Cheminformatics analysis of molecular datasets of transcription factors associated with quorum sensing in *Pseudomonas aeruginosa*

Felipe Victoria-Muñoz[1,2], Norberto Sánchez-Cruz[3], José L. Medina-Franco[3], Fabian López-Vallejo[2,*]

[1] Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Farmacia, Av. Cra 30 # 45-03, Bogotá D.C.,11001 Colombia.

[2] Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Química, Grupo de Investigación en Productos Naturales Vegetales Bioactivos, Av. Cra 30 # 45-03, Bogotá D.C.,11001 Colombia.

[3] DIFACQUIM Research group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City, 04510 Mexico

* Corresponding author: FL-V. E-mail: fhlopezv@unal.edu.co

Abstract

Transcription factors associated with quorum sensing in *P. aeruginosa* are promising targets for developing new adjuvants against infection by this pathogen. Regulation of these transcription factors offers the possibility of controlling multiple virulence factors related to them, as the development of biofilm, proteases, hydrogen cyanide, and others. Numerous molecules have been tested against those targets, but the keys responsible for antagonistic activity against these targets are unknown. In this work, we analyze the structure-activity relationships of active molecules tested against LasR, PqsR, and RhlR transcription factors to establish the structural characteristics associated. The compounds were classified as agonist, antagonist, and inactive. As part of the analysis, we conducted molecular complexity, scaffold analyses, activity cliffs analyses, and chemical space visualization to find characteristics associated with the biological activity. Several structural features were identified as important to antagonist activity in this study, highlighting molecular size and hydrogen bond acceptors.

Keywords: Quorum sensing · Pseudomonas aeruginosa · Cheminformatics · LasR/PqsR/RhlR · Activity-cliffs · Structure-activity relationships

Abbreviations:

- DS: Datasets.
- QS: Quorum sensing.
- TF's: Transcription factors.
- AI's: Autoinducers.
- SAR: Structure-activity relationships
- 3OC12-HSL: N-3-oxo-dodecanoyl homoserine lactone.
- C4-HSL: Butyryl homoserine lactone.
- PQS: 2-heptyl-3-hydroxy-4-quinolone.
- IQS: 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde.
- EC50: Half maximal effective concentration.
- IC50: Half maximal effective concentration.
- Fsp3: Fraction of sp3 atoms.
- FCC: Fraction of chiral centers.
- SD: Standard deviation.
- MACCS: Molecular access system keys.
- ECFP: Extended connectivity fingerprints.
- MCS: Maximum common substructure.
- MOA: Mechanism of action.
- t-SNE: T-distributed stochastic neighbor embedding.

Graphical Abstract
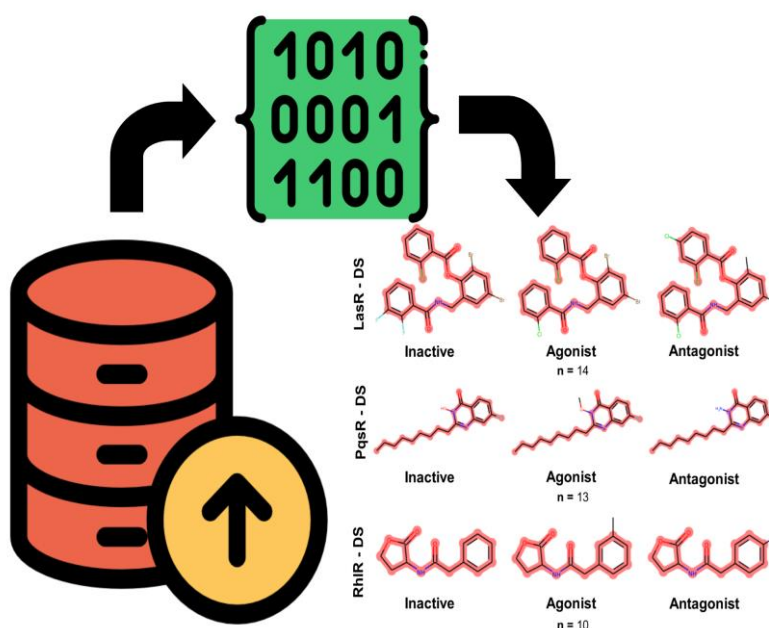
Introduction

Infection by *P. aeruginosa* is a challenge in therapeutic management. These bacteria can adapt to a hostile environment, resist the immune system, and block the action of broad-spectrum antibiotics such as third-generation cephalosporin, carbapenems, and others [1, 2]. For that reason, the World Health Organization has classified *P. aeruginosa* with a critical priority for the development of new drugs [3].

One of the critical mechanisms related to the augmented antibiotic resistance of *P. aeruginosa* is Quorum sensing (QS), which is a communication system where each bacteria can detect an existent population in a given environment [4]. Small molecules called autoinducers (AI's) regulate the QS through their binding to specific receptors. They activate the genetic machinery to produce multiple virulence factors [5].

Nowadays, there are known twenty transcription factors (TF's) involved in QS; all of them having multiple interactions with each other [6]. *Las*, *Rhl*, PQS, and IQS are related to the activation of QS and the production of multiple virulence factors (elastase, protease, rhamnolipids, pyocyanin, hydrogen cyanide) [7, 8]; *Las* and *Rhl* are systems of the Lux type, while PQS is a quinolone dependent system.

Las and *Rhl* systems are important for QS activation. These systems have an inductor protein (LasI and RhlI) and a receptor protein (LasR and RhlR); inductors synthesize AI's and receptors recognize them [9]. Lux type systems work with acyl-homoserine lactone molecules like AI's. These molecules are composed of a lactonic ring and a carbonate chain of variable size connected by an amide. N-3-oxo-dodecanoyl homoserine lactone (3OC12-HSL) is the autoinducer of the *Las* system, while butyryl homoserine lactone (C4-HSL) is an autoinducer of the *Rhl* system [4, 10] (Figure 1).

2-heptyl-3-hydroxy-4-quinolone (PQS) is the associated autoinducer to PQS system. PQS is recognized by PqsR and synthesized by multiple proteins, as PqsB, PqsC, PqsD, and others [12, 13]. 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde (IQS) is an autoinducer for IQS system. Previous studies established the activation of QS by IQS in deficiency of *Las* system, generating interest for its study [7, 14]. LasR, RhlR, and PqsR AI's are shown in Figure 1.
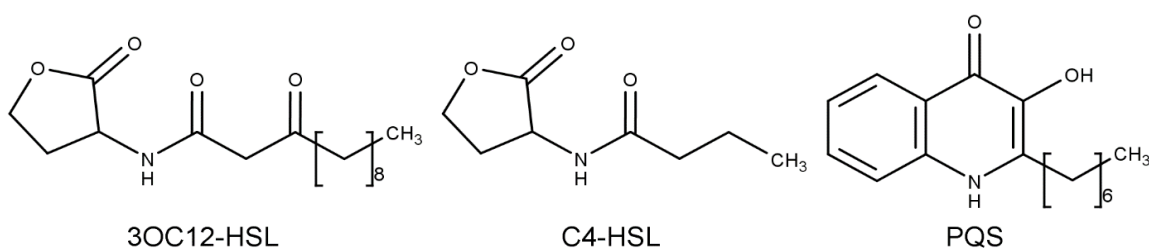
3OC12-HSL   C4-HSL   PQS

Figure 1. Molecular structures of autoinducers associated with quorum sensing in *P. aeruginosa*. N-3-oxo-dodecanoyl homoserine lactone (3OC12-HSL), butyryl homoserine lactone (C4-HSL) and 2-heptyl-3-hydroxy-4-quinolone (PQS) [4].

