Primary Target Prediction of Bioactive Molecules from Chemical Structure

Abed Forouzesh¹, Sadegh Samadi Foroushani¹*, Fatemeh Forouzesh², and Eskandar Zand¹

¹Iranian Research Institute of Plant Protection, Agricultural Research Education and Extension Organization (AREEO), Tehran, Iran
²Department of Medicine, Tehran Medical Branch, Islamic Azad University, Tehran, Iran

*To whom correspondence should be addressed. Tel: (+9821) 22400080; Fax: (+9821) 22400568; Email: ssamadi@alumni.ut.ac.ir

ABSTRACT

There are various tools for computational target prediction of bioactive molecules from a chemical structure in a machine-readable material but these tools can’t distinguish a primary target from other targets. Also, due to the complex nature of bioactive molecules, there has not been a method to predict a target and or a primary target from a chemical structure in a non-digital material (for example printed or hand-written documents) yet. In this study, an attempt to simplify primary target prediction from a chemical structure was resulted in developing an innovative method based on the minimum structure which can be used in both formats of non-digital and machine-readable materials. A minimum structure does not represent a real molecule or a real association of functional groups, but is a part of a molecular structure which is necessary to ensure the primary target prediction of bioactive molecules. Structurally related bioactive molecules with the minimum structure were considered as neighbor molecules of the query molecule. The known primary target of the neighbor molecule is used as a reference for predicting the primary target of the neighbor molecule with an unknown primary target. In results, we confirmed the usefulness of our proposed method for primary target prediction in 548 drugs and pesticides involved in four primary targets by eight minimum structures.

INTRODUCTION

Bioactive molecules such as drugs and pesticides are produced in large numbers by many commercial and academic groups around the world¹⁸. Most bioactive molecules perform their actions by interacting with proteins or other macromolecules⁵. However, for a significant fraction of them, the primary target remains unknown⁵.
Computer-aided drug design, that offers an in silico alternative to medicinal and agricultural chemistry techniques for studying the structure and predicting the biological activity of drug and pesticide candidates, has the advantages of both speed and low cost and is becoming an indispensable program of major pharmaceutical and agrochemical companies\(^6\).

With the ever-increasing public availability of bioactivity data\(^2\), it is possible to construct reliable target prediction models using statistical or machine learning methods\(^21\). However, bioactivity data hasn’t been increased in all areas. For example, the databases are rich in human targets and molecules that modulate these targets, but contain limited information when it comes to bacterial targets\(^11\).

Various methods (such as chemical structure similarity searching\(^9\), data mining/machine learning\(^17\), panel docking\(^14\) and bioactivity spectra based algorithms\(^3\)) can be employed for computational target prediction (each method adopts a particular model type). The tools created by these methods have not been able to distinguish between a primary target and other targets yet due to the focus of databases from which they have been extracted. Also, they are not equally reliable in all areas. Furthermore, chemical structures should be provided in machine-readable descriptions so that the tools can be used.

Based on our knowledge, there is no method to predict a target and or a primary target from a chemical structure in a non-digital material. In this study, an attempt to simplify primary target prediction from a chemical structure was resulted in developing an innovative method based on the minimum structure which can be used in both formats of non-digital and machine-readable materials. The proposed method has several distinctive features compared to available computational target prediction methods. Firstly, it’s easy to use. Secondly, it is highly accurate. Thirdly, it can be used appropriately in both formats of non-digital and machine-readable materials. Fourthly, it enables us to gain a deeper understanding (more informative) of the relationship between the chemical structure and the primary target.

METHODS
The proposed method steps for primary target prediction of bioactive molecules from a chemical structure include (i) query molecule, (ii) similarity searching, (iii) data collection, (iv) minimum structure identification and (v) primary target prediction (example in Supplementary Figure 9).
(I) Query molecule. A bioactive molecule such as a drug or a pesticide is used as a query molecule. The query molecule may have a known or an unknown primary target. The query molecule with known primary target can be used as a reference to predict the primary target of structurally related bioactive molecules.

(II) Similarity searching. The query molecule is used to search for the structurally related bioactive molecules with similar chemical scaffold. However, it must be born in mind that structurally related analogs may bind in a slightly or considerably different manner\textsuperscript{24}. KEGG\textsuperscript{7}, DrugBank\textsuperscript{23}, PubChem\textsuperscript{10} and ChEMBL\textsuperscript{4} databases provide common names and chemical structures for large numbers of bioactive molecules and, in some cases, their primary targets. All four databases support structure similarity searches.

(III) Data collection. The primary target information is collected for all structurally related bioactive molecules. If the primary target of the query molecule is not known, information on structure-activity relationship and pharmacophore will be collected for all structurally related bioactive molecules. If the primary target of the query molecule is known, information on structure-activity relationship and pharmacophore will be collected only for structurally related bioactive molecules with the same primary target as the query molecule.

