

# Structure-based Assessment of Homologous Analogues of Natural products: A computational approach to predict the therapeutic effects of natural products

Aditya Rao (✉ [adityaraosj@gmail.com](mailto:adityaraosj@gmail.com))

CSIR-CFTRI: Central Food Technological Research Institute CSIR <https://orcid.org/0000-0002-2546-5109>

Nandini Shetty

CSIR-CFTRI: Central Food Technological Research Institute CSIR

---

## Research Article

**Keywords:** Drug design, Synthetic drugs, Pharmacokinetics, Structural scaffolds, Activity cliff, Structural fingerprint

**Posted Date:** June 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-620117/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

The present study describes a novel strategy to screen natural products (NPs) for their therapeutic effects with the predicted mechanism of action. The method entitled '*Structure-based Assessment of Homologous Analogues of Natural products-SAHANA*' follows the comparison of NPs against prescribed synthetic chemical drugs to deduce activity cliffs and core fragments, based on the molecular properties and 2-dimensional structural similarities. The method was applied to predict the biological effect of the identified NPs as antidiabetic molecules. Selected NPs were assessed for their pharmacokinetic and pharmacodynamics properties. The biological interactions and structural stability of the bound structures were evaluated using molecular docking and molecular dynamics simulations. The study yielded NPs with significant structural similarities to prescribed drugs. Further, their binding interactions stabilized the macromolecular structure. The results envisage a strong indication that the natural products can produce therapeutic effects efficiently if administered individually. The results also encourage using the current screening strategy to identify competent natural product drugs against any disease condition ad libitum.

## 1. Introduction

Searching and identifying chemical compounds as therapeutic molecules are considered an important aspect of modern drug design. It is based on the assumption that structurally similar molecules frequently exhibit similar functions. This assumption has fuelled many drug-designing programs to develop many two-dimensional and three-dimensional similarity-based approaches, which are highly efficient [1]. However, an effective strategy should involve a close coupling of search algorithms and database implementation [2]. Several strategies have been proposed in this avenue, emphasizing core-structure and sub-structure search across the databases. This includes searching compounds by converting structures to unique string sequences like WLN (Wiswesser Line Notation), ROSDAL (Representation of Organic Structures Description Arranged Linearly), or SMILES (Simplified Molecular Input Line Entry System) and perform a simple string search [3]. The modern search approaches employ advanced computational methods and data structure modules to develop new search algorithms [4–6]. However, as these search algorithms give preference to structures and not functions, it becomes necessary to filter molecules with predictable biological effects.

Molecular similarities are generally considered in optimizing the potency and druglikeness of lead compounds to derive structure-activity relationships [1]. The mathematical models used to predict pharmacokinetic (PK) and pharmacodynamic (PD) properties can describe the drug-target association and possible mechanisms that evoke a physiological response. Based on these characters the drugs' affinity and efficacy are predicted. The therapeutic effect is generally associated with a specific target playing a critical role in the disease mechanism [7]. Validating such links and targeting them to interfere with biological processes results in controlling or managing several diseases. The underlying mechanism of action of such target molecules would be the inhibition by the drug molecule (reversible or irreversible), as an agonist or antagonist for a receptor molecule, intercalator (binder), modifier (alkylating agent) or substrate mimicking for a DNA molecule, membrane blocker or opener for cell membrane ion channels or uptake inhibitors for a transporter molecules.

### Structure resemblance as a basis to ascertain function:

The structural arrangements attribute to the biological activity of any molecule. If two molecules have a similar structure, they will most probably have a similar biological effect [1, 8, 9]. Though some converse effects are being reported [10, 11], computational chemists successfully exploited this principle to construct diverse compound libraries and select compounds for high-throughput screening experiments [8]. Such similarities have driven the synthesis of many small molecule drugs that share maximum identity with natural products [12].

Unlike synthetic compounds, which usually cover a small range of chemical diversities, natural compounds exhibit more versatile functionality. As they are synthesized from enzyme-catalyzed reactions, the asymmetric factors play a critical role in imparting their pharmacological effect. Hence, it is possible to predict their biological behavior by considering the molecular fingerprints compared with structurally similar synthetic drugs [9, 13]. As the action mechanism of chemically synthesized drugs is well defined, their structurally similar natural products are also expected to exert similar effects.

### Overview of the methodology:

'Structure-based Assessment of Homologous Analogues of Natural products- *SAHANA*' describes a novel approach to screen natural products by considering structural similarities and comparing with synthetic chemical drugs, emphasizing their therapeutic potentials. The comparison was made based on the molecular composition and two-dimensional structural similarities between natural and synthetic drugs. The resulting molecule was assessed for possible therapeutic effect and a probable mechanism of action using different *in-silico* analyses. An overview of the methodology followed in *SAHANA* is presented in figure 1.

## 2. Material And Method

### 2.1 Data collection and library construction:

Natural product activity and species source database (NPASS) offers an extensive library of natural products with experimental-derived quantitative activity data (<http://bidd.group/NPASS/index.php>) [14]. For the comparison, SMILES (Simplified molecular-input line-entry system) data library was constructed using NPASS database consisting of 26,609 natural product structures alongside a synthetic drug library composed of 85 prescribed chemical drugs selected using Drugs.com server (<https://www.drugs.com/condition/diabetes-mellitus-type-ii.html>). The SMILES were converted to structure data using Osiris DataWarrior software [15] and used for structure comparison.

### 2.2 Implementation of the method:

### a. Pharmacophore based comparison of natural products with their synthetic drugs counterparts:

The non-redundant natural product library was compared against the synthetic chemical drug library consisted of prescribed chemical drugs for type-2 diabetes (T2D) conditions. The molecular properties were used to assign the druglikeness of each molecule considering their molecular weight, cLogP, hydrogen atom donors, hydrogen atom acceptors, and rotatable hydrogen bonds. To find structurally similar compounds rather than compounds sharing a common sub-structure, core fragment-based SAR analysis was performed by considering the most central ring structure. The similarities between the fragments were assessed based on the number of fragments that both molecules have in common, divided by the number of fragments being found in any of the two structures [15]. The structures were further analyzed for structural scaffolds based on plane ring system to determine the substructures and to define the similarity cut-off during the structure comparison. The molecular properties, activity cliff, core fragments, and structural scaffolds were predicted using Osiris DataWarrior V.4.4.3 software [15].

