proteome was carried out by identification and prediction of off-targets for the given set of potential repellent natural small molecules. Off-target identification was performed using several bioinformatic tools and databases. These databases are known to have most of the identified proteins in the human proteome, thereby, covering the maximum stretch of the proteome for off-target-search and analysis.

Thus, out of ~81K proteins, off-target search revealed ~7995 proteins, after elimination of all repetitive proteins. The reason for shortlisting about ~7995 is because off-target search is specific for every ligand. The search results are provided for only those molecules that have a possibility of binding to the molecule in question. A specific selection criterion was then employed to further narrow down our off-targets that had to be taken forward for manual docking studies. The selection criteria is provided in the following sections. All repetitive targets and duplicates were eliminated again to further narrow down the proteins. The final shortlisted and identified off-targets of the repellent molecules were then docked against the given set of molecules, followed by analysis of docked complexes and molecular dynamic studies for the best docked and worst docked complexes.

OFF-TARGET PREDICTIONS

3 Prediction of off-targets of small molecules in human

The binding of the ligands to protein targets other than those for which the ligands were originally meant to bind are called off-targets. In this study, the authors predicted the off-targets in the entire human proteome for a given set of naturals to determine the effect of these natural small molecules on targets other than their own. Prediction of off-targets was carried out using various tools, including, LigTMap (https://cbbio.online/LigTMap/),PharmMapper (http://www.lilab-ecust.cn/pharmmapper/), SwissTargetPrediction (http://www.swisstargetprediction.ch/) and SuperPRED (https://prediction.charite.de/). These tools search the entire databases and finds all possible off-targets for each natural molecule in *Homo sapiens.*

3.1 LigTMap

This tool identifies off-targets from various classes of human protein targets by combining the ligand similarity searches with docking and binding similarity score analyses to predict the proteins. Every protein belongs to specific classes, based on which the targets were identified. Outcomes include PDB (Protein Databank) IDs of the targets, binding similarity scores, predicted affinity and probable binding energies of the given set of naturals to the human proteins.

3.2 PharmMapper

This tool works on the principle of pharmacophore mapping and is highly efficient due to its high-throughput approach. PharmMapper functions as a tool for recognizing off-targets and other new compounds with information on its biological activity. Thus, for further examination, all targets with Z-score ≥ 1 were screened.

3.3 SwissTarget

This tool estimates the most probable targets of the selected bioactives in a given organism, including humans. Probability values are obtained as outcomes that indicate the binding probability of the naturals to the predicted protein targets.

3.4 SuperPRED

This tool predicts the targets based on machine learning (ML) models, by screening the entire database having 6,65,000 compound-target interactions to identify the best off-targets for desired natural small molecules. The results are provided as probability scores and model accuracies of the ML model used for prediction.

The following table provides information on all the tools used for prediction of off-targets and their general working:

Table 2: Summary of tools used for off-target predictions and their general information

A	B
Tool	**Information**
Off-target predictions	
LigTMap	Identifies off-targets from various classes of human protein targets
	Combines ligand similarity search with docking and binding similarity score analysis to predict the off-targets
	Binding similarity score, the predicted affinity and probable docking score predicted
PharmMapper	High throughput approach to predict off-targets
	Provides Z-score for the predicted off-targets
SwissTarget Predictions	Predicts binding probability of natural small molecules to targets
	Probability scores are predicted
Super-PRED	Predicts based on machine learning models
	Screens database of 6.6 lakh compounds.
	Probability and model accuracy is predicted

