The Medicinal and Biological Chemistry (MBC) Library: an Efficient Source on New Hits[&]

Víctor Sebastián-Pérez,[1,§] Carlos Roca,[1,§] Mahendra Awale,[2,§] Jean-Louis Reymond,[2] Ana Martinez,[1] Carmen Gil[1] and Nuria E. Campillo[1,*]

[1]Centro de Investigaciones Biológicas (CIB, CSIC). Ramiro de Maeztu, 9, 28040-Madrid, Spain
[2]Department of Chemistry and Biochemistry. University of Bern, Freiestrasse, 3, 3012-Berne, Switzerland
Switzerland
[§]Theses authors have contributed equally to this work

[*]**Corresponding author**

Dr. Nuria E. Campillo (nuria.campillo@csic.es)

Centro de Investigaciones Biológicas (CSIC)

Ramiro de Maeztu 9, 28040 Madrid-Spain

Tel. +34 91 838 31 12

Abstract: Identification of new hits is one of the biggest challenges in drug discovery. Creating a library of well-characterized drug-like compounds is a key step in this process. Our group has developed an in-house chemical library called Medicinal and Biological Chemistry (MBC) library. This collection has been successfully used to start several medicinal chemistry programs and developed in an accumulation of more than thirty years of experience in drug design and discovery of new drugs for unmet diseases. It contains over 1,000 compounds, mainly heterocyclic scaffolds. In this work, an analysis of drug-like properties and comparative study with well-known libraries by using different computer software is here presented.

INTRODUCTION

The development of a new drug, from the lab to the market, is a complex process that can last 12–15 years and can cost in excess of $1 billion.[1] During this process there are a number of bottlenecks,[2] such as the ability to obtain drug-like compounds. Because of the length of time and the high cost involved in the drug-discovery process, there is a growing impact of chemical databases in the drug discovery industry.[3] During recent years, chemical databases have evolved from a simple repository of synthesized compounds to an important research tool for the discovery of new hits and lead compounds. This is due to the fact that the availability of samples for virtual or experimental screening is of great importance for the identification of new hits, and the quality of these samples is crucial for a successful drug discovery program.[4]

Traditionally, databases have contained only chemical structures. Such databases support the organization of information within a company or, in the most pragmatic cases, for virtual screening purposes.[5] However, in recent decades the amount of information held in databases has greatly increased, including information about structure, ADMET (absorption, distribution, metabolism, excretion and toxicity) characteristics and physicochemical parameters (such as logP values, polar surface area [PSA], and hydrogen bond acceptors and donors [HBA and HBD, respectively]). These data are very useful to use as filters in order to obtain structures with acceptable drug-like profiles.

It has been widely accepted that the physicochemical profiles of oral drugs can be evaluated using the Lipinski rule of five.[6] In this sense, multiple studies have identified a relationship between the physicochemical properties of compounds — such as size, polarity, or lipophilicity — and failure of drug candidates during the drug discovery process.[7] These studies have suggested benefits that come from controlling general physicochemical properties in terms of reducing the probability of attrition in drug candidates, and they are based on the analysis of datasets taking as references approved oral drugs and compounds in preclinical and clinical studies.[8]

Physicochemical properties are not only critical for understanding the ADME profile of compounds, but a correlation has also been found between these properties and drug promiscuity and *in vivo* toxicology results.[9-10] One of the most interesting findings of these studies, carried out by the pharmaceutical company Pfizer, led to the

observation of a link between logP and PSA values and toxicology in preclinical *in vivo* studies.[11] Furthermore, due to the importance of biological information about small molecules that can enable many types of drug discovery analyses and decision making, most chemical databases have begun to include biological assay results. This allows for relevant information to be collected and made easily accessible to biochemical researchers and for drug-discovery purposes.

Our Medicinal and Biological Chemistry (MBC) laboratory has developed an in-house chemical library, which at the time of this study contained 1,096 compounds with a standard chemical purity of at least 95% by HPLC (Figure 1). These are compounds synthesized and characterized by our group, and it is based on more than thirty years of medicinal chemistry research. Our compounds have been designed mainly as potential drugs for neurological and neurodegenerative diseases. The chemical library is available both electronically and physically on request.