### b. Similarity score cut-off limit:

Natural products exist in many of their stable analogs forms in nature. Even with minor structural variations, the analogous forms of natural products can exert unique biological effects [16]. Therefore, there was a need to set an appropriate limit to filter natural products and their analogous structures. Due to their complex structures, at higher cut-off limits, such derivatives are expected to exclude, while at lower cut-off limits, the analogous structures become inclusive (Table S1). By considering such variations, a similarity cut-off limit of 60% was fixed for the comparison.

## 2.3 Method validation:

### a. State 1: *In-silico* pharmacokinetics, pharmacodynamic and toxicology profiling:

The molecular complexity is one of the major characteristics of natural products. The stereo-specificity, chirality, and cis/ trans-configuration contributing to the structural properties of the natural product are very specific. The presence of aromatic rings, larger macrocyclic aliphatic rings, lower nitrogen content, and increased oxygen content contribute majorly to their structural complexity [17]. The oral bioavailability of natural products can be predicted by considering such structural properties. In pharmaceutical research, Rule of five or Lipinski rule-of-five (RO5) has become one of the widely used computational approaches to estimate solubility and permeability of new drug candidates [18]. Although these parameters cannot be strictly applied to filter natural products, they can be used as a 'rule of thumb' to describe the molecular properties necessary to screen a candidate drug's pharmacokinetic (PK) and pharmacodynamics (PD) [19, 20].

The oral bioavailability assessment was done using Osiris Data warrior V.4.4.3 [15] to analyze mutagenicity, tumorigenicity, reproductive effectiveness, irritant properties, druglikeness, and basic chemical properties like total molecular weight ClogP, H-acceptors, H-donors, rotatable bonds as RO5 filters. Bioactivity score was predicted using molinspiration [19]. The admetSAR server [21] was used to predict solubility, permeability, drug-likeness, GPCR ligand, ion channel modulator, protease inhibitor, kinase inhibitor, enzyme inhibitor, nuclear receptor inhibitor, aqueous solubility, TPSA (Topological polar surface area), blood-brain barrier penetration, human intestinal absorption, Caco-2 permeability, AMES toxicity, carcinogenicity and acute oral toxicity [20, 22].

### b. Stage 2: High throughput virtual screening to predict Molecular interactions:

The selected molecules were subjected to high throughput virtual screening (HTVS) using automated docking to identify their binding affinity with the selected targets [23, 24]. The information on the action mechanism of synthetic drugs was collected from Inxight: Drugs database of the NIH- National Center for Advancing Translational Sciences (NCATS) server (<https://drugs.ncats.io/>). We used automated docking as a screening method to predict the probable interactions of selected natural products with appropriate target proteins. The targets were selected based on the action mechanism of the most structurally similar synthetic drug (Table S2). The Protein structure files were obtained from the protein data bank and edited by removing heteroatoms and adding Kollmann charges as part of target preparation. For Empagliflozin and Luseogliflozin and their structurally similar natural products, Sodium-glucose co-transporter 2 (SGLT2) protein with PDB ID 2XQ2 was used for interaction studies. Similarly, for Prednisolone and its structurally similar natural products, Glucocorticoid receptor agonist with PDB ID 1M2Z was used. For each target, residues forming the binding site were identified using the PDBsum server. The grid box was set around the residues forming a binding pocket, and the Broyden-Fletcher-Goldfarb-Shanno algorithm implemented in the AutoDockVina was employed to study appropriate binding modes of the ligand in different conformations [25]. For the ligand molecules, all the torsions were allowed to rotate during docking [26]. The *in-silico* studies were performed on a local machine equipped with AMD Ryzen 5 six-core 3.4 GHz processor, 8GB graphics, and 16 GB RAM with Microsoft Windows 10 and Ubuntu 16.04 LTS dual boot operating systems.

### c. Stage 3: United-atom molecular dynamic simulation studies to predict the protein stability:

United-atom molecular dynamics (MD) simulation studies analyze the physical movement of atoms and molecules. To assess the structural stability upon ligand binding [27, 28], MD simulations were run for a time scale of 20 ns. Natural product-bound targets with the lowest binding energy were selected from each docked group and subjected to MD simulations. For simulation studies, GROMOS96 54a7 [29] force field implemented in GROMACS package (Abraham *et al.*, 2015), version 2018 was employed. Briefly, the system was solvated using the simple point charge (SPC) model [30] and neutralized using sodium and chloride ions. Energy minimization was done using the steepest descent method followed by temperature coupling (300K) using V-rescale thermostat [31] and pressure coupling ( $10^5$  Pa) using Parrinello-Rahman barostat [32]. LINCS algorithm [33] was employed to adjust bond length constraints, and particle mesh Ewald method (PME) [34] was used to evaluate electrostatic interactions. Equilibration trajectories were prepared for a time scale of 1ns under periodic boundary conditions. Final MD trajectories were prepared for a time scale of 20ns at a time step of 2fs with trajectory coordinates regularly written at 10ps intervals. Thus produced MD trajectories were analyzed using *gmx energy*, *gmx rms*, *gmx rmsf*, *gmx gyrate*, *gm xhbond*, *gmx do\_dssp*, *gmx covar*, *gmx anaeig*, and *gmx sasa* modules of GROMACS along with interaction energies in terms of electrostatic and van der Waals energy between the ligand and the macromolecule.