Information on the primary target, structure-activity relationship and pharmacophore are obtained from databases with annotated primary target (such as KEGG\textsuperscript{7}, DrugBank\textsuperscript{23}, PubChem\textsuperscript{10} and ChEMBL\textsuperscript{4} databases), scientific literature and pharmacophoric descriptors (including hydrogen bonds, hydrophobic and electrostatic interaction sites). The most recent development regarding pharmacophore alignment technique is LigandScout’s pattern matching approach\textsuperscript{25}.

(IV) Minimum structure identification. A minimum structure does not represent a real molecule or a real association of functional groups, but is a part of a molecular structure which is necessary to ensure the primary target prediction of bioactive molecules. The minimum structure is identified using data collection about the structurally related bioactive molecules. The minimum structure consists of the core and or the peripheral part. Here, the peripheral part is shown as the comment. The core plays an essential role in a bioactive molecule. Furthermore, modifying at some key position on the peripheral part can make a big change in the primary target or the activity of a bioactive molecule. Thus, the peripheral part can be useful for distinguishing the bioactive molecules based on their primary targets.

Since the minimum structure is related to the structurally related bioactive molecules and information about them, when they become available, the minimum structure can be updated to further refine it.
(V) Primary target prediction. Structurally related bioactive molecules with the minimum structure were considered as neighbor molecules of the query molecule. The known primary target of the neighbor molecule is used as a reference for predicting the primary target of the neighbor molecule with an unknown primary target.

RESULTS AND DISCUSSION

In results, we made predictions for eight groups of bioactive molecules. Here, the proposed method was employed in 548 drugs and pesticides involved in four primary targets (Tables 1-4 and Supplementary Data 1). 4-Pyridone group includes 192 bioactive molecules of DNA gyrase and topoisomerase IV inhibitors (Table 1; example in Supplementary Figure 1), 2,4(or 5)-diaminocyclohexanol group includes 138 bioactive molecules of small ribosomal subunit inhibitors (Table 2; example in Supplementary Figure 2), (4aRS,5aRS)-Sancycline group includes 34 bioactive molecules of small ribosomal subunit inhibitors (Table 2; example in Supplementary Figure 3), cytosine group includes 28 bioactive molecules of large ribosomal subunit inhibitors (Table 2; example in Supplementary Figure 4), 3-glutarimidyl group includes 17 bioactive molecules of large ribosomal subunit inhibitors (Table 2; example in Supplementary Figure 5), (1R)-propanol group includes 10 bioactive molecules of large ribosomal subunit inhibitors (Table 3; example in Supplementary Figure 6), imidazol-1-yl group includes 54 bioactive molecules of sterol 14α-demethylase inhibitors (Table 4; example in Supplementary Figure 7), and 1,2,4-triazol-1-yl group includes 75 bioactive molecules of sterol 14α-demethylase inhibitors (Table 4; example in Supplementary Figure 8).

Many of these 548 predictions were either confirmed by scientific literature, database searching or computational target prediction tools. Out of 548 predictions, 371 predictions (67.7%) were confirmed by scientific literature published in scientific journals, conferences and books (4-pyridone group with 24.5%, 1,2,4-triazol-1-yl group with 11.3%, 2,4(or 5)-diaminocyclohexanol group with 11.3%, imidazol-1-yl group with 6.9%, (4aRS,5aRS)-Sancycline group with 4.7%, cytosine group with 4.6%, 3-glutarimidyl group with 2.7% and (1R)-propanol group with 1.6%; Supplementary Data 2), 160 predictions (29.2%) were confirmed by database searching (KEGG7, PubChem10, DrugBank23 and ChEMBL4 respectively with 23.2%, 16.4%, 11.3% and 10.4%; Supplementary Data 3) for annotated primary target, 546 predictions (99.6%) were confirmed by computational target prediction tools (PASS online13 with 99.3%, SEA8 with 58%, PPB1 with 54.4%, TargetHunter20 with 44.7%, ChemProt12 with 44.5%,
PharmMapper\textsuperscript{22} with 37.8\%, SuperPred\textsuperscript{16} with 8.6\%, HitPick\textsuperscript{15} with 6.4\% and SPiDER\textsuperscript{19} with 0.4\%; Supplementary Data 4). We found no precedent for one prediction.

A lot of information is needed to identify the minimum structure because it is necessary to know which part(s) of the bioactive molecule confers the activity. Therefore, the proposed method leads to a high accuracy and a deeper understanding of the relationship between the chemical structure and the primary target. Furthermore, after identifying the minimum structure, it's been easy to use and it can be used in both formats of non-digital and machine-readable materials. Due to the complex nature of bioactive molecules, there has not been a method to predict a target and or a primary target from a chemical structure in a non-digital material yet and the proposed method is the only method that can be used for this purpose.

The proposed method is not applicable in cases where no neighbor bioactive molecules for a primary target exist, since in these situations no training on the minimum structure-based information is possible.