Find new antagonists for these TF's gives a new pathway for the *P. aeruginosa* infection management, but the structure-activity relationship of these targets doesn't know at the moment. For that reason, in this work, we search for characteristics or structural patterns associated with agonist activity, antagonist activity, or inactivity. To achieve this goal, we built a dataset of 310 molecules with reported biological activity in major public repositories from LasR, PqsR, and RhlR and performed cheminformatics studies. Here we propose some relevant keys to contribute to solving this problem.

Methods

Construction and preparation of compound datasets

Compound datasets (DS) were built with ChEMBL, version 26, PubChem, and scientific articles. For ChEMBL, the following identifiers were used: LasR - CHEMBL1075207, PqsR - CHEMBL2424500, and RhlR - CHEMBL3112386. In PubChem Bioassay, the search was done directly with the AI's. Additionally, for the previous search, only molecules IC_{50} and EC_{50} values were taken into account. Keywords' LasR', 'PqsR', 'RhlR', 'Quorum Sensing' were used to search for molecules in scientific articles in SciFinder. Then all papers were filtered by their biological activity values 'IC_{50}' or 'EC_{50}', taking into account all articles published between1998 and 2020. This search was carried out between January 2020 and July 2020.

All molecules in DS were classified as "active" and "inactive" using a cutoff value of 10 µM for LasR and PqsR, while 12 µM was used for RhlR. These cutoff values were chosen based on the mean minus one standard deviation of biological activity. Then, the following steps were applied to prepare the molecules: 1) the largest fragment in each molecule was selected; 2) non-organic molecules were removed; 3) charges

were neutralized by adding and/or removing hydrogens where possible; 4) it was ensured that the most potent acid groups were ionized first in partially ionized molecules; 5) stereochemistry was removed; 6) the most populated tautomer was selected; 7) duplicate molecules were discarded. This curation was done with RDKit [23] and MolVS [24] library.

Molecular complexity

Molecular complexity has been associated with target selectivity and success in clinical stages [25–27], here we inspect the relation between active compounds and molecular complexity. For this study, the fractions of sp^3 atoms (Fsp^3) and chiral centers (FCC) were used as complexity metrics based on topological and physicochemical descriptors [28]. Also, Complexity by Bertz [29] and DataWarrior [30] based on graph-theoretical methods were used. All metrics were calculated with RDKit [23] and DataWarrior 5.0.0 [30].

Scaffolds analysis

Scaffolds can be defined as the core of a molecule used in drug design [32–34]. For this reason, scaffolds are intuitive and informative to explore SAR and Structure-Multiple Activity Relationships (SmART) [35] of compounds data sets, identifying them supports drug design. This work explores the relationship between scaffolds and the biological activity for LasR, PqsR, and RhlR datasets. For scaffold generation, we use the definition proposed by Bemis and Murcko [36] as implemented in RDKit [23], in which a scaffold is defined as the core structure of the compound consisting of all of its rings and connecting linkers. According to their origin molecule, scaffolds were classified into three groups: agonist, antagonist, and inactive according to their origin molecule, in all DS. A maximum common substructure analysis was done on the most frequent scaffold in each DS [37, 38]. The data were collected using RDKit [23].

Activity cliffs analysis

Activity cliffs are small changes in the structure associated with unexpected large changes in biological activity [39–41]. Identification of activity cliffs is a relevant and informative analysis of the SAR because they point to specific and minor structural modifications that strongly influence biological activity. One must be

aware, however, that there might be "artificial" cliffs in compound data sets, for instance, the biological activity is not measured correctly [42]. In this work, to identify the possible presence of activity cliffs in the DS, we calculated consensus activity cliffs between multiple 2D fingerprints (MACCS Keys 166-bits, ECFP4, ECFP3, ECFP2, and path length), founding the activity cliff common in all fingerprints through of Tanimoto coefficient as is proposed by Medina-Franco and coworkers [43]. RDKit [23] and MayaChemTools [44] were employed for this analysis.

Chemical space visualization

The representation of a set of molecules in a two- or three-dimensional plot based on a set of pre-defined descriptors is usually known as visualization of the chemical space. In this work, the statistical analysis called t-distributed stochastic neighbor embedding (t-SNE) was done to represent chemical space occupied by datasets, with a perplexity equal to 10, a learning rate equal to 100, using to Jaccard metric and descriptor-based in fingerprint ECFP4. Also, constellation plots recently developed [45] were built for the datasets; this plot integrates a methodology of analogs series with the visualization of the chemical space, associate the molecular analogs to several molecules and their biological activity.

Results and discussion

Table 1 summarizes the number of compounds found in ChEMBL and PubChem Bioassays. In ChEMBL can be found molecules with activity against LasR, PqsR, and RhlR. LasR has 173 molecules reported, but only 82 have values of IC_{50} and EC_{50}, 29 molecules are reported for PqsR, of which 8 have values of EC_{50}, while RhlR possesses 31 compounds with activity being have 4 with values of IC_{50}. In comparison, PubChem Bioassay has 87, 10, and 17 compounds with activity for LasR, PqsR, and RhlR, respectively. The rest of the molecules were found in scientific articles. More than 70 percent of molecules are shared with ChEMBL. LasR is TF with a higher number of molecules. Furthermore, the LasR dataset has the highest proportion of active molecules between all datasets.

	Total	Agonist		Antagonist		Inactive	
		n	%	n	%	n	%
LasR	188	52	27,7	41	21,8	95	50,5
PqsR	54	10	18,5	15	27,8	29	53,7
RhlR	47	12	25,6	7	14,9	28	59,5

Table 1. TF's Datasets

Molecular complexity

Target selectivity is a major issue in drug discovery. It has been proposed the association between greater molecular complexity and less target promiscuity. Here we found an association between DataWarrior complexity and biological activity, showed in Figure 2.