### **Biding free energy calculations using g\_mmpbsa:**

Molecular mechanics/Poisson-Boltzmann surface area (MMPBSA) is one of the most widely used approaches for estimating the binding free energy of a protein-ligand complex. For MMPBSA calculations, new trajectory files were created from each trajectory considering the final 10 ns, with frames generated at every 200ps using the *gmx trjconv* module. The binding energy components were calculated using the *g\_mmpbsa* package [35]. For calculation of polar solvation energy, *g\_mmpbsa* relies on the Adaptive Poisson-Boltzmann Solver (APBS) program. For non-polar solvation energy calculation, the Solvent Accessible Surface Area (SASA) model was used. Molecular mechanical (MM) energy consisted of electrostatic (Eelec) and van der Waals (EvdW) contributions. The *g\_mmpbsa* package uses the following equation to calculate the binding energy of the protein-ligand complex;

$$\Delta G_{\text{Binding}} = G_{\text{Complex}} - (G_{\text{Protein}} + G_{\text{Ligand}}) \quad (I)$$

The 'G' term can be further decomposed into the following components-

$$\Delta G = \Delta E_{\text{MM}} + \Delta G_{\text{Solvation}} - T\Delta S = \Delta E_{\text{(Bonded + Non-bonded)}} + \Delta G_{\text{(Polar + Non-polar)}} - T\Delta S \quad (II)$$

Where,

$G_{\text{Complex}}$  = total free energy of the binding complex,

$G_{\text{Protein}}$  and  $G_{\text{Ligand}}$  = total free energies of protein and ligand, respectively.

$E_{\text{MM}}$  = vacuum potential energy;  $G_{\text{Solvation}}$  = free energy of solvation

In equation (II),  $E_{\text{MM}}$  term consists of  $E_{\text{bonded}}$  and  $E_{\text{non-bonded}}$  energies.  $E_{\text{non-bonded}}$  energy is the sum of van der Waals (Lennard-Jones potential function) and electrostatic (Coulomb potential function) energies.  $G_{\text{solvation}}$  consists of the sum of  $G_{\text{polar}}$  and  $G_{\text{non-polar}}$  terms, which represent the electrostatic and the non-electrostatic energies.  $G_{\text{polar}}$  is calculated based on a continuum implicit solvent model using the Poisson-Boltzmann equation [36]. The  $G_{\text{non-polar}}$  term consists of the sum of  $G_{\text{cavity}}$  and  $G_{\text{vdW}}$ .  $G_{\text{cavity}}$  is the work done by the solute to create a cavity in the solvent, and  $G_{\text{vdW}}$  is the attractive van der Waals energy between solvent and solute.  $G_{\text{non-polar}}$  accounts for the hydrophobic effect [37]. The entropic term ( $T\Delta S$ ) denotes the product of the temperature and the entropic contribution.

## **3. Results**

### **3.1 *In-silico* pharmacokinetic, pharmacodynamic, and toxicology profiling:**

The comparison within the T2D synthetic chemical drugs library, performed to support the similar actions, supported their similarities in mechanisms of action (Table S2). The comparison of the natural product library with the chemical drug library yielded 13,609 structures. The structures were further filtered based on their availability in edible sources, which yielded 166 structures [38, 39]. The most similar structures from each chemical drug group were selected for further studies (Table 1). The druglikeness estimation indicated that out of 15 molecules identified, 09 molecules with positive scores demonstrated druglikeness, while 14 molecules found to penetrate the human intestine, 09 molecules penetrating the blood-brain barrier, and none of them being the Cytochromes P450 substrate indicating the high possibility of their bioavailability (Table 2A). The result of the *in-silico* PK/PD studies showed that, out of 15 molecules, 02 were predicted to be mutagenic while none of the molecules were tumorigenic, without any reproductive effects or irritancy (Table 2B).

### **3.2 High throughput virtual screening to predict Molecular interactions:**

The *in-silico* molecular interaction studies predicted the most effective natural product to bind to the appropriate target [23, 24, 40]. The binding was compared with the structurally similar synthetic drug counterparts whose targets were used for the interaction studies. The study yielded natural products effectively binding to their respective targets (Table 3). Naringin, Kaempferol-3-neohesperidoside, Ellagitannin, and Gallotannin, showed interaction energies ranging between -11.9 to -11.0 kcal/mol. Isoquercitrin, Rutin, Hesperidin, Procyanidin, and Phlorizin showed greater interaction energies with SGLT2 ranging between -12.4 to -1.0 kcal/mol. However, Ursolic acid, Oleanolic acid, Gymnemic acid, Beta-sitosterol, and Stigmasterol showed prominent interactions with the Glucocorticoid receptor, their chemical drug counterpart; prednisolone showed higher interaction energy of -1.0 kcal/mol. The effectiveness of binding was studied by considering the protein stability using molecular dynamics simulation studies.

### **3.3 United-atom molecular dynamic simulation studies to predict protein stability:**

In the present study, MD simulations were performed to confirm the accuracy of binding resulted from docking studies. MD simulation results displayed stable conformational changes acquired by the target proteins upon ligand binding (fig 4). The MD trajectory analysis revealed that Naringin and Isoquercitrin imposed fewer RMS deviations and RMS fluctuations than their structurally similar SGLT2 inhibitors; Empagliflozin and Luseogliflozin. Similarly, Ursolic acid displayed fewer RMS deviations and RMS fluctuations compared to glucocorticoid receptor inhibitor; Prednisolone. The binding free energy calculated using the *g\_mmpbsa* module revealed better interactions between selected natural products with their respective targets than their synthetic drug counterparts (Table 4).