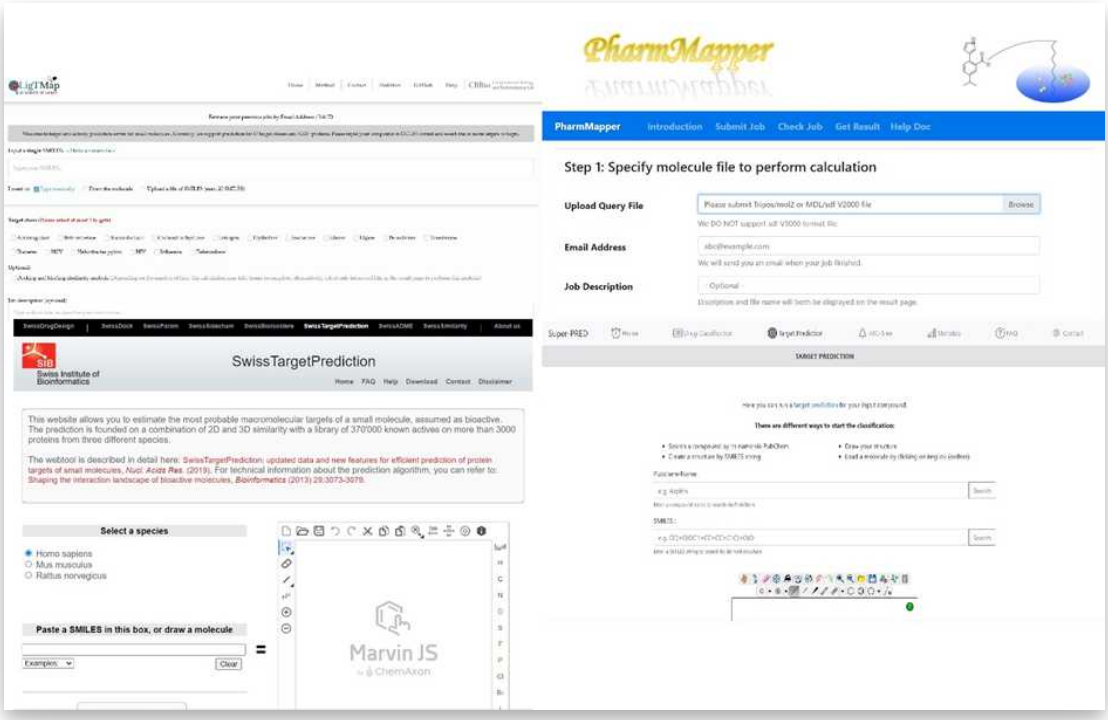

Figure 2: Off-target predictions using LigTMap, PharmMapper, SwissTarget Predictions, and SuperPRED

SELECTION CRITERIA

4 Selection criteria for off-target predictions

Every tool used for the prediction of off-targets in humans provide several protein targets with the possibility of binding via docking scores. However, since all the targets cannot be selected for manual docking purposes, the selection criteria were set to narrow down the overall number of protein targets that maybe studied further and, on whose functions/pathways the natural small molecules may have an effect.

Therefore, using the selection criteria provided below, and removal of all duplicates identified from each tool, 614 proteins were shortlisted to test effect of natural small molecules on them via manual molecular docking.

Table 3: The selection criteria for each tool used for off-target predictions and shortlisting the protein targets, along with their significance

	A	B	C	D	E

	A	B	C	D	E
	Tools used	**Overall results**	**Selected targets**	**Selection Criteria**	**Significance**
	LigTMap	150	10	LigTMap score >= 0.5	Protein targets are ranked based on LigTMap score. Hence, score greater than or equal to 0.5 has to be chosen in order to take it forward for further studies
	SuperPred-Target Prediction	2652	380	Probablity>= 75%	Probabilities are extensively used in order to rank protein targets that bind to the bioactive small molecules. Higher the probability, higher is the binding affinity of the small molecule to the protein targets
	PharmMapper	2985	478	Z Score>= 1	Z score is based on “Enhanced pharmacophore-based target prediction”. Targets that have score greater or equal to 1 may be chosen for further studies

	A	B	C	D	E
	Swiss Target Prediction	2208	540	Probablity> 0	Probabilities are extensively used in order to rank protein targets that bind to the bioactive small molecules. Targets that have a probability value that is greater than 0 may be considered.
	Total	7995	1408	-	-
	Total number of target proteins selected	1408	614	After removal of Duplicates	
	Final Number of proteins docked	614			

MOLECULAR DOCKING

5 Molecular docking

Docking studies of a given set of ligands against shortlisted human protein targets were carried out using Schrodinger. All 614 proteins were prepared individually using the protein preparation wizard in Schrodinger. All the given natural small molecules along with positive and negative controls were prepared via ligprep wizard. All the conformers that were generated during ligand preparation were considered for docking. For this, cross docking technique was employed, where multiple protein and multiple ligands docking was carried out for all shortlisted human protein targets (obtained from screening of the human proteome) against the given set of ligands.