The majority of molecules active against PqsR have the highest DataWarrior complexity and some of them show pIC50 upper than 7. In contrast, RhlR compounds are concentrated in the lower left of figure 2, where molecules with low activity and poor complexity are showed. Finally, LasR active compounds could not be associated with specific complexity values because this dataset has a high molecular diversity as are evidenced in posterior analysis. Antagonist molecules in all datasets don't have a clear behavior associated with complexity in metrics studied.
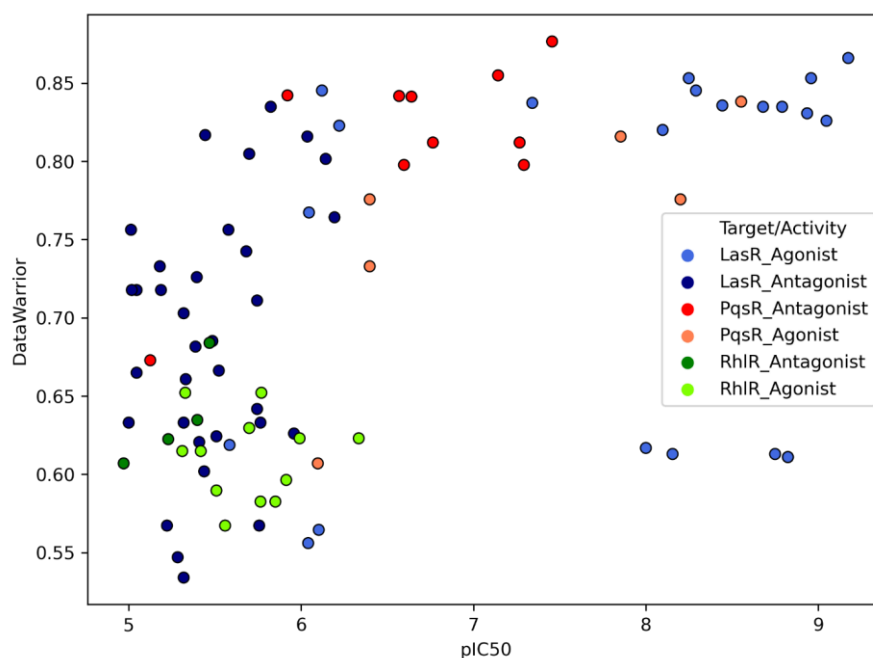

Figure 2 Distribution of biological activity values against DataWarrior complexity. In blue, red and green are represented molecules with activity against LasR, PqsR and RhlR respectively while light colors represent agonist compounds and dark colors represent antagonist compounds.
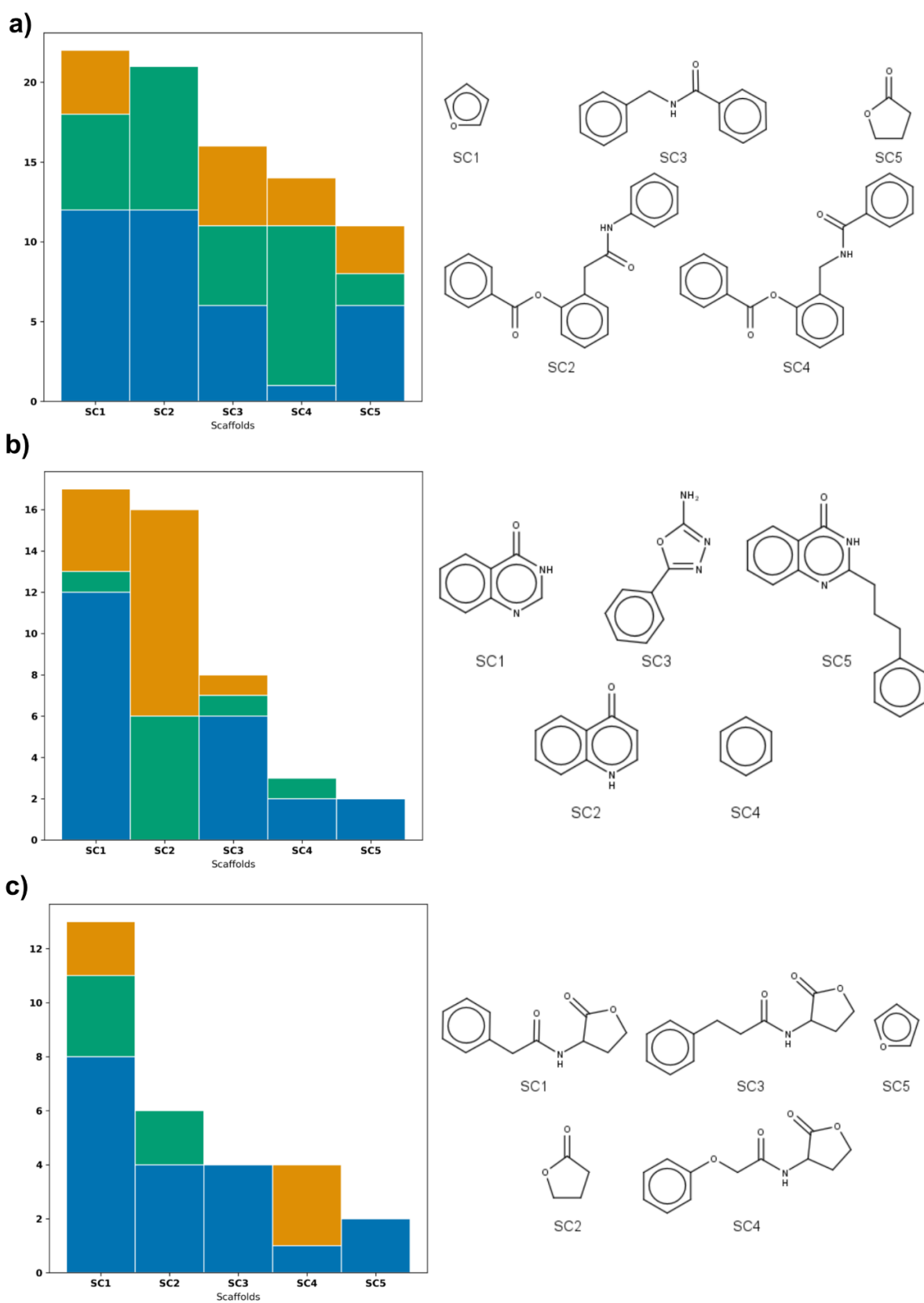
Figure 3 First five Bemis - Murcko Scaffold frequency in datasets, a) represent LasR dataset, b) PqsR dataset and c) RhlR dataset; activity type is showed by colors, inactivity (blue), agonist (green), and antagonist molecules (yellow).

Scaffold analysis

Figure 3 shows scaffolds frequency in DS on the left side, while the scaffolds structure is shown on the right. The frequency of scaffolds found for DS varies according to DS size (Figure 3). Nevertheless, the first five most frequent scaffolds represent around seventy percent of the total, highlighting a variety of types of biological activity in that scaffolds. In LasR scaffolds, it is found that five first have molecules with activity, being the most frequently agonist activity, which is shown in figure 3 a); rings of six members connected between them through a small chain with amide, are the most relevant for this DS.