## **4. Discussion And Conclusion**

The present research describes a novel method to screen natural products by considering their molecular composition and 2-dimensional structural arrangements, followed by their effectiveness against a specific disease condition. The method relies on the central foundation of medicinal chemistry that structurally similar molecules will have similar biological effects [1, 8]. As an initial part of the study, the natural products were compared against antidiabetic synthetic chemical drug library based on their molecular and 2-dimensional structural similarities. The identified natural products were checked for druglikeness, pharmacokinetics, and pharmacodynamics properties by comparing them against a large group of already reported datasets [19, 21]. The biological interactions involved in the action mechanism of synthetic drugs were extended to their structurally similar natural products. The interaction studies described the binding interactions of the selected molecules against suitable drug targets. The effectiveness of binding was studied using MD simulation, reporting the most effective drug-like natural product to treat T2D conditions. Based on the molecular interactions, and target's structural stability induced by natural products, it is convincing to state that the selected natural products will exert effects similar to their synthetic drug counterparts.

Several approaches have proposed chemical compounds search and identification methods based on conventional string sequence search to modern data structure module-based search algorithms [3–6]. The similarity measurements generally consider either structural representations involving physicochemical properties, topological indices, molecular graphs, pharmacophore features, molecular shapes, molecular fields, or quantitative measures involving Tanimoto coefficient, Dice index, cosine coefficient, Euclidean distance, Tversky index [1]. The 3-dimensional confirmation-based searching methods include pharmacophore modeling, shape similarity, molecular field-based methods, 3D fingerprints to compare chemical compounds. Popular chemical databases like PubChem uses percent similarity measures employing Tanimoto equation [41] and a dictionary-based fingerprint, analogous to the Molecular ACCess System structure-based keys (MACCS) [42]. Tools like SwissTargetPredictor predicts the targets for bioactive molecules based on a combination of 2D and 3D similarity measures but fails to predict the biological effects if the prediction accuracy is significantly lower for molecules with unknown bioactivity [9].

The OSIRIS DataWarrior program, employed in the current study to assess molecular similarity, uses chemical similarity-based methods that rely on substructure fragments to biological similarity considering 3D-geometry and binding [15]. The current method uses these calculations uniquely to screens natural products against specific disease conditions by targeting a drug target. Unlike SwissTargetPredictor, the present method can be employed even for molecules whose bioactivity is unknown. Further, it proceeds for bioactivity assessment based on molecular interactions and structural stability using molecular docking and MD simulations.

## 5. Limitations And Other Considerations

Our findings encourage using the current strategy to screen natural products for any disease condition ad libitum. The method can also define the new roles of natural products with therapeutic benefits, reducing the time and cost of lead discovery and validation. The methodology described in *SAHANA* relies entirely on computational calculations. Though the results are supported by statistical accuracy, emphasis must be given to experimental validations for precise action mechanisms. Another limitation of the current method is that it cannot be applied to conditions that do not have any validated action mechanisms for their prescribed drugs. As the prediction relies on the structural similarities between natural products and the available drugs, the information on well-defined action mechanisms is a prerequisite for *SAHANA*. Nevertheless, the prediction made using *SAHANA* can act as a 'rule of thumb' in screening a large dataset of natural products against any disease condition.

## Abbreviations

ADME- Absorption, Distribution, Metabolism, and Excretion; APBS- Adaptive Poisson- Boltzmann Solver; HTVS- High Throughput Virtual Screening; MACCS- Molecular ACCess System; MD- Molecular Dynamics; MM- Molecular mechanical; MMPBSA- Molecular mechanics/Poisson-Boltzmann surface area; NCATS- National Center for Advancing Translational Sciences; NPASS- Natural product activity and species source database; NPs - Natural products; PD- Pharmacodynamics; PDB- Protein Data Bank; PK Pharmacokinetics; PME- Particle Mesh Ewald method; Rg- Radius of Gyration; RMSD- Root Mean Square Deviation; RMSF- Root Mean Square Fluctuation; R05- Rule-of-Five; ROSDAL- Representation of Organic Structures Description Arranged Linearly; SASA- Solvent Accessible Surface Area; SMILES- Simplified Molecular Input Line Entry System; SPC- Simple Point Charge; T2D- Type 2 diabetes; TPSA- Topological polar surface area; WLN- Wiswesser Line Notation.

## Declarations

**Funding:** This work is supported by the Department of Science and Technology (DST)- Science and Engineering Research Board (SERB), Govt. of India (Grant number: PDF/2018/00237).

**Availability of data and materials:** All the data used during the current study are available from the corresponding author on reasonable request.

**Competing interests:** The authors declare that they have no conflicts of interest.

**Author contributions:** ARSJ: designed and conceived the study, performed the research and wrote the manuscript. NPS: participated in the results discussion and technical support. Both the authors read and approved the final manuscript.

**Ethics approval and consent to participate:** Not applicable.

**Acknowledgment:** The authors thank the Department of Science and Technology (DST) - Science and Engineering Research Board (SERB), Govt. of India for their financial support.