REFERENCES

1. Awale, M. & Reymond, J. L. The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. J. cheminform. 9, 11 (2017).
2. Bento, A. P. et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res. 42, D1083–D1090 (2014).
3. Cheng, T., Li, Q., Wang, Y. & Bryant, S. H. Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining. J. Chem. Inf. Model. 51, 2440–2448 (2011).
4. Gaulton, A. et al. The ChEMBL database in 2017. Nucleic Acids Res. 45, D945–D954 (2017).
5. Gfeller, D., Michielin, O. & Zoete, V. Shaping the interaction landscape of bioactive molecules. Bioinformatics 29, 3073–3079 (2013).
6. Huang, H. J. et al. Current developments of computer-aided drug design. J. Taiwan Inst. Chem. Eng. 41, 623–635 (2010).
7. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353–D361 (2017).
8. Keiser, M. J. et al. Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. 25, 197–206 (2007).
9. Keiser, M. J. et al. Predicting new molecular targets for known drugs. Nature 462, 175–181 (2009).
10. Kim, S. et al. PubChem substance and compound databases. Nucleic Acids Res. 44, D1202–D1213 (2016).
11. Koutsoukas, A. et al. From in silico target prediction to multi-target drug design: current databases, methods and applications. *J. Proteomics* **74**, 2554–2574 (2011).

12. Kringelum, J. et al. ChemProt-3.0: a global chemical biology diseases mapping. *Database* **2016**, bav123 (2016).

13. Lagunin, A., Stepanchikova, A., Filimonov, D. & Poroikov, V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics* **16**, 747–748 (2000).

14. Li, H. et al. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **34**, W219–W224 (2006).

15. Liu, X., Vogt, I., Haque, T. & Campillos, M. HitPick: a web server for hit identification and target prediction of chemical screenings. *Bioinformatics* **29**, 1910–1912 (2013).

16. Nickel, J. et al. SuperPred: update on drug classification and target prediction. *Nucleic Acids Res.* **42**, W26–W31 (2014).

17. Nidhi, , Glick, M., Davies, J. W. & Jenkins, J. L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **46**, 1124–1133 (2006).

18. Orchard, S. et al. Minimum information about a bioactive entity (MIABE). *Nature Rev. Drug Discov.* **10**, 661–669 (2011).

19. Reker, D., Rodrigues, T., Schneider, P. & Schneider, G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. USA* **111**, 4067–4072 (2014).

20. Wang, L. et al. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.* **15**, 395–406 (2013).

21. Wang, Z., Liang, L., Yin, Z. & Lin, J. Improving chemical similarity ensemble approach in target prediction. *J. Cheminform.* **8**, 1–10 (2016).

22. Wang, X. et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res.* **45**, W356–W360 (2017).

23. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

24. Poulos, T. L. & Howard, A. J. Crystal structures of metyrapone-and phenylimidazole-inhibited complexes of cytochrome P-450cam. *Biochemistry* **26**, 8165–8174 (1987).
25. Wolber, G., Dornhofer, A. A. & Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aided Mol. Des.* **20**, 773–788 (2006).
Table 1. The bioactive molecules with primary target of bacterial type IIA topoisomerase (DNA gyrase and topoisomerase IV) inhibition predicted by the minimum structure.

| Minimum structure | Bioactive molecules |
|-------------------|---------------------|
| 4(1H)-Pyridinone or 4-Pyridone |

![Diagram of molecular structure](image)

Comments:
- $X_1 \neq H$ (weak inhibitory activity)
- $X_1 = \text{-COO; isothiazol-3(2H)-one (a) fused to the core}$
- Active ingredient ≠ no fused aromatic ring or chain attached to an aromatic ring at $X_8$ (here $X_8$ means six-membered aromatic ring fused to the core at $X_5$) in position 6; the core fused to more than two (if $X_1 = \text{-COO}$) or three (if $X_1 = \text{isothiazol-3(2H)-one fused to the core}$) aromatic rings.