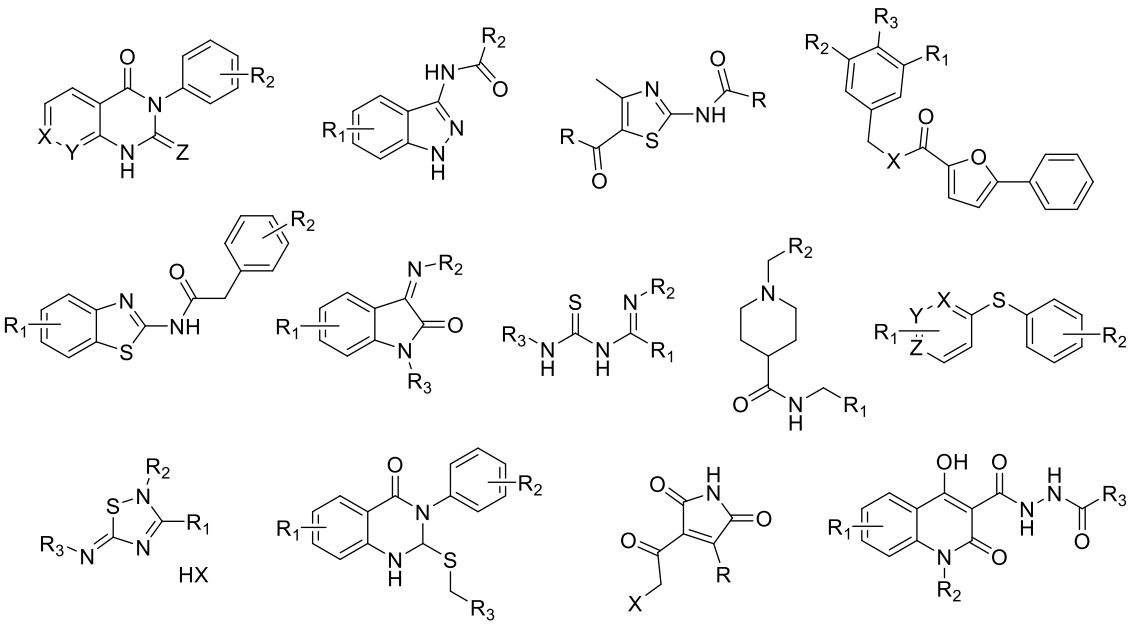

Figure 1. Representative set of the different families of MBC library

We are continuing to expand our chemical library by developing novel chemical series for different purposes that will target neglected or infectious diseases, among others. To illustrate the utility of the MBC library, we have identified, based both on different phenotypic and virtual screening campaigns, new targets and drug candidates for the treatment of Parkinson's disease[12-13] and new protein kinase[14] or phosphodiesterease[15-16] inhibitors with great value for neurological diseases. The main aim of this work is to characterize the MBC library, with particular attention to drug-

likeness and physicochemical properties, and to compare the MBC library with selected well-known chemical databases.

MATERIAL AND METHODS

Ligand preparation

All of the compounds from our in-house chemical library were exported to an excel format file (see Supporting Information) and were imported into Maestro 9.9[17] visualizer. The preparation of the library and the 2D-to-3D conversion was performed using the LigPrep[18] tool, a module of the Schrödinger software package. LigPrep allows different preparation steps of molecules such as the addition of hydrogen atoms, neutralization of charged groups, generation of ionization states, alternative chiral centers, low-energy ring conformations, options for generating multiple states and possible tautomers, followed by energy minimization using OPLS-2005 force field.[19-20] In order to carry out our studies, no possible ionizationed compounds were generated (as the compounds were in the most suitable ionization state for physiological pH conditions), all of the compounds were desalting, and no tautomers were generated. Moreover, one stereoisomer and one low energy ring conformation was generated per ligand. The last step was to minimize the compounds.

Ligand characterization

All of the prepared compounds were analyzed using the Qikprop[21] module of the small-molecule drug discovery suite in the Schrödinger software package. ADME properties were predicted using the QikProp program. QikProp was able to calculate and predict a total of 44 properties that helped to filter compounds with clear-cut, undesirable properties for drug discovery. Examples of such properties included molecular weight, molecular volume, number of HBD, number of HBA, PSA, QPlogPo/w (predicted octanol/water partition coefficient) and violations related to Lipinski's rule of five[22] and Jorgensen's rule of three.[23] Table 1 displays the parameters and corresponding ranges.

Table 1. Qikprop parameters and their corresponding ranges

Property of descriptor	Description	Range or recommended values
Mol_MW	Molecular weight of the molecule.	130–725
QPlogPo/w	Predicted octanol/water partition coefficient.	-2.0–6.5
#stars	Number of property or descriptor values that fall outside of the 95% range of similar values for known drugs. A large number of stars suggest that a molecule is less drug-like than molecules with few stars. The following properties and descriptors are included in the determination of the number of stars: MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, accptHB, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, QlogS, QPLogKhsa, QPlogBB, #metabol.	0–5
#rotor	Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds.	0–15
accptHB	Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.	2–20
donorHB	Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.	0–6
QPlogS	Predicted aqueous solubility, log S. S in mol/dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid	-6.5–0.5
Percentage HumanOral Absortion	Predicted human oral absorption. The prediction is based on a quantitative multiple linear regression model.	>80% is high and <25% is very poor
QPlogBB	Predicted brain/blood partition coefficient. Predictions are for drugs delivered orally.	-3.0 S in mol dm–31.2

Abbreviations: MW (molecular weight); IP (PM3 calculated ionization potential); EA (PM3 calculated electron affinity); SASA (solvent accessible surface area); FOSA (Hydrophobic component of the SASA); FISA (Hydrophilic component of the SASA); PISA (π component of the SASA); WPSA (Weakly polar component of the SASA); PSA (Van der Waals surface area of polar nitrogen and oxygen atoms); donorHB (dono hydrogen bond); accptHB (acceptor hydrogen bond); glob (Globularity descriptor); QPpolrz (Predicted polarizability); QPlogPC16 (Predicted hexadecane/gas partition coefficient); QPlogPoct (Predicted octanol/gas partition coefficient); QPlogPw (Predicted water/gas partition coefficient); QPlogPo/w (Predicted octanol/water partition coefficient); QlogS (Predicted aqueous solubility); QPLogKhsa (Prediction of binding to human serum albumin);QPlogBB (Predicted brain/blood partition coefficient); #metabol (Number of likely metabolic reactions).

Databases

DrugBank version 4.3, ChEMBL version 20, ZINC, and PubChem databases were downloaded from respective websites in 2016.

Processing of molecules

All cheminformatics calculations were performed using a bespoke java program developed based on the JChem chemistry library from ChemAxon Pvt. Ltd (https://www.chemaxon.com/). Molecules were processed as non-stereo simplified molecular-input line-entry systems (SMILES), counter ions were removed, valence errors were checked, and protonation states of molecules were adjusted at pH 7.4. All duplicate molecules were removed in the context of each database. The molecular properties such as heavy atom count (HAC), molecular weight (MW), HBD and HBA, octanol/water partition coefficient (logP), rotatable bond count (RBC), fraction of aromatic (fAromA) and sp3 carbons (fsp3) were calculated using various plugins — such as HBDAPlugin, TopologyAnalyserPlugin, and logPPlugin — from the JChem library. The number of unique and common molecules between different databases were calculated based on SMILES string comparisons.