Docking was carried out by first importing the ligands to Schrodinger Maestro. The maximum ligand size was set to 500 atoms. OPLS forcefield was applied. Epik (Schrödinger Release 2022-3: Epik, Schrödinger, LLC, New York, NY, 2021.) was used to generate all possible ligand conformations at a neutral pH of 7.0 ± 2. Once the conformers were generated, all selected proteins were prepared using the protein preparation wizard in Schrodinger. All 614 proteins were pre-processed, by optimization of the H-bond and structure minimization. The water molecules were all deleted and the prepared proteins were employed for docking using XGlide in Schrodinger.

5.1 Cross-docking

The proteins were all defined in the receptor section and the prepared ligands in the ligand section in Schrodinger. The grid box was generated and all active sites were detected using the site-map for the protein. Setting the grid box so as to cover the entire protein indicates that the whole protein will be searched for the best binding pockets when docking is performed. Cross-docking between 614 proteins and the given set of ligands was performed using high-throughput virtual screening (HTVS) and the binding energies were obtained. A thorough analysis of the best docked complexes revealed an interaction profile that aided in performing a comprehensive study on the best and poorly bound complexes that could be taken forward for simulation studies.

5.2 Analysis of cross-docking

Cross-docking results were analyzed using 2-dimnesional interaction software called LigPlot+ to determine the amino acid interactions, including hydrogen bonding, hydrophobic bonding, etc. against the human protein targets. The docked complexes whose stabilities were required to be checked were simulated at 100ns in Desmond Schrodinger. The interactions of the given set of ligands with the human proteins were analyzed to determine their effects.

MOLECULAR DYNAMIC SIMULATIONS

6 Molecular dynamic simulations at 100ns

To further establish the relationship between the protein and the natural molecule, molecular dynamic simulations at 100ns were performed to demonstrate and understand the binding stabilities of the natural small molecules to human proteins. Thus, the best docked and the poorly docked complexes were simulated at 100ns at 1000 frame trajectory. The Maestro workspace in Desmond, Schrodinger (Schrödinger Release 2022-3: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.), was used for simulations due to its high scalability, reliability, precision, accuracy and rapidity. Prior to molecular dynamic simulations, the pre-processing of the proteins was carried out by assessing the structural errors and refining the same. Optimization was then performed for the interacting complexes via steepest descent algorithms for 500 times, followed by removal of water molecules. The simulation environment was built using the system builder in Maestro, by setting the solvent model to TIP3P, and defining the boundary boxes via an orthorhombic box that restricts and compresses the interacting complexes to 5 Å at the x, y and z axes. This system was built to constrain the simulation at 100ns within the boundary box. Addition of sodium and chloride ions neutralized the entire system.

Neutral pH of 7.0 was maintained for the system. 100ns simulations were then carried out with the trajectory recording set to 0.1ns and normal temperature and pressure (NPT) was

selected as the ensemble thermodynamic parameters.

6.1 Analysis of 100 ns simulations

Analysis of the simulations were performed based on the assumption that the docked complexes had to be unstable since the effect of natural small molecules on the human protein is being tested for developing a safety profile. Thus, various parameters were scrutinized for all simulations using the simulation interaction diagram tools.