Three phenyl system is present in the second and fourth most populated scaffolds in figure 3a but the position of amide in chain allows antagonistic activity for the fourth scaffold. Meanwhile, the first and fifth scaffolds are based in the lactonic ring common in autoinducer and present all of the activities; being inadequate for similarity searching, but the rest of them would be useful.
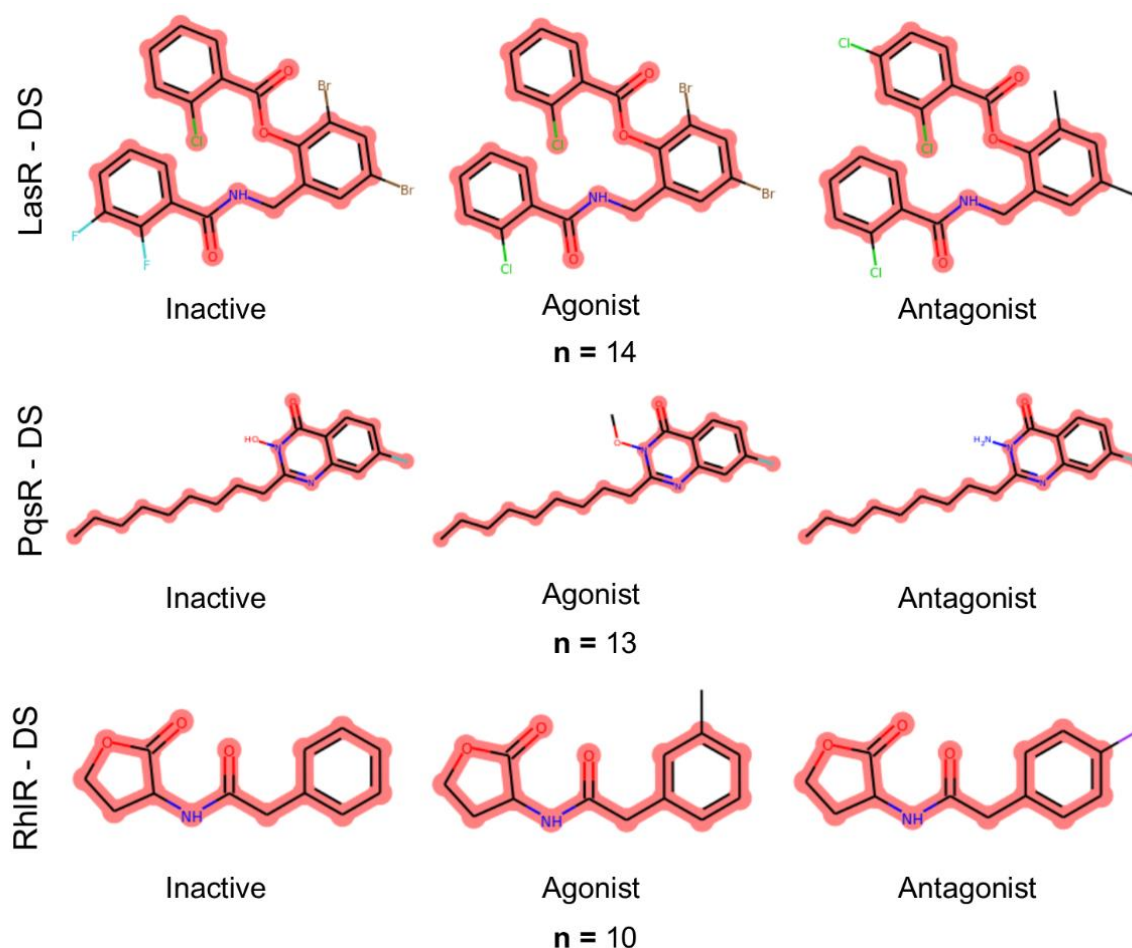

Figure 4 Representative examples of maximum common substructures. n represents the number of molecules with this scaffold.

Scaffolds of PqsR DS are composed of two rings fused of six members, characteristics of natural autoinducer. All the molecules with quinoline ring (SC2) have an activity less than 10μM (see figure 3 b)), making it a suitable candidate for the similarity searching of new molecules with biological activity; another characteristic of this scaffolds is that the first two represent more than fifty percent of molecules, exposing a few diversity in this DS. The biological activity of the first five scaffolds found in RhlR DS is lower than the activity of scaffolds in the rest of DS, here all scaffolds have a lactonic ring, but the scaffold (SC4) with the largest percent of antagonist molecules, also presents a chain with an amide, an ether, and a phenyl.

Analysis of the maximum common substructure (MCS) allows finding a small structural difference between molecules with different types of activity. All the molecules shown in Figure 4 for LasR DS were reported by O'Reilly and collaborators [19], they have a triphenyl scaffold showing the activity related to a specific ring substitution, the inactive molecule is ortho and meta disubstituted by fluorine in the first ring, the agonist and antagonist are monosubstituted in the ortho position by the chlorine in the first ring, but the antagonist is also monosubstituted in the para position in the third ring. These substitutions can change the inductive effect generating different interactions in the active site [46].

Ilangovan et al. reported the compounds associated with biological activity for PqsR DS in 2013 [47]. These compounds differ from each other only in the substitution of nitrogen in the quinazoline ring: the inactive compound has a hydroxyl, the agonist has ethyl, and the antagonist has an amine; finding the association of biological activity related to several acceptors and donors of hydrogen bonds in this position. For molecules of this analysis linked to RhlR DS, it is possible to establish a steric effect in the active site with the type of activity, as is showed in Figure 4. This analysis allows establishing for these scaffolds a mechanism of action cliffs (MOA-cliffs) [48].
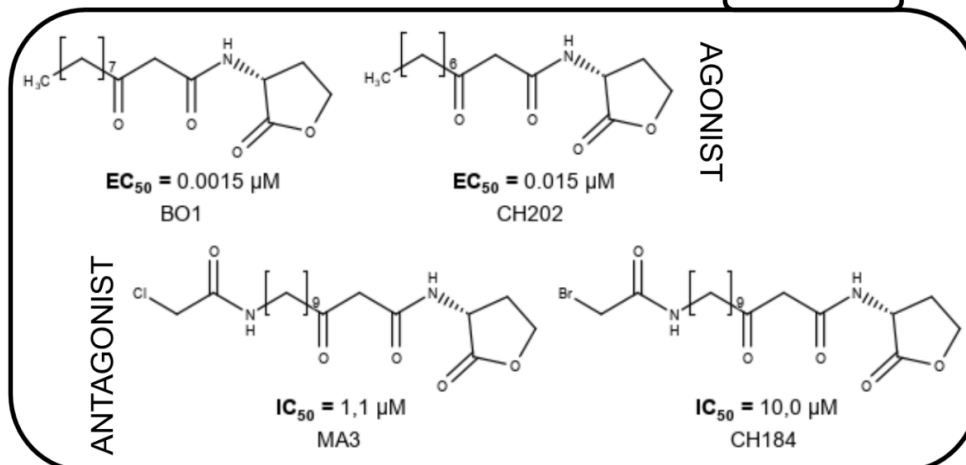
Activity cliffs