Scaffold analysis

Bemis–Murcko scaffold (BMS) analysis was performed for MBC and other publicly available databases using a bespoke program that utilized the JChem chemistry library from ChemAxon Pvt. Ltd. For all molecules in each database, BMS scaffolds were calculated using the "StructuralFrameworksPlugin" from JChem and then stored in a unique SMILES format. Finally, the numbers of unique and common scaffolds between different databases were computed based on SMILES string comparisons.

RESULTS AND DISCUSSION

Characterization of MBC chemical library

As mentioned in the introduction, the successful development of new drugs critically depends on the ADME/tox properties of chemical compounds. These properties are crucial to be able to narrow the search for promising new chemical entities[24] in the early phases of drug discovery. By monitoring these properties during lead optimization, medicinal chemists may be able to reduce the exaggerated attrition rate in the drug discovery process.

With this concern in mind, our research group has been developing the MBC library. The main feature of this in-house chemical library is the common therapeutic profile of the compounds (that is, most of them are designed for the treatment of neurological and neurodegenerative diseases). Currently, 1,096 compounds are part of the database, which correspond to different alkyl and heterocyclic chemical families (Figure 1).

The information about the drug-like properties that are present in the MBC library was analyzed using the QikProp module[21] (Schrödinger Software Modules). After all of the compounds were prepared and characterized, we calculated the different physicochemical properties for further analyses (Table 2).

More than twenty relevant molecular descriptors were calculated by QikProp and were used to define the stars parameter, as detailed in Table 1. A large number of stars (that is, a number greater than five, as determined by the stars index) suggests that a molecule does not have the characteristics required of a desired drug; a value less than five on the stars index indicates a molecule that is similar to the vast majority of drugs used in a clinical setting. The values of the most relevant descriptors from the MBC library are shown in Table 2. Of the 1,096 compounds, only 12 molecules fall outside of the 95% range for values for known drugs, which represents only 1% of the compounds in our chemical library. In that sense we can affirm that, according to these predictions, 99% of the compounds that are present in our library have a physicochemical profile that is in line with known drugs that have been approved for human pharmacological treatments.

Table 2. Pharmacokinetic properties of MCB library analyzed by QikProp module.

LIPINSKI'S RULE OF FIVE	
0 violations	84.5% (927)*
1 violation	98.3% (1078)
JORGENSEN'S RULE OF THREE	
0 violations	76.7% (841)
1 violation	99.2% (1088)
Molecular Weight (Da)	
0/200	8.6% (94)
201/300	37.3% (409)
301/400	40.0% (438)
401/500	12.1% (133)
>500	2.0% (22)
Number of rotatable bonds	
0-5	82.5% (904)

Prediction of LogP	
≤5	77.6% (851)
>5	22.4% (245)
Prediction of LogS	
-12.0/-7.0	6.3% (69)
-6.9/-3.0	72.2% (791)
-2.9/2.0	21.5% (236)
Number of Heavy Atoms	
0/15	15.6% (171)
16/30	78.4% (859)
>30	6% (66)
Number of Aromatic rings	
0/1	11.8% (129)
2/3	79.4% (870)

6-10	16.5% (181)
>10	1.0% (12)
Number of donor Hydrogen bonds groups	
≤5	100% (1096)
>5	0% (0)
Number of acceptor Hydrogen bonds groups	
≤10	99% (1085)
>10	1% (11)
>3	8.8% (96)
Prediction of BBB pass	
-3.0/-1.0	16.4% (180)
-0.9/1.0	83.6% (916)
Prediction of Percent of Human Oral Absorption	
0%-50%	1.5% (17)
51%-75%	9.0% (98)
76%-100%	89.5% (981)

* Numbers into brackets point the nº of compounds.

When we analyzed Table 2 in detail, it showed that most of the molecules meet the criteria for each physicochemical parameter. Of the 1,096 compounds, 84.5% of the molecules have no violations of Lipinski's rule of five, and over 98% have less than two violations (that is, either zero violations or one violation); molecules with these ranges are considered to have a great profile for drug-likeness. In the same sense, for Jorgensen's rule of three, according to predictions, over 76% have no violations while more than 99% present with either zero violations or one violation.

Properties as size of the molecules, capacity to form hydrogen bond, lipophilicity or flexibility are important to establish a good drug-like profile. Figure 2 presents different 3D-plots that show the variability of the compounds in the MBC library in terms of physicochemical properties (on the x and y-axis) and the stars parameter index (colored). Analyses of the distribution of these variables as MW against lipophilicity (logP), RBC, or HBA and HBD has revealed that almost all of the compounds within the MCB library meet the criteria (Table 1) to be drug-like compounds.

Lipophilicity influences a number of physiological properties including transport through cell membranes, rate of metabolism, and interaction with receptor binding sites. Because of that, logP is a key parameter for the drug discovery process (Figure 2). The dispersion of logP against MW shows that logP values of most of the compounds are within the range of -1.0 to 6.0.

Molecular flexibility is an important property that is dependent on the RBC. This parameter contains information on a compound's conformational space. This implicit information is indirect and very limited, but it suggests that conformational

behavior matters not only in pharmacodynamics events — such as drug target recognition — but also from an ADME perspective. It has been reported that problems with high molecule flexibility cause a decrease in bioavailability[25] and the rate of transport across cell membranes, which limits the achievable binding affinity to the pharmacological target. The distribution of the RBC shows the degree of conformational flexibility. Most of the molecules show an RBC between two and eight. Similar dispersions are shown with the other properties, such as HBA (<10) and HBD (>5), which the values are into the expected range for drug-like compounds (Figure 2).