A 57132, A 57241, A 57274 (A 62917), A 60919 (BRN 4276829), A 62251 (A 57531; PD 137954), A 62255, A 62824, A 65326, ACH 702, Acorafloxacin (avarofloxacin), ADDNC (A 65485), Alalevonidifloxacin, Alatrofloxacin, Amifloxacin, Antofloxacin, AT 4929, Balofloxacin, BAY Y-3118 free base, Besifloxacin, Binfloxacin, BMY 40062, BMY 40597, BMY 42230, BMY 43261, BMY 43748, BMY 45243, BMY 45706, BRN 4913428 (PD 131199), Cadrofloxacin, Cetefloxacin, Chinfloxacin, CI 990 (PD 131112), Ciprofloxacin, Clinafloxacin, CP 100964, CP 104830, CP 105532 (PD 125275), CP 115953, CP 115955, CP 135803, CP 67015, CP 67804, CP 74667, CP 92121, CP 99433, Danofloxacin, DC 159a free base, Delafloxacin, Desfluorociprofloxacin (SQ 4004), Difloxacin, DJ 6783, DK 507k, DN 9494, Droxacin, E 3604, E 3846, E 4441, E 4474, E 4480, E 4497, E 4501, E 4502, E 4527, E 4528, E 4534, E 4535, E 4695, Ecnofloxacin, EN 272, Enoxacin, Erofloxacin, Esafloxacin, FA 103, Fandofloxacin, Finafloxacin, Fleroxacin, Flumequine, Garenoxacin, Gatifloxacin, Gemifloxacin, Grepafloxacin, Ibafloxacin, Irloxacin, K 12, KB 5246, KPI 10 free base (WQ 3810), Lascufloxacin, Levofloxacin, Levonadifloxacin arginine (WCK 771), Lomefloxacin, Marbofloxacin, Merafloxacin, Metioxate, MF 5101, MF 5103, MF 5112 free base, MF 5126, MF 5137, MF 5143, MF 5168, Miloxacin, Moxifloxacin, Nadifloxacin, Nalidixic acid, Nemonoxacin, Norfleroxacin, NSFQ 104, NSFQ 105, Olamufloxacin, Oxolinic acid, Ozenoxacin, Pauflufloxacin, PD 111834, PD 112388, PD 114111, PD 115311, PD 116507, PD 117596, PD 118362, PD 119344, PD 129626, PD 131628, PD 135042 (AM 1147), PD 135144 (BMY 33315), PD 13716, PD 138312, PD 140248, PD 163449, PD 164488, Pefloxacin, Pipemidic acid, Piromidic acid, Piroxacin, Predofloxacin, Prulifloxacin, PubChem CID-11531032, PubChem CID-11566845, PubChem CID-11610627, PubChem CID-11996318, PubChem CID-11996799, PubChem CID-11997263, PubChem CID-25028697, PubChem CID-44408626, PubChem CID-44408894, PubChem CID-44408894, PubChem CID-44409010, PubChem CID-53236573, PubChem CID-53236796, QA 241 free base, RO 13-5478, RO 14-9578, Rosoxacin, Rufloxacin, S 25932, S 31076, Sarafloxacin, Sefloxacin, Sparfloxacin, T 14097, Temafloxacin, Troloxacin, Tosalifloxacin, Trovatofloxacin, Uloxifloxacin, Vefubloxacin (benofloxacin), VG 6/1, WCK 1152 free base, WIN 57273, WIN 57294, WIN 58161, WQ 2743, WQ 2756, WQ 2908, WQ 2942, WQ 3330, Y 688, Zabofloxacin.
Table 2. The bioactive molecules with primary target of small ribosomal subunit inhibition predicted by the minimum structure.

| Minimum structure | Bioactive molecules |
|-------------------|---------------------|
| 2,4(or 5)-Diaminocyclohexan-1-ol or 2,4(or 5)-Diaminocyclohexanol | 1-Epididimicin, A 396I (SS 56D), Ambistrin (streptoduocin), Amikacin, Apramycin (nebramycin II), Aprosamine, Arbekacin, Astromicin (fortimicin A), Astromicin B (fortimicin B), Bekanamycin (kanamycin B; nebramycin V), Betamicin (gentamicin B), Butikacin, Butirosin A, Butirosin B, Dactimicin, Destomycin A, Destomycin B, Dibekacin, Dihydrostreptomycin, Etimicin, Fortimicin AE, Fortimicin AH, Fortimicin AI, Fortimicin AK, Fortimicin AL, Fortimicin AM, Fortimicin AN, Fortimicin AO, Fortimicin AP, Fortimicin AQ, Fortimicin AS, Fortimicin C, Fortimicin D, Fortimicin E (fortimicin KH), Fortimicin KE, Fortimicin KF, Fortimicin KG, Fortimicin KL, Fortimicin KR, Framycetin (neomycin B), Geneticin (gentamicin G-418), Gentamicin A, Gentamicin A1, Gentamicin A2, Gentamicin A3, Gentamicin B1, Gentamicin C1, Gentamicin C2, Gentamicin C3, Gentamicin C4, Gentamicin X1, Hybrimycin A, Hybrimycin A1, Hybrimycin B, Hybrimycin B1, Hybrimycin C, Hybrimycin C1, Hybrimycin D, Hygromycin B (A 396II), Iseamicin, Istamycin A (sannamycin A), Istamycin A0 (sannamycin B), Istamycin A1, Istamycin A2, Istamycin A3, Istamycin AO, Istamycin AP (sannamycin E), Istamycin B, Istamycin B0, Istamycin B1, Istamycin B2, Istamycin C, Istamycin C0, Istamycin C1, Istamycin KL, Istamycin Li, Istamycin X0 (sannamycin G), Istamycin Y0 (sannamycin H), Kanamycin (kanamycin A), Kanamycin C, Kanamycin D, Kanamycin X, Lividamine (nebramycin IX), Lividomy cin, Lividomycin B (3'-deoxytromomycin I), Mannosylparomomycin, Micromycin (gentamicin C2a), Neamine (neomycin A; nebramycin X), Nebramine (nebramycin VIII), Nebramycin III, Nebr amycin IV, Nebramycin V', Nebramycin XI, Nebramycin XII, Nebramycin XIII, Neomycin C, Neomycin F (paromomycin II), Netilimicin, NK 1001, Oxyapramycin (nebramycin VII), Paromamine (neomycin D), Paromomycin (paromomycin I; neomycin E), Pentisomicin, Plazomycin, Propiakcin, Piranakacin, Ribostamycin, Saccharocin (KA 5685), Sannamycin C, Sannamycin F, Sannamycin J, Sannamycin K, Sannamycin KR, Sannamycin L, Seldomy c in, Seldomy cin 1 (seldomy cin factor 1), Seldomy cin 2 (seldomy cin factor 2), Seldomy cin 3 (seldomy cin factor 3), Seldomy cin 5 (seldomy cin factor 5), Sisomicin, Sisomicin B, Sisomicin D, Spectinomycin, Sparacin A, Sparacin B, Sparacin C, Sparacin D, Sparacin E, SS 56A, SS 56B, SS 56C, Streptomycin, Streptozionicid (streptozionicid), Tobramycin (nebramycin VI), Tro spectomycin, Verdamicin, Verdamicin C1, Verdimycin |