Figure 5 shows the structure of representative activity cliffs in DS, for all TF's, it is evident the relation of agonist and antagonist molecules, supported by previous analysis. Regarding LasR agonist, it is possible to see a change of EC_{50} value by a variation in one-carbon of a lateral chain, BO1 has eleven carbons, while CH202 has ten carbons, but the potency decreases in ten orders of magnitude; furthermore, natural autoinducer possesses a lateral chain with twelve carbons; all of these allow us to understand the
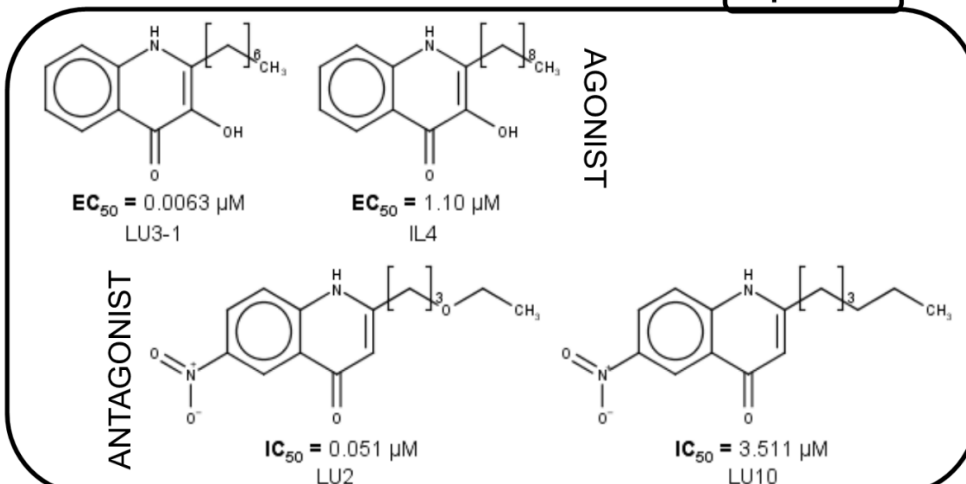
importance of size for the agonist activity. The molecular pair found for LasR antagonist, MA3 and CH184, differ between them in the final halogen, MA3 has chlorine while CH184 has bromine; but antagonist activity for CH184 is a ten percent respect to MA3; relating the size of the atoms in a chain with the activity; a smaller atomic size generates more antagonist activity.

LU3-1 and IL4 are the molecular pair found for PqsR agonist, LU3-1 has a chain with seven carbons while IL4 has a chain with nine carbons, but agonist activity is around one hundred and seventy times greater for LU3-1, finding relationships between the chain size and the agonist activity. In another perspective, compounds with code LU2 and LU10 were activity cliffs related to antagonist activity, these have a lateral chain with six carbons, but LU2 has oxygen in position four, while for LU10 all are carbons; this variation generates a loss of potency in sixteen orders of magnitude; this loss should be explained by the lack of a hydrogen bond acceptor.

LasR - DS

AGONIST

EC$_{50}$ = 0.0015 μM
BO1

EC$_{50}$ = 0.015 μM
CH202

ANTAGONIST

IC$_{50}$ = 1,1 μM
MA3

IC$_{50}$ = 10,0 μM
CH184

PqsR - DS

AGONIST

EC$_{50}$ = 0.0063 μM
LU3-1

EC$_{50}$ = 1.10 μM
IL4

ANTAGONIST

IC$_{50}$ = 0.051 μM
LU2

IC$_{50}$ = 3.511 μM
LU10

RhlR - DS

AGONIST

EC$_{50}$ = 0.463 μM
BO2-4

EC$_{50}$ = 3.820 μM
BO11

ANTAGONIST

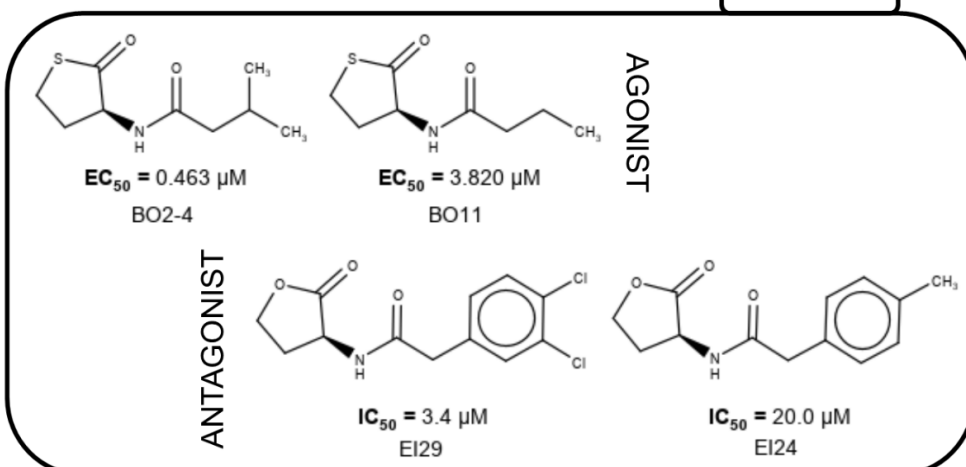IC$_{50}$ = 3.4 μM
EI29

IC$_{50}$ = 20.0 μM
EI24

Figure 5 Activity cliffs found in datasets, here the agonist has shown at left and the antagonist at the right.

12

Activity cliffs for RhlR agonist are molecules with a thiolactone ring and a lateral chain, here the chain change in methyl group in position four for BO2-4 while BO11 does not have it, activity for BO2-4 is eight times greater than BO11; furthermore, the natural autoinducer of RhlR has an EC_{50} (8,95μM) lower than these compounds, and its difference is a lactonic ring versus thiolactone ring, as reported by Boursier and coworkers [49]. The substitution of the phenyl ring in molecular pairs found for antagonist activity generates a change of this activity, while EI29 is disubstituted by two chlorines, EI24 is monosubstituted by a methyl group, linking the size and inductive effects with this activity [18].

Chemical space

Through visualization of the distance between molecules using structural parameters as a method of classification, give a full vision of molecular datasets. For this reason, we represent a chemical space of DS relating to a type of biological activity generating a global view of molecules.