RMSD

Modifications in the values of RMSD help analyze the stability of the docked complex and the mode of binding. Root mean square deviation is computed as the change in the displacement of a certain selection of atoms for a specific time frame with respect to the reference. This was estimated and analyzed for all 1000 frames within the set trajectory. The formula for RMSD is:

The RMSD for frame *x* is:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i'\left(tx\right)\right) - r_i\left(t_{ref}\right)\Big)2}$$

Where,
N = total number of atoms
r' = particular position of atoms in the x frame after superimposition to the frame of reference,
t_{ref} = time of reference
frame x = recorded at a time of tx

RMSF

The root mean square fluctuation calculates the average particle deviation, such as the amino acid residue of the protein, over time, from a position of reference, which is typically the time averaged position of the particle. Comprehending protein and ligand RMSFs help in analyzing the conformational changes of the protein and/or ligand during the simulation event. The formula for ligand RMSF where fit of the ligand on the protein is estimated is:

The RMSF for frame *x* is:

$$\mathrm{RMSF_x} = \sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t=1}}^{\mathrm{T}}\left(\mathrm{r_i'}\left(\mathrm{t}\right)\right) - \mathrm{r_i}\left(\mathrm{t_{ref}}\right)\right)2}$$

Where,
T = trajectory time over which the RMSF is calculated
t ref = the reference time (usually for the first frame, and is regarded as the zero of time)
r = position of atom i in the reference at time t ref
r' = position of atom i at time t after superposition on the reference frame

The formula for protein RMSF is:
The RMSF for frame *x* is:

$$RMSF_x = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_i'\left(t\right)\right) - r_i\left(t_{ref}\right)\right)2 >}$$

Where,
T = trajectory time over which the RMSF is calculated
t ref = the reference time (usually for the first frame, and is regarded as the zero of time)
r = position of atom i in the reference at time t ref
r' = position of atom i at time t after superposition on the reference frame

Aside these, the ligand properties, hydrophobic and hydrogen bond interactions, ionic bonds, water bridges, etc. were also studied. The interactions between the proteins and ligands in terms of the solvent accessible surface area (SASA), polar surface area (PSA) and radius of gyration were also studied to comprehend the stability of the interacting complex at the molecular and structural level. Thus, the simulation parameters provided assistance in building the final safety profile of natural molecules.

TOOLS' SENSITIVITY AND SPECIFICITY

7 Sensitivity and specificity of the tools used in the protocol

The below table provides the specificity and sensitivity of all the tools used in the study.

Table 4: Sensitivity and specificity of the tools and software used in the protocol

A	B	C	D
Sl. No.	Tools	Specificity	Sensitivity
1	LigTMap	Provides docked structures as results which helps in drug repurposing and computer-aided drug design	Success rate and accuracy-70%.
2	PharmMapper	The tool/database houses many pharmacophore databases such as potential drug target database, DrugBank, BindingDB, TargetBank and many more. These databases include more than 7000 receptor-based pharmacophore model information along with 1500 drug targets information.	The tool/database has the ability of performing high-throughput screening of the whole target database and gives potential results in 1 hour.
3	SwissTarget Prediction	Based on the combination of 3D and 2D similarity measures of the molecules, the web server accurately predicts the targets for bioactive small molecules	The success rate and accuracy is 60%

	A	B	C	D
	4	Super-PRED	The web server has been updated its ATC and target prediction from structure similarity to machine learning based on functional groups and small molecules' mode/mechanism of actions	The accuracy of specific prediction is 80.5%
	5	Protox-II	Tool uses a very large dataset consisting of 1,00,000 known drug molecules and their toxicity profiles.	The accuracy of the developed model is 86%
	6	SwissADME	The web server is highly specific in providing small molecules' pharmacokinetics profiles using more robust and speedy computational methods	The accuracy of the web server when benchmarked with other similar tools is 88%
	7	Toxicity Estimation Software Tool (T.E.S.T)	The tool gives us comparison between both experimentally derived value and predicted value which has the least difference.	TEST tool has an accuracy of 80% in terms of predicting acute toxicity and 70% in terms of predicting carcinogenicity