Comment: The active ingredient comprises a ring attached to the core directly or indirectly.

(4S,12aS)-4-(Dimethylamino)-3,10,12,12a-tetrahydroxy-1,11-dioxo-1,4,4a,5,6,11,12a-octahydrotetracene-2-carboxamide or (4aR5,5aR5)-Sancycline

7-Iodosancycline, Amicycline, Apicycline, Bromotetracycline (bromotetracycline), Chlorotetracycline (chlorotetracycline), Clomocycline, Demeclocycline, Demecycline, DMG-DMDOT (DMG-DM DOT), DMG-MINO, Doxycycline, Eravacycline, Etamocycline, Glycocycline, Guanecycline, Lymecycline, Meclocycline, Meglucycline, Metacycline (methacycline), Minoycline, Morphocycline, Nitrocycline, Omadacycline, Oxytetracycline, Pecocycline, Penimpeiclycine, Penimocycline, Pipacycline, Rolitetracycline, Sancycline, Sarocycline, Tetracycline, Tigecycline, TP 271
Table 3. The bioactive molecules with primary target of large ribosomal subunit inhibition predicted by the minimum structure.

| Minimum structure | Bioactive molecules |
|-------------------|---------------------|
| 4-Aminopyrimidin-2(1H)-one or Cytosine | Amicetin (allomycin), Antelmycin (anthelmycin), Arginomycin, Bagougeramine A, Bagougeramine B, Bamicetin, Blastidin H, Blastidin S, Cytimidine, Cytomycin (saitomycin), Cytosamine, Cytosaminomycin A, Cytosaminomycin B, Cytosaminomycin C, Cytosaminomycin D, Gougerotin, Mildiomycin, Mildiomycin B, Mildiomycin C, Mildiomycin D, Mildiomycin M, Norplicacetin, Oxamicetin, Oxyplicacetin (cytosaminomycin E), Plicacetin (amicetin B), Rodaplutin, SCH 36605, SF 2457 |
| ![Image of minimum structure](image1) | ![Image of active molecules](image2) |
| Comment: The active ingredient contains at least one of the following components without a phosphorus group including 5-amino-5,6-dihydro-2H-pyran-2-yl (a), 5-aminotetrahydro-2H-pyran-2-yl (b) or 4-formamidobenzoyl (c) | |
| ![Image of 2,6-Dioxo-4-piperidinyl or 3-Glutaramidyl](image3) | 9-Methylstreptimidone (S 632A3), Acetoxycycloheximide (streptovitacin E-73), Actiketal, Actiphenol (actinophenol), Cycloheximide, Epiderstatin, Inactone, Isoactocyheximide, Isomigrastatin, Lactimidomycin, Naramycin B, Neoisocycloheximide, S 632A3, Streptimidone (S 632A1), Streptovitacin A, Streptovitacin B, Streptovitacin C3 |
| ![Image of 2,6-Dioxo-4-piperidinyl or 3-Glutaramidyl](image4) | ![Image of active molecules](image5) |
| ![Image of (1R)-Propan-1-ol or (1R)-Propanol](image6) | Azidamfenicol, Bromamphenicol (bromoamphenicol), Cetofenicol (cetophenicol), Chloramphenicol, Florfenicol, Flornfenicol, Monooiodoamphenicol, Racefenicol (racephenicol), Tevenel, Thiamphenicol, WIN 5094-2 |
| ![Image of (1R)-Propan-1-ol or (1R)-Propanol](image7) | ![Image of active molecules](image8) |
| ![Image of (1R)-Propan-1-ol or (1R)-Propanol](image9) | ![Image of active molecules](image10) |
| ![Image of (1R)-Propan-1-ol or (1R)-Propanol](image11) | ![Image of active molecules](image12) |