Chemical space visualization occupied by compounds in DS, based on ECFP4 fingerprints (being this chosen because have a better resolution than others), show agonist and antagonist molecules to be close in space, due to that type of compounds are similar between them, according to the results previously shown [48]. In the lower right corner of the plot, molecules with five-membered rings are shown, usually lactonic and thiolactone rings. In this region of the chemical space, can be found compounds of LasR and RhlR DS, agonist and antagonist of LasR are closer to each other than molecules of RhlR since the spatial location is given by the structure.
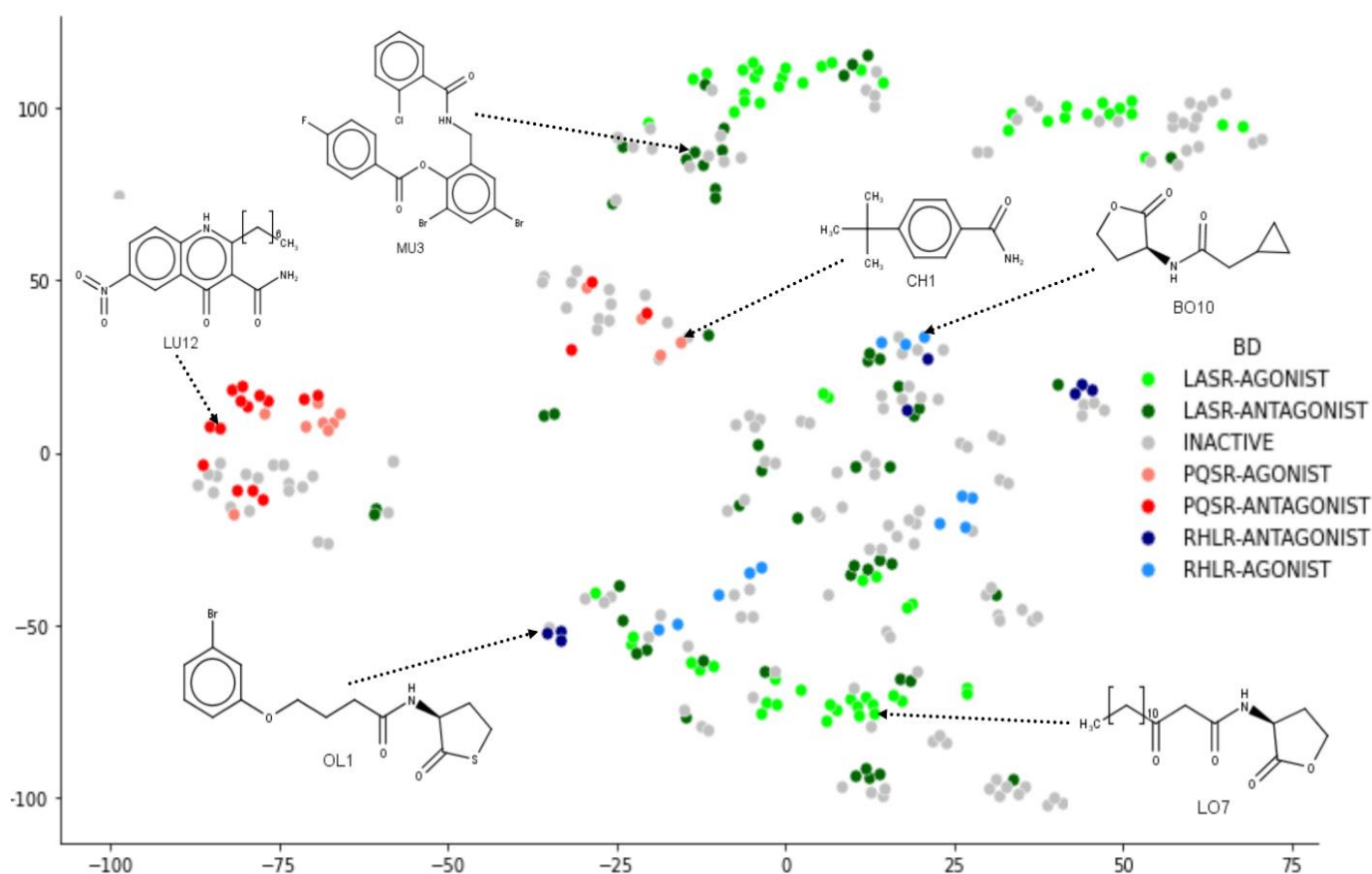
Figure 6 Structural space of databases. Visual representation of the chemical space obtained by t-distributed stochastic neighbor embedding (t-SNE). All molecules are classified as activity, inactive (gray), LasR agonist (light green), LasR antagonist (dark green), PqsR agonist (pink), PqsR antagonist (red), RhlR agonist (light blue), RhlR antagonist (dark blue).

In the upper right corner of Figure 6, molecules with triphenyl scaffold are found, characteristic of LasR DS. Here to found activity cliffs, nevertheless, a group of antagonist molecules is displayed separated from the leading group, being these a critical starting point for optimization of antagonist activity or similarity searching in big databases. Compounds of PqsR DS are displayed in the center of the plot, quinolinic and quinazolines rings are usual scaffolds by this activity.

In the visualization of the chemical space in Figure 6, all DS close in the space, especially LasR and RhlR DS. This is an indication that it is possible to find a molecule with antagonist activity against both TF's. But it is important to see the close relationship between agonist and antagonist in all DS, hence the challenge to find molecules with a specific activity on these TF's.

Constellation plots

A Constellation plot is a hybrid approach of visualization of chemical space and substructure analysis. As described in detail elsewhere, the main analog series of a compound data set are generated. The structural similarity of the analogs series is represented in a two-dimensional representation of chemical space such as t-SNE. Constellation plots represent a method to visualize and analyze StAR**s** (**St**ructure-**A**ctivity **R**elationship**s**) in chemical space [50, 51].
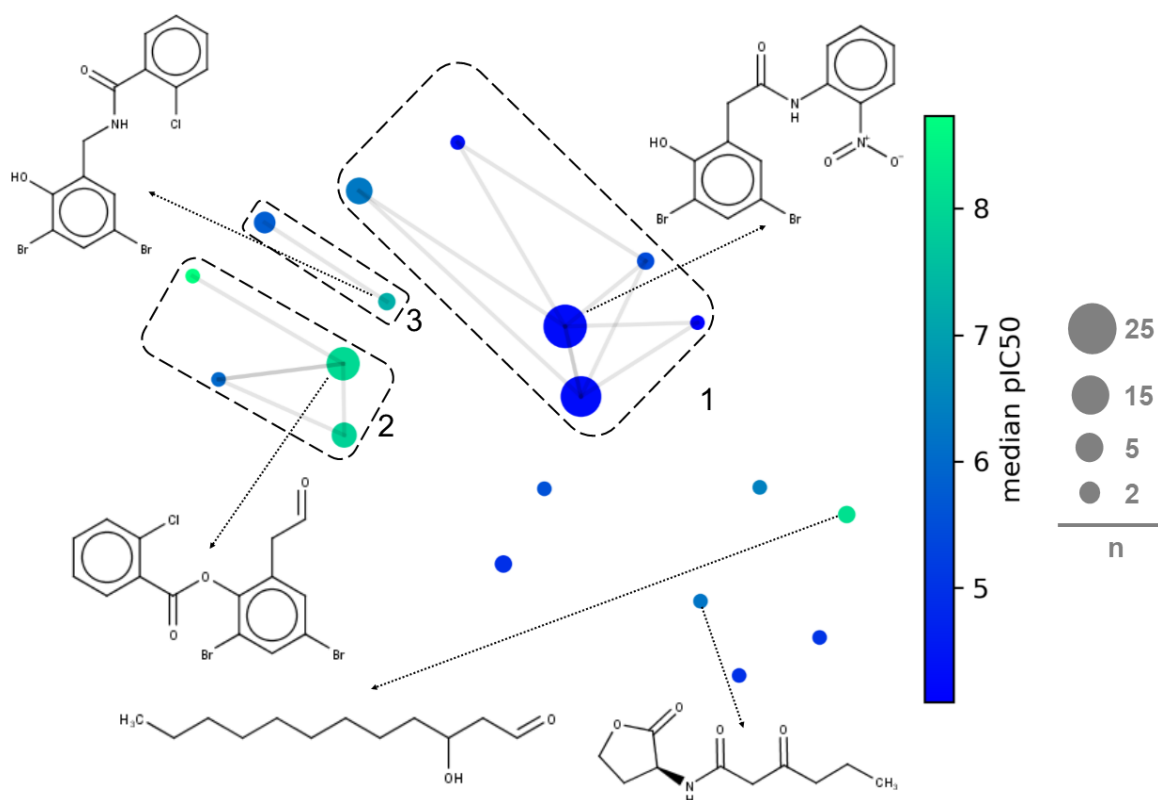

Figure 7 General constellation plot for a dataset of LasR. Here the most representative cores are shown. This plot has the same parameters of figure 5 (t-SNE, ECFP4, Tanimoto Matrix), each point represents a core, the size of circles indicates the "n" number of compounds annotated, connected circles are cores sharing compounds, grids represents analog series every point, and color shows the average of pIC$_{50}$.