	A	B	C	D
	8	Schrodinger Glide	Glide tool has a specificity of 94%. The tool has an accurate model of predicting binding sites and virtual screening protocol.	Glide tool is more accurate (twice) than the GOLD, FlexX and Surflex methods. Tool has 87% sensitivity.
	9	Desmond	Tool has 92% specificity and provides parallel scalability, scientific accuracy in terms of simulation throughput.	Tool has 85% sensitivity

FINAL SAFETY PROFILE

8 Development of final safety profile

The final safety profile was then developed based on all the parameters considered and the results obtained. The natural small molecules that have low levels of toxicity or are predicted to be relatively safer from these tools and tests, are recommended for further use in wet lab, while those that affect the human protein pathways are studied further to analyse their effects. Additionally, given a large number of natural small molecules, once the entire profile is built, they may be categorized into different set of safe natural small molecules to provide appropriate recommendations for wet lab studies.

In the present protocol, after interpreting and understanding various parameters such as toxicity classes, carcinogenicity, mutagenicity, hepato-toxicity, immunotoxicity, cytotoxicity, various Tox 21 nuclear receptor signaling pathways, Tox-21 stress response pathways, Ames mutagenicity, developmental toxicity, pharmacokinetic properties such as GI absorption, blood brain barrier permeation, skin permeability, and molecular docking abilities and simulations in dynamic environments, the safety profile was built. Those that qualified the maximum number of parameters were categorized as the best set of safe natural small molecules, that could be recommended to be taken forward for further studies in the laboratory. In addition to this, the kind of toxicities were also taken into consideration for designing a safety profile, and not just the number, since the toxicity kind plays a crucial role in the altering the human protein pathways. The final profile built can be analyzed using graphical representations such as bar charts for better visualizations.

Expected result

The figures below show two ways of expressing the expected outcomes of the safety profile.

Ligands	Toxicity class 1	Toxicity class 2	Toxicity class 3	Toxicity class 4	Toxicity class 5	Carcinogenicity	Mutagenicity	Hepatoxicity	Immunotoxicity	Cytotoxicity	Tox21-Nuclear receptor singallig pathways	Tox21-Stress response pathways	Ames Mutagenicity	Developmental toxicity	GI absorption	BBB permeation	Skin permeation	Docking & MDS studies
A	-	-	Yes	-	-	Yes	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
B	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
C	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
D	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
E	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	No	Yes	No	Stable
F	-	-	-	Yes	-	Yes	No	No	Yes	No	No	No	No	Yes	Yes	Yes	No	Stable
G	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
H	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
I	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
J	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
K	-	-	-	-	Yes	No	No	No	Yes	No	No	No	No	Yes	No	No	No	Stable
L	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
M	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Unstable
N	-	-	-	Yes	-	No	No	Yes	No	No	No	No	No	No	Yes	Yes	No	Unstable
O	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
P	Yes	-	-	-	-	Yes	No	No	No	No	No	No	No	Yes	No	Yes	No	Unstable
Q	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
R	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
S	-	-	Yes	-	-	No	Yes	No	No	No	No	No	Yes	Yes	No	No	No	Stable
T	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
U	-	-	-	Yes	-	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	No	Unstable
V	-	-	-	Yes	-	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable
W	-	-	-	-	Yes	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	Stable

Figure 3: Expected safety profile built for a given set of natural small molecules after considering all parameters. The criteria that the molecules have qualified are given in green, those that are in warning or require further analysis are given in red.