Comments:
- \( X_1 = X_3 = X_5 = H \)
- \( X_2 \neq H \); straight-chain longer than 12-membered chain; components without -OH or -CO
- Active ingredient = if there is a macrocyclic lactone ring, this ring will have at least one of the following components as part of the ring including (1R)-1-methyl-3-hydroxybutan-1-yl formate (a), (1R)-1-methylbut-3-en-1-yl formate (b), (1R)-1-methyl-3-aminobutan-1-yl formate (c) or (1R)-1-methyl-3-oxobutan-1-yl formate (d)
- \( (1R) \)-Propan-1-ol ≠ cyclic bonds
- \( X_1 = \) aromatic ring
- \( X_2 = -\text{NCO} - \text{NSO}_2 \)
- \( X_3 = \text{O}, \text{F}, \text{Cl}, \text{Br}, \text{I} \)
Table 4. The bioactive molecules with primary target of sterol 14α-demethylase inhibition predicted by the minimum structure.

| Minimum structure | Bioactive molecules |
|-------------------|---------------------|
| 1H-Imidazol-1-yl  | 1-Dodecylimidazole (N-dodecylimidazole), AFK 108, Aliconazole, Asartaconazole, Azalanstat, BAY C-9263, BAY D-9603, Beciconazole, Bifonazole, Brolaconazole, Butaconazole, Ciconazole, Climbazole, Clotrimazole, Croconazole, Democonazole, Dichlorophenyl imidazolidoxolan (eluobl), Doconazole, Eberconazole, Econazole, Fenapanil, Fenticonazole, Flutrimazole, Imazalil (enilconazole), Isoconazole, Ketaminaize, Ketoconazole, Lanoconazole, Lonconazole, Lorbonazole, Luliconazole, MH 0685, Miconazole, Neticonazole, OK 8705, OK 8801, Omoconazole, Orconazole, Oxiconazole, Oxipoconazole, Parconazole, Pefurazoate, PR 967-234, Prochloraz, R 31000, Sertaconazole, SM 4470, SSF 105, Sulconazole, Ticonazole, Triflumizole, UK 38667, Valiconazole, Zinoconazole, Zoficonazole |
| or Imidazol-1-yl  |                                                                  |
|                   |                                                                  |
| N                  |                                                                  |
| X\(_1\) = hydrophobic group (directly or indirectly) |
| X\(_1\) ≠ methylbenzonitrile (a); methylphenylacetonitrile (b) |
| X\(_2\) = X\(_3\) = X\(_4\) = H |
| Active ingredient = if there is a fused aromatic ring, the molecule will have an aromatic ring with a halogen in position 2 or 4 |
| Active ingredient ≠ carbanilate (c); an imidazol-1-yl at one end of the molecule and a carboxylate (-COO) at the other end; single (S)-enantiomer in the molecule with one atom stereocenter |
| a                  |                                                                  |
| b                  |                                                                  |
| c                  |                                                                  |

| 1H-1,2,4-Triazol-1-yl | 1,2,4-Triazol-1-yl |
|----------------------|-------------------|
| or 1,2,4-Triazol-1-yl |                                                                  |
| X\(_5\) = hydrophobic group (directly or indirectly) |
| X\(_5\) ≠ methylbenzonitrile (a); methylphenylacetonitrile (b) |
| X\(_6\) = X\(_7\) = H |
| Active ingredient = if there is a fused aromatic ring, the molecule will have an aromatic ring with a halogen in position 2 or 4 |
| Active ingredient ≠ carbanilate (c); single (S)-enantiomer in the molecule with one atom stereocenter |
| a                  |                                                                  |
| b                  |                                                                  |
| c                  |                                                                  |

Comments:
X\(_1\) = hydrophobic group (directly or indirectly)
X\(_1\) ≠ methylbenzonitrile (a); methylphenylacetonitrile (b)
X\(_2\) = X\(_3\) = X\(_4\) = H
Active ingredient = if there is a fused aromatic ring, the molecule will have an aromatic ring with a halogen in position 2 or 4
Active ingredient ≠ carbanilate (c); an imidazol-1-yl at one end of the molecule and a carboxylate (-COO) at the other end; single (S)-enantiomer in the molecule with one atom stereocenter
**Comments:**

- $X_5 \neq \text{H}$ (weak inhibitory activity)
- $X_4 = -\text{COO}$; isothiazol-3(2H)-one (a) fused to the core
- $X_5$ = aromatic ring

Active ingredient: non-fused aromatic ring or chain attached to an aromatic ring at $X_4$ (here $X_4$ means six-membered aromatic ring fused to the core at $X_4$) in position 5; the core fused to more than two (if $X_5 = -\text{COO}$) or three (if $X_5 = \text{isothiazol-3(2H)-one fused to the core}$) aromatic rings.
**Supplementary Figure 1.** An example of identifying a bioactive molecule with the primary target of DNA gyrase and topoisomerase IV inhibition by the minimum structure in 4-pyridone group. Molecule 1 contains five aromatic rings fused to the core (4-pyridone), molecule 2 contains one methyl at X₈, molecule 4 doesn’t contain -COO or isothiazol-3(2H)-one fused to the core at X₃, molecule 5 contains hydrogen at X₁ and molecule 6 contains one methyl attached to the aromatic ring (phenyl) at X₈ (here X₈ means six-membered aromatic ring fused to the core at X₅) in position 6. Therefore, the primary target of molecules 1, 2, 4, 5 and 6 is not DNA gyrase and topoisomerase IV inhibition.
2,4(or S)-Diaminocyclohexanol