Following the methodology proposed by Naveja and Medina-Franco, constellation plots were built [45]. These analyses gave us the relation of standard cores in the DS with biological activity and the number of compounds they have. Figure 7 shows the constellation plot for LasR DS, here are represented nineteen cores grouped in ten analog series they contained one hundred and six compounds (around half of the dataset).

The first three analog series display small structural changes in diphenyl rings, both the first and third series show diphenyl rings connected by a chain of three members with an amide group in the middle of them. The second series has diphenyl rings connected by a chain of two members with an ester group in the middle of them, that series has a better activity profile that cores an excellent base for searching other compounds or a useful scaffold for the optimization of biological activity.

Other essential cores for this activity are a carbonate chain of twelve members and a lactonic ring, the carbonate chain has an excellent mean activity for this target, but the lactonic ring does not; this is because the lactonic ring has multiple compounds with a harmful activity while carbonate chain has a significant number of compounds with functional activity. PqsR and RhlR DS could not be analyzed with this method because the rules of fragmentation used, does not involve carbonate chains or rings [43].

Conclusions

Although antagonist molecules in all DS don't was associated with molecular complexity, nevertheless, actives compounds have a relation with this analysis. PqsR actives molecules shown a high DataWarrior complexity, in RhlR DS all molecules were correlated with low DataWarrior Complexity, of contrast LasR actives compounds were associated with high and low values of complexity, making impossible a specific association with DataWarrior complexity.

For LasR antagonist compounds, it was identified that the side chain size should range between 10 and 12 members in homoserine lactone scaffolds. In contrast, in triphenyl scaffolds, the substitution of rings generates a specific type of activity. In the molecules active against PqsR the size of the side chain in quinazoline rings and substitution in the nitrogen of the third position derive in a determinate type of activity. Active homoserine lactone scaffolds for RhlR show that phenyl or ethyl substitution in the side chain gives a type of activity.

Analysis of maximum common substructure and visualization of the chemical space indicated that there is no scaffold associated with an exclusive type of activity in all TFs. Still, they are a good starting point since most of them are active. The LasR constellation plot allowed selecting a diphenyl scaffold as one of the best scaffolds to optimize because more than 25 molecules with biological activity on this target contain it.

Nevertheless, it is important to highlight the ambiguity of these characteristics because, in some cases, it should generate a change in the mode of action, as described in MOA-cliffs, making it difficult to optimize these compounds. To establish the structural keys with a type of activity, it is recommended to extend several molecules tested against TF's and to use molecular docking and dynamics for a deeper understanding at the molecular level.

Compliance with ethical standards

The authors declare that there is no financial or commercial conflict of interest.

References

1. Peña C, Suarez C, Tubau F, et al (2009) Carbapenem-resistant Pseudomonas aeruginosa: Factors influencing multidrug-resistant acquisition in non-critically ill patients. Eur J Clin Microbiol Infect Dis 28:519–522. https://doi.org/10.1007/s10096-008-0645-9

2. Potron A, Poirel L, Nordmann P (2015) Emerging broad-spectrum resistance in Pseudomonas aeruginosa and Acinetobacter baumannii: Mechanisms and epidemiology. Int J Antimicrob Agents 45:568–585. https://doi.org/10.1016/j.ijantimicag.2015.03.001

3. Tacconelli E, Carrara E, Savoldi A, et al (2018) Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis 18:318–327. https://doi.org/10.1016/S1473-3099(17)30753-3

4. Kalia VC (2015) Quorum Sensing vs Quorum Quenching: A Battle with No End in Sight. Springer India, New Delhi

5. March Rosselló GA, Eiros Bouza JM (2013) Quorum sensing en bacterias y levaduras. Med. Clin. (Barc). 141:353–357

6. Huang H, Shao X, Xie Y, et al (2019) An integrated genomic regulatory network of virulence-related transcriptional factors in Pseudomonas aeruginosa. Nat Commun 10:. https://doi.org/10.1038/s41467-019-10778-w

7. Lee J, Zhang L (2015) The hierarchy quorum sensing network in Pseudomonas aeruginosa. Protein Cell 6:26–41. https://doi.org/10.1007/s13238-014-0100-x

8. Schuster M, Peter Greenberg E (2006) A network of networks: Quorum-sensing gene regulation in Pseudomonas aeruginosa. Int J Med Microbiol 296:73–81. https://doi.org/10.1016/J.IJMM.2006.01.036

9. Kostylev M, Kim DY, Smalley NE, et al (2019) Evolution of the Pseudomonas aeruginosa quorum-sensing hierarchy . Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1819796116

10. Miller MB, Bassler BL (2001) Quorum Sensing in Bacteria. Annu Rev Microbiol 55:165–199. https://doi.org/10.1146/annurev.micro.55.1.165

11. Pesci EC, Milbank JBJ, Pearson JP, et al (1999) Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 96:11229–11234. https://doi.org/10.1073/pnas.96.20.11229

12. Gallagher LA, McKnight SL, Kuznetsova MS, et al (2002) Functions required for extracellular quinolone signaling by Pseudomonas aeruginosa. J Bacteriol 184:6472–6480. https://doi.org/10.1128/JB.184.23.6472-6480.2002

13. Déziel E, Lépine F, Milot S, et al (2004) Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. Proc Natl Acad Sci U S A 101:1339–1344. https://doi.org/10.1073/pnas.0307694100

14. Lee J, Wu J, Deng Y, et al (2013) A cell-cell communication signal integrates quorum sensing and stress response. Nat Chem Biol 9:339–343. https://doi.org/10.1038/nchembio.1225

15. Ahumedo Monterrosa M, Galindo JF, Vergara Lorduy J, et al (2019) The role of LasR active site amino acids in the interaction with the Acyl Homoserine Lactones (AHLs) analogues: A computational study. J

Mol Graph Model 86:113–124. https://doi.org/10.1016/j.jmgm.2018.10.014

16. Kamal AAM, Petrera L, Eberhard J, Hartmann RW (2017) Structure-functionality relationship and pharmacological profiles of: Pseudomonas aeruginosa alkylquinolone quorum sensing modulators. Org Biomol Chem 15:4620–4630. https://doi.org/10.1039/c7ob00263g

17. Boursier ME, Moore JD, Heitman KM, et al (2018) Structure–Function Analyses of the N -Butanoyl l -Homoserine Lactone Quorum-Sensing Signal Define Features Critical to Activity in RhlR. ACS Chem Biol 13:2655–2662. https://doi.org/10.1021/acschembio.8b00577