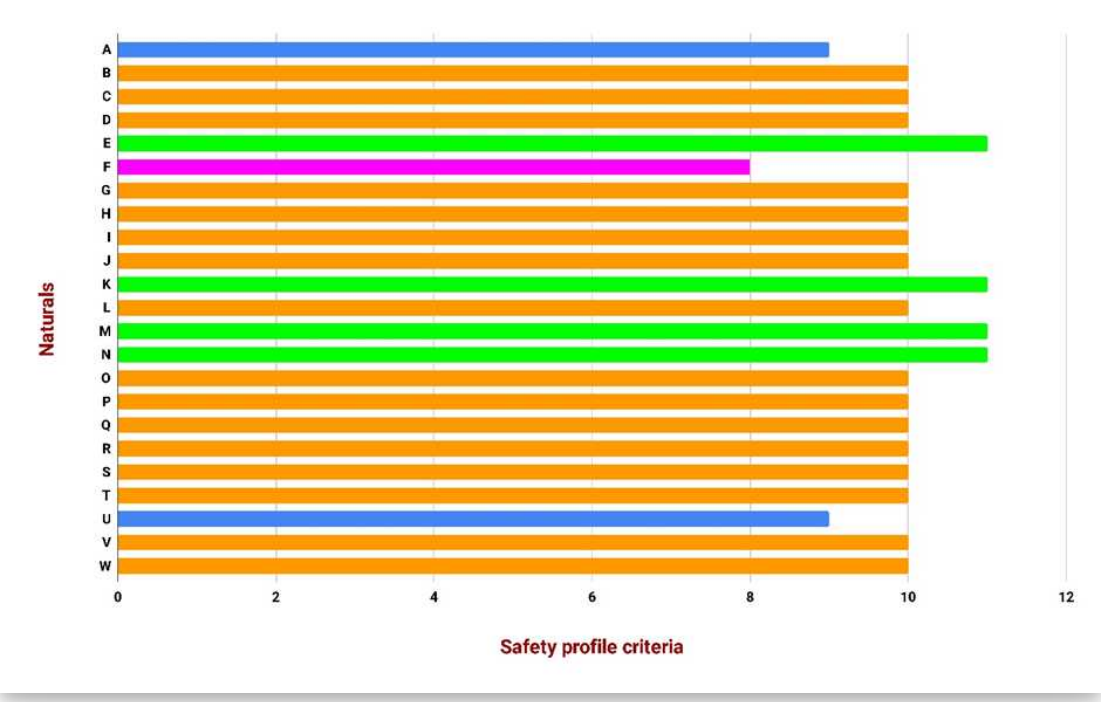

Figure 4: Expected graph of natural small molecules v/s the criteria they fulfilled. The graph indicates the final set of recommendations as per safety profile for a given set of natural small molecules. Green indicates 1st set of recommendations, followed by orange, blue and pink bars. The colors maybe altered as per user requirements.

CONCLUSIONS

9 This protocol has been tried and tested against a given set of natural small molecules to assess their probable safety and develop a profile for the same. The methodology involves moving from predicting the oral, aspiration and dermal toxicity values via *in-silico* tools to identifying all possible off-targets in humans for the potential repellent molecules, by screening through the whole human proteome and shortlisting the targets based on a set of selection criteria mentioned in the protocol. Once the final number of possible interacting proteins that the molecules bind to are finalized, their interactions can be studied statically and dynamically via molecular docking and 100ns simulations. The aim here is to have those molecules that do not interact and affect the human protein pathways since they will be typically used in repellent/insecticide formulations. However, for better clarity, all best and poorly bound complexes can be simulated. Considering all parameters, an evaluation of the safety parameters profile maybe designed and a set of best recommendations maybe provided. This study also has future perspectives in terms of performing free energy of binding studies to obtain further insights into the complex stabilities and to use additional tools to further confirm these predictions. This protocol can be used for building an *in-silico* safety profile for not only natural repellent molecules, but also for any molecule whose potential safety against human proteome has to be tested.

DECLARATIONS

10 **Conflict of Interest**

Authors Vartul Panhalkar, Sonia Sharma and Manas Sarkar are employees of Reckitt Benckiser India, Gurgaon, Haryana. The remaining authors declare no conflicts of interest.

Funding

This research work was funded by Reckitt Benckiser India Pvt. Ltd., Gurgaon, Haryana.

Acknowledgement

We would like to acknowledge Dr. Anala M, Professor, Department of Information Science and Engineering, R V College of Engineering, Bangalore, India, for providing us Titan X GPU for high performance computing. We would also like to thank our research interns Aajnaa Upadhaya, Amulya Rao, Riya Sharma and Abhipsa Rath, from RV College of Engineering, who helped immensely in data collection and management.