Comment:
The active ingredient comprises a ring attached to the core directly or indirectly.
**Supplementary Figure 2.** An example of identifying a bioactive molecule with the primary target of small ribosomal subunit inhibition by the minimum structure in 2,4(or 5)-diaminocyclohexanol group. Molecules 1, 3, 4, 5 and 6 don’t contain 2,4(or 5)-diaminocyclohexanol (the core). Therefore, the primary target of molecules 1, 3, 4, 5 and 6 is not small ribosomal subunit inhibition.
Supplementary Figure 3. An example of identifying a bioactive molecule with the primary target of small ribosomal subunit inhibition by the minimum structure in (4aRS,5aRS)-Sancycline group. Molecules 2, 3, 4, 5 and 6 don’t contain (4aRS,5aRS)-Sancycline (the core). Therefore, the primary target of molecules 2, 3, 4, 5 and 6 is not small ribosomal subunit inhibition.
Comment:
The active ingredient contains at least one of the following components without a phosphorus group including 5-amino-5,6-dihydro-2H-pyran-2-yl (a), 5-aminotetrahydro-2H-pyran-2-yl (b) or 4-formamidobenzoyle (c)

\[
\begin{align*}
\text{Comment:} \\
\text{The active ingredient contains at least one of the following} \\
\text{components without a phosphorus group including 5-amino-5,6-} \\
\text{dihydro-2H-pyran-2-yl (a), 5-aminotetrahydro-2H-pyran-2-yl (b) or} \\
4\text{-formamidobenzoyle (c)}}
\end{align*}
\]
**Supplementary Figure 4.** An example of identifying a bioactive molecule with the primary target of large ribosomal subunit inhibition by the minimum structure in cytosine group. Molecules 3 and 5 don’t contain cytosine (the core) and molecules 1, 4 and 6 don’t contain at least one of the following components including 5-amino-5,6-dihydro-2H-pyran-2-yl, 5-aminotetrahydro-2H-pyran-2-yl or 4-formamidobenzoyl. Therefore, the primary target of molecules 1, 3, 4, 5 and 6 is not large ribosomal subunit inhibition.
Large ribosomal subunit inhibition

3-Glutarylidy

Connections:
X₅ = X₃ = X₆ = H
X₂ ≠ H, straight-chain longer than 12-membered chain; components without -OH or -CO
Active ingredient: if there is a macrocyclic lactone ring, this ring will have at least one of the following components as part of the ring including (1R)-1-methyl-3-hydroxybutan-1-yl formate (a), (R)-1-methylbutan-3-en-1-yl formate (b), (1R)-1-methyl-3-aminoheptan-1-yl formate (c) or (1R)-1-methyl-3-exobutan-1-yl formate (d)
**Supplementary Figure 5.** An example of identifying a bioactive molecule with the primary target of large ribosomal subunit inhibition by the minimum structure in 3-glutarimidyl group. Molecules 2, 3, 4 and 5 don’t contain 3-glutarimidyl (the core) and molecule 1 contains components other than hydrogen at X₁, X₃ and X₅. Furthermore, molecule 1 contains components without -OH or -CO at X₄. Therefore, the primary target of molecules 1, 2, 3, 4 and 5 is not large ribosomal subunit inhibition.
Large ribosomal subunit inhibition

(LR)-Propanol

Comments:
(LR)-Propanol ≠ cyclic bonds
X₃ = aromatic ring
X₂ = -NCO₂ -NSO₂
X₁ = O, F, Cl, Br, I
**Supplementary Figure 6.** An example of identifying a bioactive molecule with the primary target of large ribosomal subunit inhibition by the minimum structure in (1R)-propanol group. Molecules 1, 2, 3 and 6 don’t contain (1R)-propanol (the core) and molecule 5 doesn’t contain -NCO or -NSO₃ at X₂. Therefore, the primary target of molecules 1, 2, 3, 5 and 6 is not large ribosomal subunit inhibition.
Sterol 14α-demethylase inhibition

Imidazol-1-yl

Comments:
X_1 = hydrophobic group (directly or indirectly)
X_2 ≠ methyl/bromonitrile (a), methyl/phenylisocyanate (b)
X_2 = X_3 = H
Active ingredient: if there is a fixed aromatic ring, the molecule will have an aromatic ring with a halogen in position 2 or 4
Active ingredient: α-carboxylate (c): an imidazol-1-yl at one end of the molecule and a carboxylate (-COO) at the other end; single (S)-enantiomer at the molecule with one atom stereocenter
**Supplementary Figure 7.** An example of identifying a bioactive molecule with the primary target of sterol 14α-demethylase inhibition by the minimum structure in imidazol-1-yl group. Molecules 1 and 2 contain a fused aromatic ring but these molecules don’t have an aromatic ring with a halogen in position 2 or 4, molecule 3 has a carbanilate, molecule 4 doesn’t contain a hydrophobic group (directly or indirectly) at $X_1$ and molecule 6 contains an imidazol-1-yl at one end of the molecule and a carboxylate (-COO) at the other end. Therefore, the primary target of molecules 1, 2, 3, 4 and 6 is not sterol 14α-demethylase inhibition.
Sterol 14α-demethylase inhibition
1,2,4-Triazol-1-yl