18. Eibergen NR, Moore JD, Mattmann ME, Blackwell HE (2015) Potent and Selective Modulation of the RhlR Quorum Sensing Receptor by Using Non-native Ligands: An Emerging Target for Virulence Control in *Pseudomonas aeruginosa*. ChemBioChem 16:2348–2356. https://doi.org/10.1002/cbic.201500357

19. O'Reilly MC, Dong S-H, Rossi FM, et al (2018) Structural and Biochemical Studies of Non-native Agonists of the LasR Quorum-Sensing Receptor Reveal an L3 Loop "Out" Conformation for LasR. Cell Chem Biol 25:1128-1139.e3. https://doi.org/10.1016/J.CHEMBIOL.2018.06.007

20. Soheili V, Tajani AS, Ghodsi R, Bazzaz BSF (2019) Anti-PqsR compounds as next-generation antibacterial agents against Pseudomonas aeruginosa: A review. Eur. J. Med. Chem. 26–35

21. Mendez D, Gaulton A, Bento AP, et al (2019) ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Res 47:D930–D940. https://doi.org/10.1093/nar/gky1075

22. Kim S, Chen J, Cheng T, et al (2019) PubChem 2019 update: Improved access to chemical data. Nucleic Acids Res 47:D1102–D1109. https://doi.org/10.1093/nar/gky1033

23. (2013) RDKit: Open-source cheminformatics. http://www.rdkit.org

24. (2016) MolVS: Molecule Validation and Standardization. https://molvs.readthedocs.io/en/latest/index.html

25. Lovering F, Bikker J, Humblet C (2009) Escape from flatland: Increasing saturation as an approach to improving clinical success. J Med Chem 52:6752–6756. https://doi.org/10.1021/jm901241e

26. Lovering F (2013) Escape from Flatland 2: Complexity and promiscuity. Medchemcomm 4:515–519. https://doi.org/10.1039/c2md20347b

27. Clemons PA, Bodycombe NE, Carrinski HA, et al (2010) Small molecules of different origins have

distinct distributions of structural complexity that correlate with protein-binding profiles. Proc Natl Acad Sci U S A 107:18787–18792. https://doi.org/10.1073/pnas.1012741107

28. Méndez-Lucio O, Medina-Franco JL (2017) The many roles of molecular complexity in drug discovery. Drug Discov. Today. 22: 120-126. https://doi.org/10.1016/j.drudis.2016.08.009

29. Bertz H. S (1981) The First General Index of Molecular Complexity. J Am Chem Soc 103:3599–3601

30. 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

31. McKinney W (2010) Data Structures for Statistical Computing in Python. PROC 9th PYTHON Sci CONF (SCIPY 2010) 51–56. https://doi.org/10.3828/ajfs.41.3.62

32. Hu Y, Stumpfe D, Bajorath J (2011) Lessons learned from molecular scaffold analysis. J Chem Inf Model 51:1742–1753. https://doi.org/10.1021/ci200179y

33. Stumpfe D, Dimova D, Bajorath J (2016) Computational method for the systematic identification of analog series and key compounds representing series and their biological activity profiles. J Med Chem 59:7667–7676. https://doi.org/10.1021/acs.jmedchem.6b00906

34. Zdrazil B, Guha R (2018) The Rise and Fall of a Scaffold: A Trend Analysis of Scaffolds in the Medicinal Chemistry Literature. J Med Chem 61:4688–4703. https://doi.org/10.1021/acs.jmedchem.7b00954

35. Saldívar-González FI, Naveja JJ, Palomino-Hernández O, Medina-Franco JL (2017) Getting SMARt in drug discovery: Chemoinformatics approaches for mining structure-multiple activity relationships. RSC Adv. 7:632–641. https://doi.org/10.1039/C6RA26230A

36. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893. https://doi.org/10.1021/jm9602928

37. Golbamaki A, Franchi AM, Gini G (2017) The Maximum Common Substructure (MCS) Search as a new tool for SAR and QSAR. In: Advances in QSAR Modeling. Challenges and Advances in Computational Chemistry and Physics, 24th ed. pp 149–165

38. Xu YJ, Johnson M (2002) Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. J Chem Inf Comput Sci 42:912–926. https://doi.org/10.1021/ci025535l

39. Maggiora GM (2006) On outliers and activity cliffs - Why QSAR often disappoints. J Chem Inf Model 46:1535. https://doi.org/10.1021/ci060117s

40. Cruz-Monteagudo M, Medina-Franco JL, Pérez-Castillo Y, et al (2014) Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? Drug Discov Today 19:1069–1080. https://doi.org/10.1016/j.drudis.2014.02.003

41. Hu Y, Bajorath J (2013) Activity profile relationships between structurally similar promiscuous compounds. Eur J Med Chem 69:393–398. https://doi.org/10.1016/j.ejmech.2013.08.044

42. Medina-Franco JL (2013) Activity cliffs: Facts or artifacts? Chem Biol Drug Des 81:553–556. https://doi.org/10.1111/cbdd.12115

43. González-Medina M, Prieto-Martínez FD, Owen JR, Medina-Franco JL (2016) Consensus Diversity Plots: a global diversity analysis of chemical libraries. J Cheminform 8:1–11. https://doi.org/10.1186/s13321-016-0176-9

44. Sud M (2016) MayaChemTools: An open source package for computational drug discovery. J Chem Inf Model 56:2292–2297. https://doi.org/10.1021/acs.jcim.6b00505

45. Naveja JJ, Medina-Franco JL (2019) Finding constellations in chemical space through core analysis. Front Chem 7:1–10. https://doi.org/10.3389/fchem.2019.00510

46. Wermuth CG (2011) The Practice of Medicinal Chemistry, 3rd Ed. 982

47. Ilangovan A, Fletcher M, Rampioni G, et al (2013) Structural Basis for Native Agonist and Synthetic Inhibitor Recognition by the Pseudomonas aeruginosa Quorum Sensing Regulator PqsR (MvfR). PLoS Pathog. https://doi.org/10.1371/journal.ppat.1003508

48. Hao M, Bryant SH, Wang Y (2016) Cheminformatics analysis of the AR agonist and antagonist datasets in PubChem. J Cheminform 8:37. https://doi.org/10.1186/s13321-016-0150-6

49. Boursier ME, Combs JB, Blackwell HE (2019) N -Acyl l -Homocysteine thiolactones are potent and stable synthetic modulators of the RhlR quorum sensing receptor in *Pseudomonas aeruginosa*. ACS Chem Biol 14:186–191. https://doi.org/10.1021/acschembio.8b01079

50. Medina-Franco JL, Naveja JJ, López-López E (2019) Reaching for the bright StARs in chemical space. Drug Discov Today 24:2162–2169. https://doi.org/10.1016/j.drudis.2019.09.013

51. Medina-Franco JL, Sánchez-Cruz N, López-López E, Díaz-Eufracio BI. (2021) Progress on open chemoinformatic tools for expanding and exploring the chemical space. J. Comput.-Aided Mol. Des. 2021, in press. https://doi.org/10.1007/s10822-021-00399-1