## References

1. Kumar A, Zhang KYJ (2018) Advances in the development of shape similarity methods and their application in drug discovery. *Front. Chem.* 6
2. Wang H, Volarath P, Harrison R (2005) An approach in building a chemical compound search engine in oracle database. In: Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. pp 2839–2842
3. Gasteiger J, Engel T (2003) *Chemoinformatics: A Textbook*. Wiley-VCH Verlag GmbH & Co. KGaA
4. Volarath P, Wang H, Fu H, Harrison R (2004) Knowledge-based algorithms for chemical structure and property analysis. In: Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. pp 3011–3014
5. Ullmann JR (1976) An Algorithm for Subgraph Isomorphism. *J ACM* 23:31–42. <https://doi.org/10.1145/321921.321925>
6. Cordella LP, Foggia P, Sansone C, Vento M (1999) Performance evaluation of the VF graph matching algorithm. In: Proceedings - International Conference on Image Analysis and Processing, ICIAP 1999. pp 1172–1177
7. Schneider G (2013) *Madame Curie Bioscience Database*. In: Prediction of Drug-Like Properties. Landes Bioscience, Austin (TX)
8. Martin YC, Kofron JL, Traphagen LM (2002) Do structurally similar molecules have similar biological activity? *J Med Chem* 45:4350–4358. <https://doi.org/10.1021/jm020155c>
9. Gfeller D, Grosdidier A, Wirth M et al (2014) SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Res* 42:. <https://doi.org/10.1093/nar/gku293>
10. Borman S (2010) SIMILAR MOLECULES, OPPOSITE EFFECTS. *Chem Eng News* 88:11. <https://doi.org/10.1021/cen-v088n022.p011>
11. Wisén S, Bertelsen EB, Thompson AD et al (2010) Binding of a small molecule at a protein-protein interface regulates the chaperone activity of Hsp70-Hsp40. *ACS Chem Biol* 5:611–622. <https://doi.org/10.1021/cb1000422>
12. Baker DC, Chu C (1993) *Nucleosides and Nucleotides as Antitumor and Antiviral Agents*, 1st edn. Springer US
13. Capecchi A, Probst D, Reymond JL (2020) One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J Cheminform* 12:. <https://doi.org/10.1186/s13321-020-00445-4>
14. Zeng X, Zhang P, He W et al (2018) NPASS: Natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res* 46:D1217–D1222. <https://doi.org/10.1093/nar/gkx1026>
15. Sander T, Freyss J, Von Korff M, Rufener C (2015) DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J Chem Inf Model* 55:460–473. <https://doi.org/10.1021/ci500588j>
16. Liu J-H, Yu B-Y (2010) Biotransformation of Bioactive Natural Products for Pharmaceutical Lead Compounds. *Curr Org Chem* 14:1400–1406. <https://doi.org/10.2174/138527210791616786>
17. Huigens RW, Morrison KC, Hicklin RW et al (2013) A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products. *Nat Chem* 5:195–202. <https://doi.org/10.1038/nchem.1549>
18. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25
19. Jarrahpour A, Motamedifar M, Zarei M et al (2010) Petra, Osiris, and Molinspiration Together as a Guide in Drug Design: Predictions and Correlation Structure/Antibacterial Activity Relationships of New N-Sulfonyl Monocyclic  $\beta$ -Lactams. *Phosphorus Sulfur Silicon Relat Elem* 185:491–497. <https://doi.org/10.1080/10426500902953953>
20. Aditya R, Venugopal T, Jayanna N et al (2020) Bioactive isolates of *Morus* species as antibacterial agents and their insilico profiling. *Lett Drug Des Discov* 17:. <https://doi.org/10.2174/1570180817999201104120815>
21. Cheng F, Li W, Zhou Y et al (2012) AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *J Chem Inf Model* 52:3099–3105. <https://doi.org/10.1021/ci300367a>
22. Raghavendra S, Aditya Rao SJ, Kumar V, Ramesh CK (2015) Multiple ligand simultaneous docking (MLSD): A novel approach to study the effect of inhibitors on substrate binding to PPO. *Comput Biol Chem* 59:. <https://doi.org/10.1016/j.compbiolchem.2015.09.008>
23. Keiser MJ, Roth BL, Armbruster BN et al (2007) Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 25:197–206. <https://doi.org/10.1038/nbt1284>
24. Campillos M, Kuhn M, Gavin AC et al (2008) Drug target identification using side-effect similarity. *Science* 321:263–266. <https://doi.org/10.1126/science.1158140>
25. Trott O, Olson AJ (2010) Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
26. Aditya Rao SJ, Ramesh CK, Raghavendra S, Paramesha M (2020) Dehydroabietylamine, A Diterpene from *Carthamus tinctorious* L. Showing Antibacterial and Anthelmintic Effects with Computational Evidence. *Curr Comput Aided Drug Des* 16:231–237. <https://doi.org/10.2174/1573409915666190301142811>
27. Lin J-H (2012) Accommodating Protein Flexibility for Structure-Based Drug Design. *Curr Top Med Chem* 11:171–178. <https://doi.org/10.2174/156802611794863580>
28. Salsbury FR (2010) Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr Opin Pharmacol* 10:738–744
29. Schmid N, Eichenberger AP, Choutko A et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* 40:843–856. <https://doi.org/10.1007/s00249-011-0700-9>
30. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction Models for Water in Relation to Protein Hydration. pp 331–342
31. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:. <https://doi.org/10.1063/1.2408420>

32. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182–7190. <https://doi.org/10.1063/1.328693>
33. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A Linear Constraint Solver for molecular simulations. *J Comput Chem* 18:1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
34. Essmann U, Perera L, Berkowitz ML et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593. <https://doi.org/10.1063/1.470117>
35. Kumari R, Kumar R, Consortium OSDD, Lynn A (2014) g \_ mmpbsa - A GROMACS tool for MM-PBSA and its optimization for high-throughput binding energy calculations. *J Chem Inf Model* 54:1951–1962
36. Baker NA, Sept D, Joseph S et al (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037–10041. <https://doi.org/10.1073/pnas.181342398>
37. Richmond TJ (1984) Solvent accessible surface area and excluded volume in proteins. *J Mol Biol* 178:63–89. [https://doi.org/10.1016/0022-2836\(84\)90231-6](https://doi.org/10.1016/0022-2836(84)90231-6)
38. Naveen J, Baskaran V (2018) Antidiabetic plant-derived nutraceuticals: a critical review. *Eur J Nutr* 57:1275–1299
39. Yeh GY, Eisenberg DM, Kaptchuk TJ, Phillips RS (2003) Systematic review of herbs and dietary supplements for glycemic control in diabetes. *Diabetes Care* 26:1277–1294
40. Janakirama ARS, Shivayogi SM, Satyanarayana JK, Kumaran RC (2020) Characterization of isolated compounds from *Morus* spp. and their biological activity as anticancer molecules. *BiolImpacts*. <https://doi.org/10.34172/bi.2021.09>
41. Tanimoto T (1957) IBM Internal Report 17th
42. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42:1273–1280. <https://doi.org/10.1021/ci010132r>