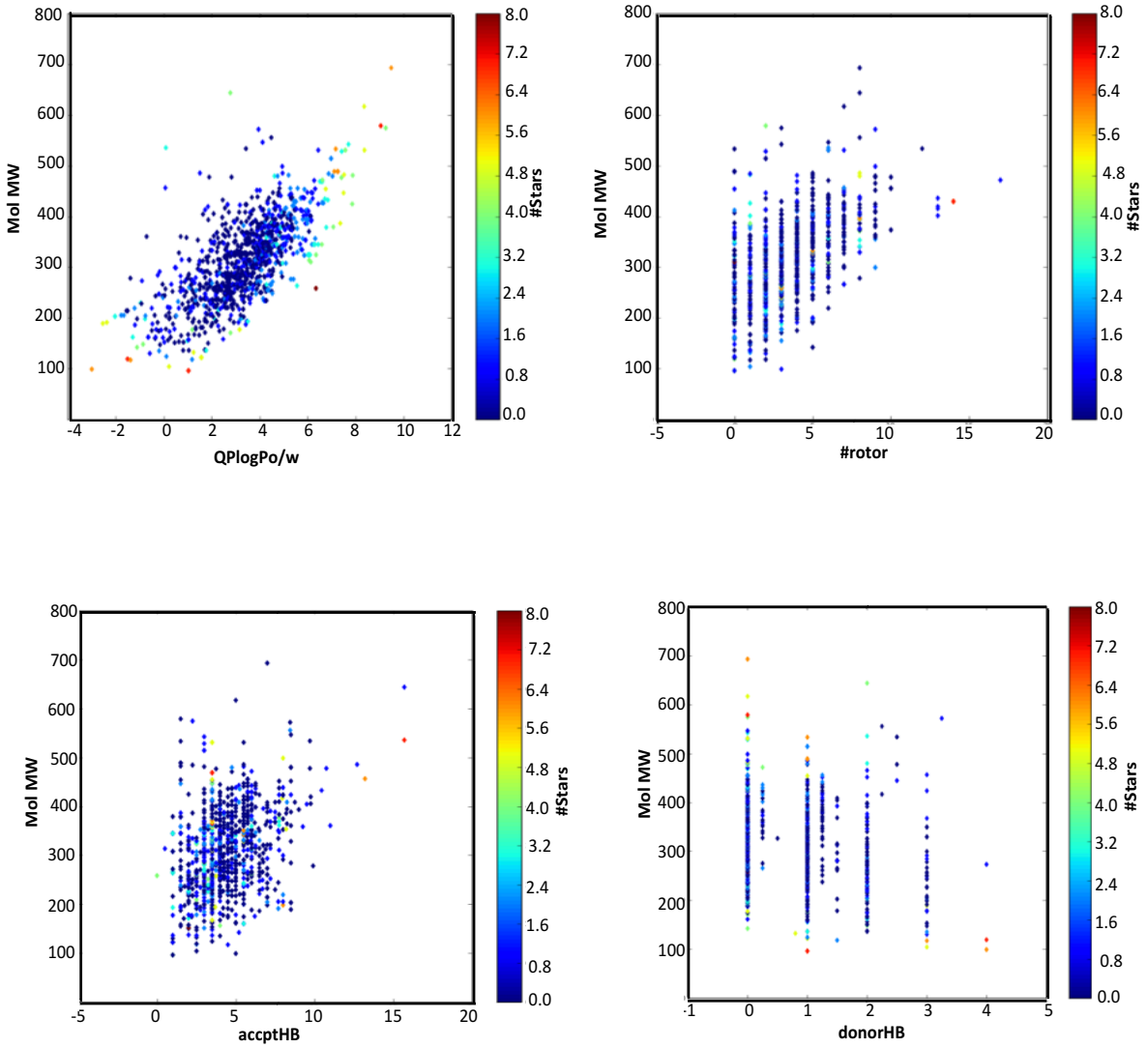

Figure 2. Dispersion of the MBC chemical compounds regarding MW and stars parameter (coloured) and log P (up, left), rotatable bonds (up, right) hydrogen bonds acceptors (down, left) and hydrogen bonds donors (down, right) respectively.

The solubility (logS) of a compound influences absorption and blood brain barrier permeation, among other properties. Figure 3 shows the distribution of MBC

chemical compounds with regard to logS. We used this information to analyze its influence on oral absorption (OA) and the brain–blood partition coefficient (logBB). When we analyzed these plots together with the information in Table 1, we concluded that 89% of the compounds in MBC library show more than 76% of absorption for oral drugs, and approximately 84% are within the recommended range for the predicted logBB (–0.9 to 1.0). Regarding the solubility of the compounds and their influence ~~on~~ in ADME properties, it can be seen that solubility values are in an acceptable range, which potentially translates to a good oral absorption and good logBB parameters.

The solubility (log S) of a compound influences on absorption and blood brain barrier permeation, among other properties. Figure 3 shows the distribution of MBC chemical compounds regarding log S in order to analyze its influence oral absorption (OA) and brain/blood partition coefficient (logBB). Analyzing these plots together with the Table 1 we can conclude that the 89% of the compounds in MBC library present over 76% of absorption for oral drugs and around 84% fall within the recommended range for the predicted logBB (–0.9 to 1.0). Regarding the solubility of the compounds and its influence in ADME properties, it can be seen that the solubility is in an acceptable range, would potentially translate into a good oral absorption and logBB parameters.