Comments:

- $X_i$: hydrophobic group (directly or indirectly)
- $X_i \neq$ methylbenzonitrile (a); methylphenylacetanilide (b)
- $X_i = X_i' = H$

Active ingredient – if there is a fused aromatic ring, the molecule will have an aromatic ring with a halogen in position 2 or 4
Active ingredient ≠ carbonilale (c); single (S)-enantiomer in the molecule with one atom stereocenter
Supplementary Figure 8. An example of identifying a bioactive molecule with the primary target of sterol 14α-demethylase inhibition by the minimum structure in 1,2,4-triazol-1-yl group. Molecule 1 contains components other than hydrogen at X1, molecule 3 has a methylbenzonitrile at X1, molecule 4 contains a fused aromatic ring but this molecule doesn’t have an aromatic ring with a halogen in position 2 or 4, molecule 5 contains one atom stereocenter and single (S)-enantiomer and molecules 1 and 6 don’t contain a hydrophobic group (directly or indirectly) at X1. Therefore, the primary target of molecules 1, 3, 4, 5 and 6 is not sterol 14α-demethylase inhibition.
Cyclopropylcyclobutane

Active ingredient ≠ AR, cyclopentadiene
Supplementary Figure 9. The process for predicting the primary target from chemical structure based on the minimum structure in a hypothetical example. Molecule 1 is used as the query molecule. Seven structurally related molecules to molecule 1 are found by similarity searching. Since the primary target of the query molecule (molecule 1) is unknown, information on the primary target, structure-activity relationship and pharmacophore are collected for the eight structurally related molecules. Then we consider the following assumptions:

Molecules 3 and 4 have the primary target of Y. Molecule 3 consists of a square attached to a pentagon with two double bonds, and molecule 4 consists of a triangle attached to a pentagon with two double bonds. Based on information on the structure-activity relationship of molecule 3, the double bonds on the pentagon with two double bonds (cyclopentadiene) must not be reduced. Therefore, the core of these two molecules is the common and important part, named the pentagon with two double bonds (cyclopentadiene). The core of molecules 3 and 4 in the presence of square and triangle is attributed to the pentagon with two double bonds. As a result, the pentagon with two double bonds has priority over the square and the triangle in expressing the primary target of the molecule.

Molecules 5 and 7 have the primary target of Z. Molecule 5 consists of a square attached to a six-membered aromatic ring (benzene) and a pentagon with two double bonds (cyclopentadiene), and molecule 7 consists of a triangle attached to a five-membered aromatic ring (pyrrole) and a square. Based on information on the structure-activity relationship of molecule 7, the aromatic ring plays an important role in the activity. Therefore, the core of these two molecules is the common and important part, named the aromatic ring (the aromatic ring in molecules 5 and 7 is characterized by LigandScout with AR). There is the pentagon with two double bonds in molecules 3, 4 and 5, but the core of molecule 5 is attributed to the aromatic ring. As a result, the aromatic ring has priority over the pentagon with two double bonds (cyclopentadiene) in expressing the primary target of the molecule.

Molecules 2 and 6 have the primary target of X. Molecule 2 consists of a triangle attached to a square, and molecule 6 consists of a triangle attached to two squares. Because there is no molecule with the primary target of X having the square or the triangle alone, so the core of these two molecules is the common part, named the triangle attached to the square. The triangle attached to the square is also present in molecule 7, but the core of this molecule is attributed to the aromatic ring. As a result, the aromatic ring has priority over the triangle attached to the square in expressing the primary target of the molecule.

Molecules 1 and 8 have an unknown target. Molecule 1 consists of a square attached to a triangle and a pentagon, and molecule 8 consists of a square attached to two triangles. There is also a triangle attached to a square (or a
square attached to a triangle) in molecules 2, 6 and 7. In addition, the results showed that the aromatic ring and possibly the pentagon with two double bonds (cyclopentadiene) have priority over the triangle attached to the square in expressing the primary target of the molecule, but molecules 1 and 8 don’t contain any the aromatic ring or the pentagon with two double bonds. Therefore, the core of molecules 1 and 8 consists of the triangle attached to the square (cyclopropylcyclobutane) and the peripheral part does not consist the aromatic ring and the pentagon with two double bonds (cyclopentadiene). As a result, the primary target of molecules 1 and 8 is identical to the primary target of their neighbor molecules, named molecules 2 and 6, and the primary target of the two molecules is predicted X.

Abbreviations: AR, aromatic ring; H, hydrophobic interaction; HBD, hydrogen bond donor.