## Tables

**Table 1:** Natural products structurally similar to T2D prescribed drugs and their similarity score.

Prescribed T2D drugs	Structurally similar Natural products	Similarity score
<b>Empagliflozin</b>	Naringin	0.6875
	Kaempferol-3-neohesperidoside	0.6084
	Hesperidin	0.6953
	Ellagitannin	0.6582
	Gallotannin	0.6582
	Phlorizin	0.6258
	Procyanidin	0.7029
	Mangiferin	0.8138
<b>Luseogliflozin</b>	Isoquercitrin	0.6238
	Naringin	0.6826
	Kaempferol-3-neohesperidoside	0.6250
	Rutin	0.6126
	Hesperidin	0.6957
	Procyanidin	0.7483
	Phlorizin	0.6610
	Mangiferin	0.8782
<b>Prednisolone</b>	Ursolic acid	0.7660
	Oleanolic acid	0.7695
	Gymnemic acid	0.6193
	Beta-sitosterol	0.6716
	Stigmasterol	0.6716

**Table 2:** A) Molecular properties and Pharmacokinetics prediction of natural products filtered in for screening against T2D condition \*.

Identified NPs	cLogP	Mol.wt	H-Acceptors	H-Donors	Rotatable Bonds	Total Surface Area	Polar Surface Area	Druglikeness	Human intestinal absorption	Caco-2 permeability	Blood brain barrier
Beta-sitosterol	12.864	815.269	07	03	25	664.41	105.45	-27.511	0.884+	0.851-	0.613+
Ellagitannin	0.0531	636.47	18	11	10	418.64	310.66	0.19723	0.758+	0.874-	0.511+
Gallotannin	0.0531	636.47	18	11	10	418.64	310.66	0.19723	0.866+	0.874-	0.511+
Gymnemic acid	4.2394	847.048	14	07	11	603.23	229.74	-6.9506	0.779+	0.922-	0.539+
Hesperidin	-0.814	610.563	15	08	07	410.61	234.29	2.0396	0.816+	0.895-	0.946-
Isoquercitrin	-0.3469	464.378	12	08	04	303.75	206.6	-3.6679	0.646+	0.939-	0.697-
Kaempferol-3-neohesperidoside	-1.9742	740.662	19	11	08	485.32	304.21	0.68673	0.649+	0.910-	0.867-
Mangiferin	-0.4299	422.341	11	08	02	270.83	197.37	-3.0467	0.886+	0.916-	0.647+
Naringin	-0.744	580.537	14	08	06	388.35	225.06	0.64246	0.753+	0.918-	0.841-
Oleanolic acid	0.055	436.412	10	07	07	305.5	177.14	-4.8731	0.402-	0.826-	0.562+
Phlorizin	2.7421	594.523	13	10	04	393.21	229.99	0.1505	0.907+	0.878-	0.511+
Procyanidin	-1.2573	610.519	16	10	06	397.71	265.52	1.9337	0.732+	0.917-	0.854-
Rutin	7.603	412.699	01	01	05	335.28	20.23	1.2217	0.991+	0.795+	0.973-
Stigmasterol	7.603	412.699	01	01	05	335.28	20.23	1.2217	0.985+	0.835+	0.776+
Ursolic acid	6.0021	456.708	03	02	01	337.87	57.53	-3.658	0.884+	0.851-	0.613+

\* Details of data interpretation is given in supplementary file

**Table 2:** B) Pharmacodynamics prediction of natural products selected for screening against T2D condition \*.

Identified NPs	Mutagenic	Tumorigenic	Reproductive effective	Ocular irritancy	Aerobic biodegradability	Ames toxicity score	Carcinogen	Hepatotoxicity
Beta-sitosterol	NONE	NONE	NONE	0.990-	0.869-	0.804-	0.965-	0.575-
Ellagitannin	NONE	NONE	NONE	0.839-	0.835+	0.814-	0.744-	0.600+
Gallotannin	NONE	NONE	NONE	0.839-	0.833+	0.814-	0.744-	0.600+
Gymnemic acid	NONE	NONE	NONE	0.906-	0.910-	0.929-	0.686-	0.650+
Hesperidin	NONE	NONE	NONE	0.914-	0.883-	0.897-	0.699-	0.650+
Isoquercitrin	NONE	NONE	NONE	0.732-	0.625-	0.577+	0.714-	0.625+
Kaempferol-3-neohesperidoside	HIGH	NONE	NONE	0.905-	0.828-	0.581-	0.730-	0.800+
Mangiferin	HIGH	NONE	NONE	0.752-	0.781-	0.723+	0.725-	0.575+
Naringin	NONE	NONE	NONE	0.929-	0.711-	0.692-	0.702-	0.575+
Oleanolic acid	NONE	NONE	NONE	0.908-	0.962-	0.849-	0.596-	0.700-
Phlorizin	NONE	NONE	LOW	0.811-	0.933+	0.785-	0.746-	0.550+
Procyanidin	NONE	NONE	NONE	0.773-	0.926-	0.928-	0.578-	0.550-
Rutin	NONE	NONE	NONE	0.897-	0.833-	0.511-	0.671-	0.700+
Stigmasterol	NONE	NONE	NONE	0.967-	0.980-	0.913-	0.588-	0.775-
Ursolic acid	NONE	NONE	NONE	0.938-	0.962-	0.849-	0.596-	0.700-

\* Details of data interpretation is given in supplementary file

**Table 3:** Molecular interactions between the selected natural products with targets of their structurally similar prescribed drugs.