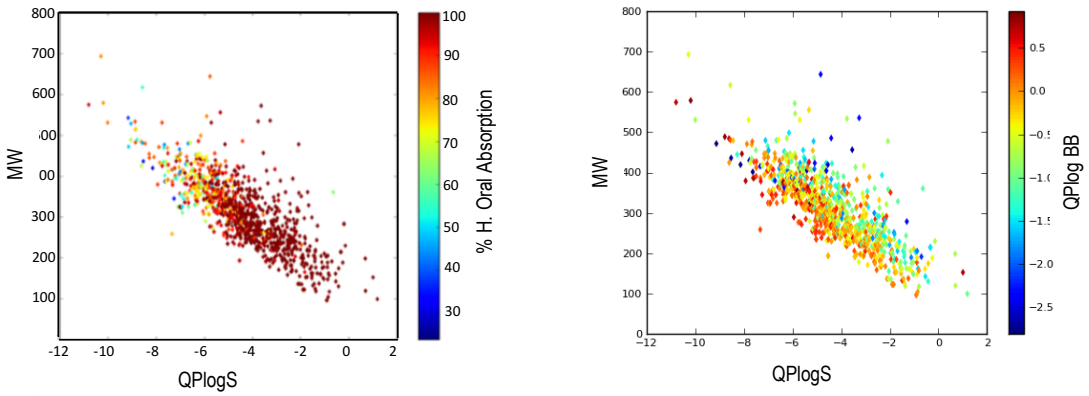

Figure 3. Dispersion of the MBC chemical compounds regarding solubility properties. Plot of MW (y-axis) and log S (x-axis) and oral absorption percentage is shown coloured (left). Plot of log S (y-axis) and logBB (x-axis)) and oral absorption percentage is shown coloured (right).

Comparative study

To complete the characterization of the MBC library, four different well-known chemical libraries that are often used in drug discovery processes were selected for comparison: ZINC, ChEMBL, PubChem, and the DrugBank. ZINC[26-27] is a freely available database of commercially available compounds. It was developed by the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF), and it contains a constantly growing number of 3D structures that are ready to dock. The new version, ZINC15, contains over 120 million purchasable "drug-like" compounds, of which around 30 million are available for immediate delivery. Each molecule also contains molecular properties (size, calculated logP values, hydrogen bond, or rotatable bonds) and purchasable information.

ChEMBL[28] is an open-source database developed by the European Bioinformatics Institute (EMBL-EBI) in Cambridge (UK). The data is manually collected from the literature and further standardized to optimize the data quality and utility across a wide range of chemical, biological, and drug discovery research problems. ChEMBL also contains structures and annotation from the U.S. Food and Drug Administration (FDA). Information about approved products (from the FDA Orange Book), including dosage information and administration routes, is included in the database. Also included are screening results and bioactivity data from other public databases such as PubChem Bioassay. This database contains binding, functional, and ADMET information for a large number of drug-like bioactive compounds. Currently, the database contains over 13.9 million bioactivity measurements for more than 1.9 million compounds and over 11,000 protein targets.

PubChem[29-30] is a database of chemical molecules and their activities in different biological assays. The system is maintained by the National Center for Biotechnology Information (NCBI), which is part of the U.S. National Institutes of Health (NIH). PubChem is comprised of three linked databases: PubChem Compound, PubChem Substance, and PubChem Bioassay. PubChem Compound contains the structure of over 82 million pure and characterized compounds and their molecular properties. PubChem Substance contains descriptions of chemical samples — for example, mixtures, extracts, complexes, and uncharacterized substances — and links to articles, protein 3D structures, and biological screening results available in PubChem BioAssay.

The DrugBank[31] is a freely available database that includes detailed drug information and data related to drug targets such as sequence, 3D structure, and

metabolic pathways. The database contains 8,206 drug entries, of which 1,991 are FDA-approved small-molecule drugs, 207 FDA-approved biotech (protein/peptide) drugs, 93 nutraceuticals, and over 6,000 experimental drugs. Additionally, 4,333 non-redundant protein sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields, with half of the information being devoted to drug and chemical data and the other half to drug-target or protein data**.**

Comparisons between our MBC library and the databases described above were carried out using a bespoke java program. The visualization of the properties was performed using molecular quantum numbers (MQN)-Mapplet software.[32] This software allowed us to visualize a 42-dimensional property space defined by 42 MQN integer value descriptors; these descriptors count different categories of atoms, bonds, polar groups, and topological features, and they categorize molecules by size, rigidity, and polarity. Figure 4 shows an example of these colors maps indicating the occupancy of the compounds in terms of properties.

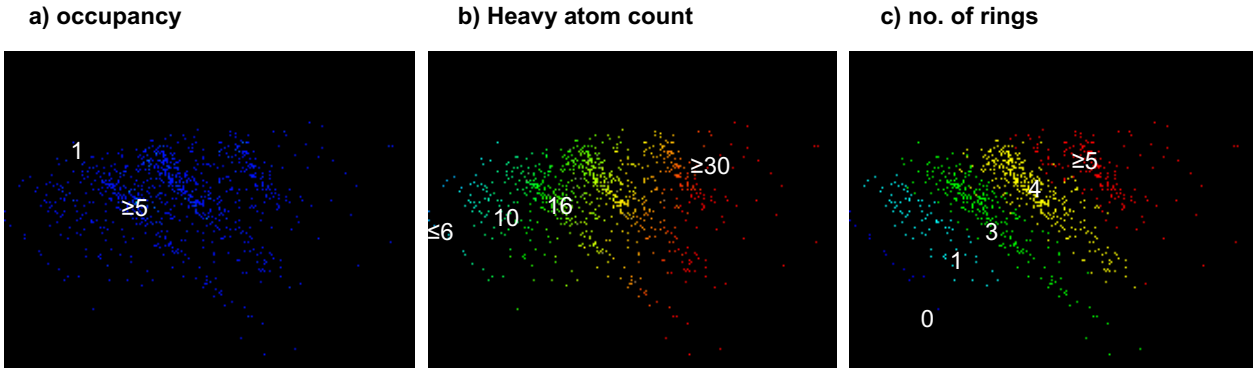

Figure 4. MQN PC1-PC2 maps of MBC library. Maps are color coded according to occupancy of compounds (a), heavy atom count (b) and no. of rings (c) in molecules. Color changes from blue to cyan to green to yellow to red with the increasing property values. PC1 and PC2 cover 62% and 19% of variance, respectively.

Figure 5 displays a comparison of different properties (heavy atom count, MW, HBSA and HBD, logP, RBC, fraction of aromatic atoms, and the fraction of sp3 carbons [sp3C]) between our library and the commercial ones.

According to the data obtained for MW, most molecules of the chemical libraries that were analyzed were in the range of Lipinski's rule of five (that is, less than 500 Da); in comparison, most marketed drugs have values of less than 500 Da. Moreover, the MBC database is the database that fits best with this parameter (Figure 5b).

In the case of HBD and HBA properties, these are also in agreement with Lipinski's rule of five and in the range of more than 95% of drugs in clinical use, taking into account values of 0.0–6.0 for HBD and 2.0–20.0 for HBA. The curve obtained for the MBC chemical library is very similar to previously published curves (Figure 5c and 5d). Regarding lipophilicity, the logP parameter remained in the range of -2.0 to 6, as was expected for more than 95% of known drugs. Also, for the MBC chemical library, the distribution of these values was almost identical when compared to the other databases (Figure 5e). Furthermore, some authors agree that a rotatable bond of less than seven in a molecule is the optimal value for this parameter in order to be considered a lead for possible drug development. The MBC database was found to be the best one when this parameter was analyzed (Figure 5f).

Finally, the fsp3C and the fraction of aromatic carbons (fArom) were analyzed and compared. These two parameters relate to two important physical properties: melting point and solubility. Molecules that are more highly complex, as measured by saturation, have the capacity to access greater chemical space. The 3D structure that is conferred by the saturation may also result in greater selectivity. Furthermore, saturation increased the likelihood of higher solubility and lower melting points, which are properties that are more likely to lead to drugs that are clinically successful (Figure 5g and 5h).

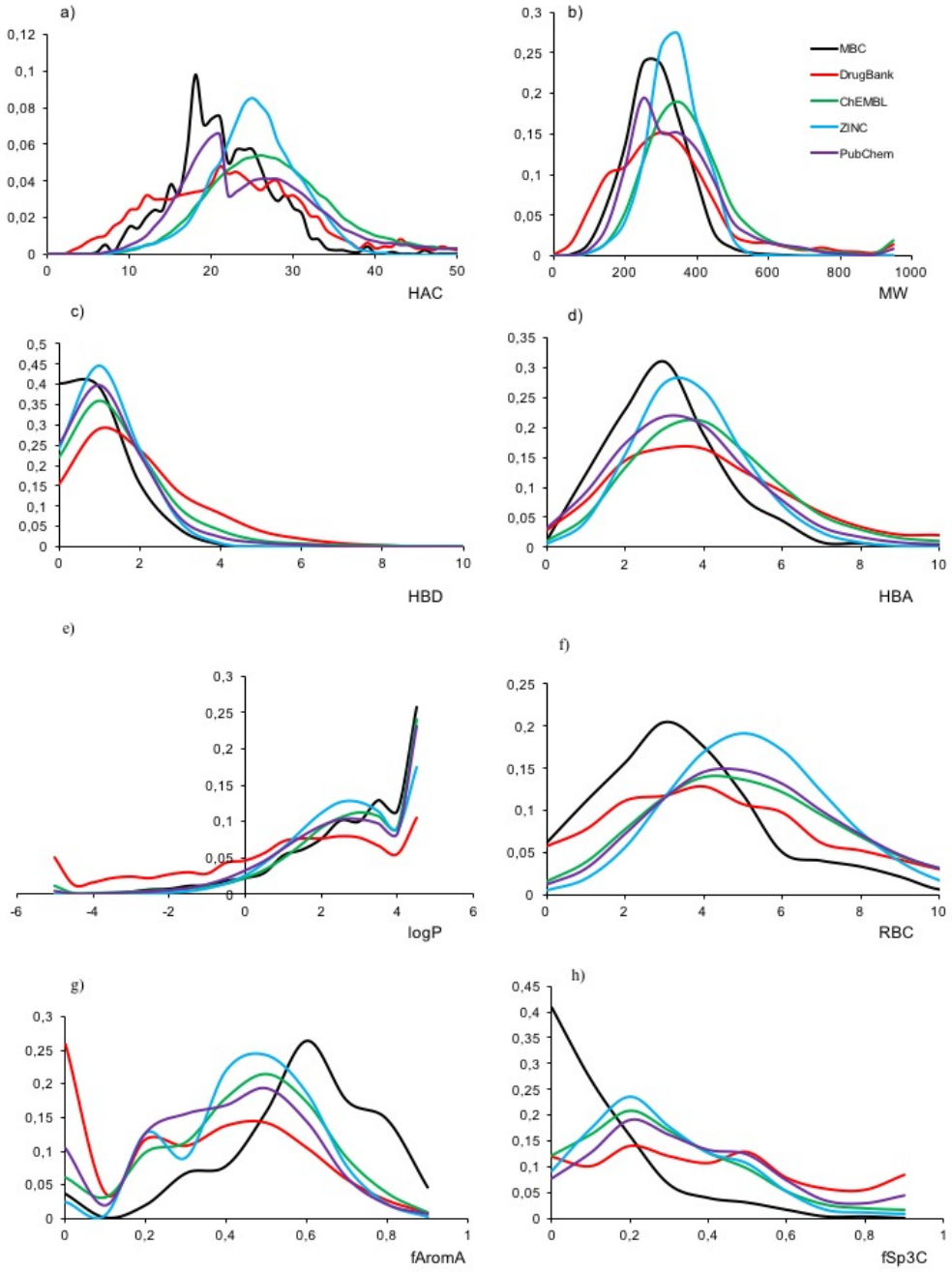

Figure 5. Histogram of molecular properties showing comparison of MBC library with publicly available chemical databases such as DrugBank, ChEMBL, ZINC and PubChem. The labels of y-axis of all histogram is the fraction of database and x-axis is a) Heavy Atom count (all non-hydrogen atoms in molecule), b) Molecular weight, c) Hydrogen Bond Donor atoms, d) Hydrogen Bond Acceptor atoms, e) calculated logP, f) Rotatable Bond Count, g) fraction of aromatic atoms in molecule (0=no aromatic atoms, 1=all atoms in molecule are aromatic), h) fraction of Sp3 hybridized carbons in molecule.