Prescribed drug and the identified NPs	Interaction Energy (kcal/mol)	Number of H-bonds	Interacting Residues
<b>Interactions with SGLT2 protein (PDB ID 2XQ2)</b>			
Empagliflozin	-9.3	07	ASN64, SER91, ALA259, ASN260, TYR263, TRP264, GLN428,
Naringin	-11.9	07	GLU68, SER91, TYR138, ASN260, TRP264, TYR269, SER365
Kaempferol-3-neohesperidoside	-11.7	06	ASN142, TYR263, ASN267, SER364, SER368, GLN428
Ellagitannin	-11.1	10	ASN64, SER66, GLN69, SER91, TYR263, TRP264, ASN267, TYR269, SER368, GLY428
Gallotannin	-11.0	09	ALA63, ASN64, SER91, TYR263, TRP264, ASN267, TYR269, GLN428, ASN260
Mangiferin	-9.3	07	TYR138, TYR263, ASN267, TYR269, SER368, GLN428, SER435
<b>Interactions with SGLT2 protein (PDB ID 2XQ2)</b>			
Luseogliflozin	-7.0	03	TYR263, ALA63, ALA62
Isoquercitrin	-12.4	05	GLU68, GLN69, SER145, ASN267, GLN428
Rutin	-11.4	09	ASN64, GLN69, SER91, TYR139, ASN142, TRP264, TYR269, ASN260, TYR263
Hesperidin	-11.2	08	ASN64, ILE65, SER91, ASN142, TRP264, ASN260, TYR263, SER365
Procyanidin	-9.7	05	TYR269, ARG273, SER364, SER365, SER372
Phlorizin	-10.1	08	ASN64, GLU68, SER91, ASN142, TYR263, TRP264, ASN267, GLN428
<b>Interactions with Glucocorticoid receptor (PDB ID 1M2Z)</b>			
Prednisolone	-10.0	02	TRP610, ARG614
Ursolic acid	-8.9	00	-
Oleanolic acid	-9.0	01	GLN615
Gymnemic acid	-7.9	02	ARG558, THR562
Beta-sitosterol	-7.4	01	GLU537
Stigmasterol	-6.8	01	VAL538

**Table 4:** Calculated MD parameters for native and ligand bound T2D drug targets obtained after the simulation along with binding energies and the contributing energy terms of prescribed drug and its most similar natural product calculated using g\_mmpbsa module.

Gromacs Modules	SGLT2 protein			SGLT2 protein			Glucocorticoid receptor	
	Native Protein	Empagliflozin	Naringin	Native Protein	Luseogliflozin	Isoquercitrin	Native Protein	Prednisolone
Potential Energy (x 10 <sup>-6</sup> )	-1.581	-1.580	-1.580	-1.581	-1.5810	-1.585	-0.994	-1.138
RMSD (nm)	0.289	0.285	0.246	0.289	0.304	0.295	0.256	0.337
RMSF (nm)	0.145	0.111	0.109	0.145	0.128	0.119	0.113	0.123
Rg (nm)	2.334	2.333	2.341	2.334	2.326	2.324	1.865	1.876
SASA (nm <sup>2</sup> )	231.55	228.55	230.16	231.55	226.66	224.78	135.96	138.92
Secondary Structure	797.12	797.14	800.35	797.12	790.90	784.23	375.16	371.85
Coul-SR (kJ/mol)	-	-153.05	-112.01	-	-7.23	-69.11	-	-10.19
LJ-SR(kJ/mol)	-	-160.61	-184.64	-	-141.83	-97.86	-	-53.06
<b>MMPBSA Module</b>								
Binding Energy (kJ/mol)	-	-120.20	-127.71	-	-46.48	-151.83	-	-31.14
SASA Energy (kJ/mol)	-	-23.41	-26.35	-	-17.09	-12.58	-	-7.11
Polar Solvation Energy (kJ/mol)	-	206.21	213.18	-	47.93	118.10	-	48.97
Electrostatic Energy (kJ/mol)	-	-103.99	-69.08	-	-4.55	-44.91	-	-7.61
van der WAALS Energy (kJ/mol)	-	-206.01	-238.46	-	-172.78	-112.44	-	-65.39

## Figures

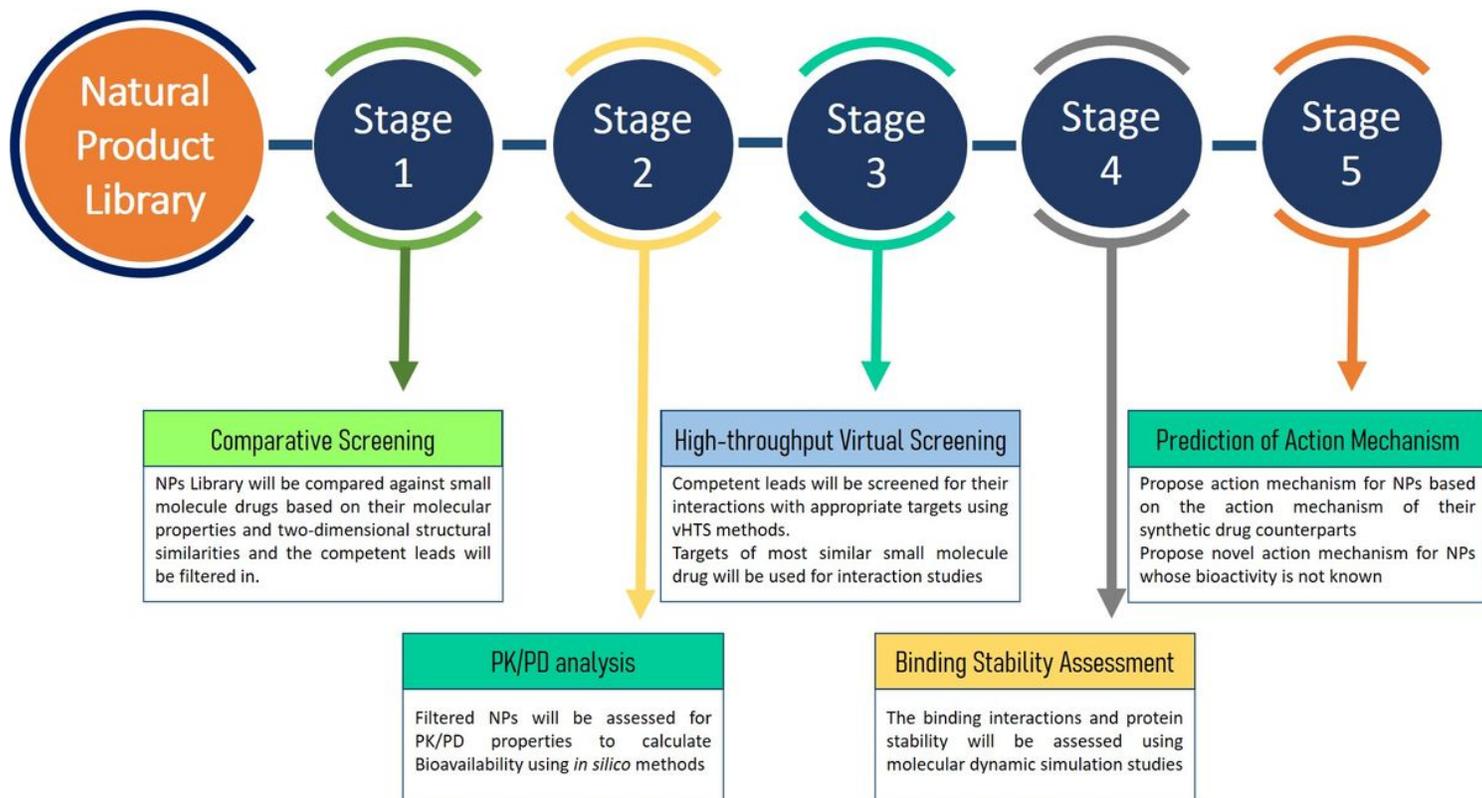


Figure 1

Overview of the methodology of SAHANA

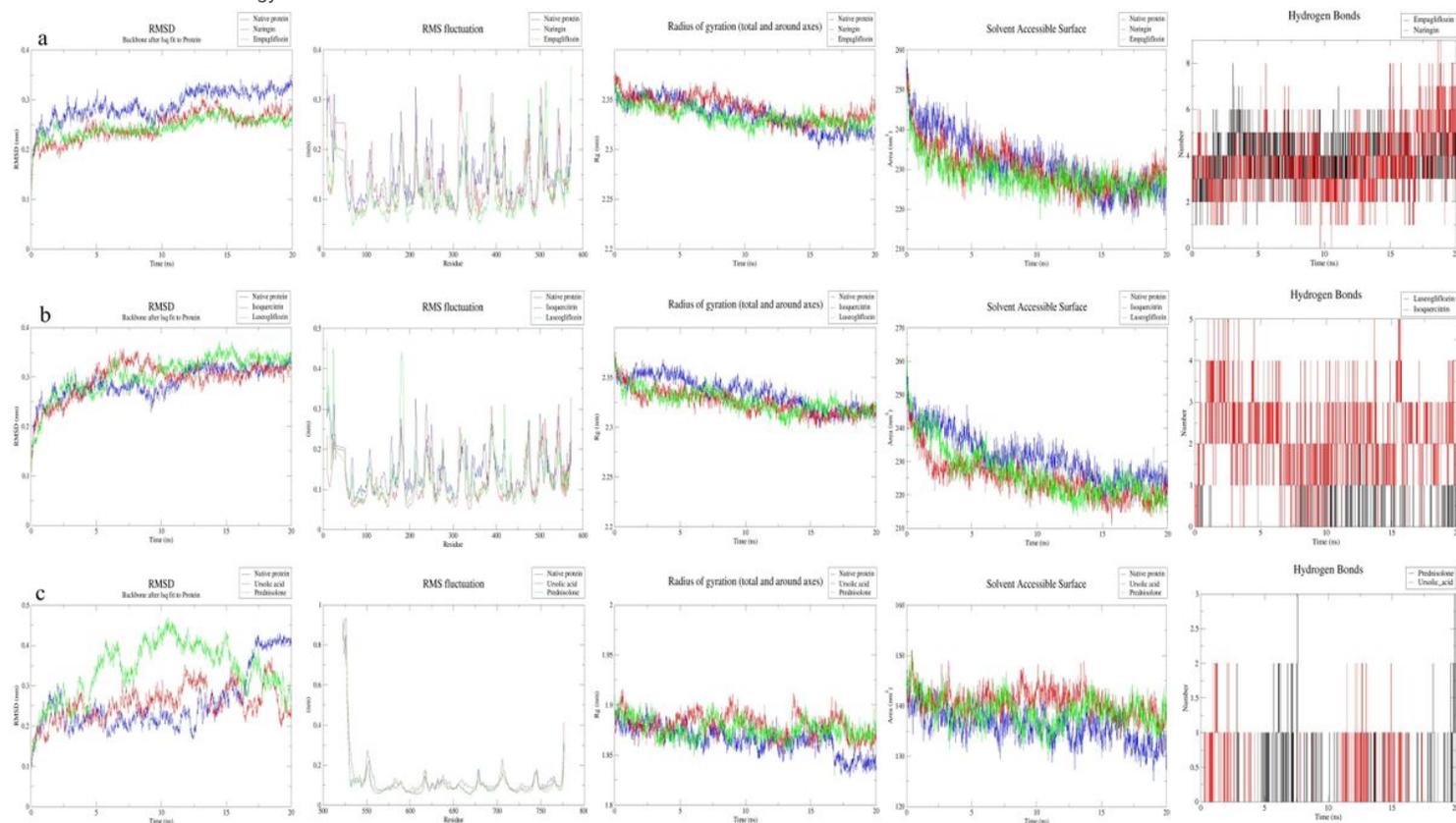


Figure 2

RMSD, RMSF, Rg, SASA, and inter-molecular H-bond plots obtained from MD trajectories analysis of native, Natural product bound, and prescribed chemical drug-bound structures of SGLT2 receptor (a & b) and glucocorticoid receptor (c).

## Supplementary Files

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

- [Supplementarydata06062021.docx](#)