To characterize our library in terms of diversity and novelty, an analysis of the different scaffolds present in the MBC was performed. We compared the diversity of

the library using various fingerprints. The results showed that 49 different chemical scaffolds were found in the analysis (Figure 6).

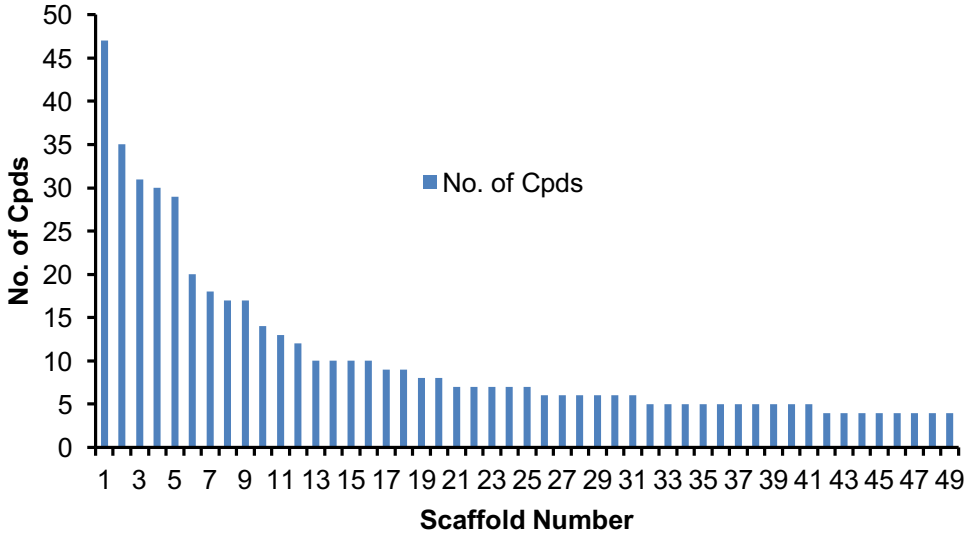

Figure 6. Bemis-Murcko scaffold distribution in MBC library

Finally, we investigated the number of compounds in the MBC library that were novel with respect to other databases. Table 3 shows the number compounds that were unique and the number of compounds and scaffolds that held a commonality between the MBC library and the other libraries (Table 3). This study described the novelty of the MBC in terms of structures and scaffolds. For example, we can see that of the 1,096 compounds in the MBC library, 361 compounds were only in the MBC library; this number takes into account the merger of all the compounds of the four commercial databases (65,516,032). Regarding new scaffolds, the MBC library had 71 unique scaffolds in the context of the number of scaffolds combined in all four of the other databases (9,827,834). Therefore, we can affirm that the MBC chemical library has both novel compounds and scaffolds when considered in the context of other well-known databases.

Table 3. Comparison of MBC with other chemical libraries

Database[1]	Number of compounds	Compounds UNQ	Compounds COMMON
MBC	1,096	0	1,096
DrugBank	6,637	1,058	3
ChEMBL.20	1,311,227	725	336
ZINC	12,189,492	759	302
PubChem	52,008,676	428	633
Merge	65,516,032	361	700

Database[2]	Number of BMS scaffolds	Scaffolds UNQ	Scaffolds COMMON
MBC	444	0	444
DrugBank	3,153	398	46
ChEMBL.20	368,739	178	266
ZINC	2,198,481	185	259
PubChem	7,257,461	84	360
Merge	9,827,834	71	373

[1]Based on smile code; [2]Based on BMS scaffolds, BMS: Bemis Murcko Scaffold; UNQ: unique nº of compounds or scaffolds in MBC library in comparison with other databases; COMMON: nº of compounds or scaffolds present or shared by MBC library and other databases.

CONCLUSION

Our aim is to shorten the drug discovery process by producing high-quality drug-like compounds that generate valuable data from screening programs. The collection of small molecules characterized here, referred to as the MBC library, is a unique collection of small molecules that have enriched drug-like properties. These molecules have been designed entirely by the medicinal chemists in our group. The MBC library contains over 1,000 handcrafted chemical compounds. The analysis reported here, based on computational studies using QuikPro and MQN-Mapplet software, showed that this collection of compounds is of high quality in terms of diversity and drug-like properties, and they are suitable for producing high-quality starting points and for enabling quick starts to drug discovery programs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: xxxxxxxx

Copy of excel file of the MBC database

AUTHOR INFORMATION

Corresponding Author

*E-mail: nuria.campillo@csic.es

ORCID

Nuria E. Campillo: 0000-0002-9948-2665

ACKNOWLEDGMENTS

Financial support from MINECO (grant no. CTQ2015-66313 and SAF2015-65740) and the EC 7th Framework Programme (FP7-HEALTH-2013-INNOVATION-1, PDE4NPD no. 602666). We also thanks to MECD (Grant of V.S, FPU15/01465).

&This paper is dedicated to our friend and colleague Dr. Santiago Conde

REFERENCES

1. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L., How to Improve R&D Productivity: the Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discovery* **2010,** *9*, 203-214.
2. Kola, I.; Landis, J., Can the Pharmaceutical Industry Reduce Attrition Rates? *Nat. Rev. Drug Discovery* **2004,** *3*, 711-715.
3. Miller, M. A., Chemical Database Techniques in Drug Discovery. *Nat. Rev. Drug Discovery* **2002,** *1*, 220-227.
4. Dandapani, S.; Rosse, G.; Southall, N.; Salvino, J. M.; Thomas, C. J., Selecting, Acquiring, and Using Small Molecule Libraries for High-Throughput Screening. *Curr. Protoc. Chem. Biol.* **2012,** *4*, 177-191.
5. Shoichet, B. K., Virtual Sreening of Chemical Libraries. *Nature* **2004,** *432*, 862-865.
6. Lipinski, C. A., Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000,** *44*, 235-249.
7. Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J., Probing the Links Between in Vitro Potency, ADMET and Physicochemical Parameters. *Nat. Rev. Drug Discovery* **2011,** *10*, 197-208.

8. Gleeson, M. P., Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J. Med. Chem.* **2008,** *51*, 817-834.

9. Peters, J. U.; Schnider, P.; Mattei, P.; Kansy, M., Pharmacological Promiscuity: Dependence on Compound Properties and Target Specificity in a Set of Recent Roche Compounds. *ChemMedChem* **2009,** *4*, 680-686.

10. Leeson, P. D.; Springthorpe, B., The Influence of Drug-Like concepts on Decision-Making in Medicinal Chemistry. *Nat. Rev. Drug Discovery* **2007,** *6*, 881-890.

11. Price, D. A.; Blagg, J.; Jones, L.; Greene, N.; Wager, T., Physicochemical Drug Properties Associated with in Vivo Toxicological Outcomes: a Review. *Expert Opin. Drug Metab. Toxicol.* **2009,** *5*, 921-931.

12. Morales-Garcia, J. A.; Redondo, M.; Alonso-Gil, S.; Gil, C.; Perez, C.; Martinez, A.; Santos, A.; Perez-Castillo, A., Phosphodiesterase 7 Inhibition Preserves Dopaminergic Neurons in Cellular and Rodent Models of Parkinson Disease. *PLoS One* **2011,** *6*, e17240.

13. Morales-Garcia, J. A.; Susin, C.; Alonso-Gil, S.; Perez, D. I.; Palomo, V.; Perez, C.; Conde, S.; Santos, A.; Gil, C.; Martinez, A.; Perez-Castillo, A., Glycogen Synthase Kinase-3 Inhibitors as Potent Therapeutic Agents for the Treatment of Parkinson Disease. *ACS Chem. Neurosci.* **2013,** *4*, 350-360.

14. Conde, S.; Perez, D. I.; Martinez, A.; Perez, C.; Moreno, F. J., Thienyl and Phenyl Alpha-halomethyl Ketones: New Inhibitors of Glycogen Synthase Kinase (GSK-3beta) from a Library of Compound Searching. *J. Med. Chem.* **2003,** *46*, 4631-4633.

15. Redondo, M.; Palomo, V.; Brea, J.; Perez, D. I.; Martin-Alvarez, R.; Perez, C.; Paul-Fernandez, N.; Conde, S.; Cadavid, M. I.; Loza, M. I.; Mengod, G.; Martinez, A.; Gil, C.; Campillo, N. E., Identification in Silico and Experimental Validation of Novel Phosphodiesterase 7 Inhibitors with Efficacy in Experimental Autoimmune Encephalomyelitis Mice. *ACS Chem. Neurosci.* **2012,** *3*, 793-803.

16. Garcia, A. M.; Brea, J.; Morales-Garcia, J. A.; Perez, D. I.; Gonzalez, A.; Alonso-Gil, S.; Gracia-Rubio, I.; Ros-Simo, C.; Conde, S.; Cadavid, M. I.; Loza, M. I.; Perez-Castillo, A.; Valverde, O.; Martinez, A.; Gil, C., Modulation of cAMP-specific PDE without Emetogenic Activity: New Sulfide-like PDE7 Inhibitors. *J. Med. Chem.* **2014,** *57*, 8590-8607.

17. Maestro 9.9; Schrödinger, LLC, New York: 2014.

18. LigPrep 3.1; Schrödinger, LLC, New York: 2014.

19. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.

20. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005,** *26*, 1752-1780.

21. QikProp, v. 4.1; Schrödinger, LLC, New York: 2014.

22. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug. Deliv. Rev.* **2001,** *46*, 3-26.

23. Jorgensen, W. L., Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009,** *42*, 724-733.

24. Branch, S. K.; Agranat, I., New Drug Designations for New Therapeutic Entities: New Active Substance, New Chemical Entity, New Biological Entity, New Molecular entity. *J. Med. Chem.* **2014,** *57*, 8729-8765.

25. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular Properties that Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002,** *45*, 2615-2623.

26. Irwin, J. J.; Shoichet, B. K., ZINC-a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005,** *45*, 177-182.

27. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G., ZINC: a Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012,** *52*, 1757-1768.

28. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Kruger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P., The ChEMBL Bioactivity Database: an Update. *Nucleic Acids Res.* **2014,** *42*, D1083-1090.

29. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H., PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016,** *44*, D1202-1213.

30. Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J., PubChem BioAssay: 2017 Update. *Nucleic Acids Res.* **2017,** *45*, D955-D963.

31. Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J., DrugBank: a Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006,** *34*, D668-672.

32. Awale, M.; van Deursen, R.; Reymond, J. L., MQN-Mapplet: Visualization of Chemical Space with Interactive Maps of DrugBank, ChEMBL, PubChem, GDB-11, and GDB-13. *J. Chem. Inf. Model.* **2013,** *53*, 509-518.

The Medicinal and Biological Chemistry (MBC) Library: an efficient source on new hits[&]

Víctor Sebastián-Pérez,[1,§] Carlos Roca,[1,§] Mahendra Awale,[2,§] Jean-Louis Reymond,[2] Ana Martinez,[1] Carmen Gil[1] and Nuria E. Campillo[1,*]

[1]Centro de Investigaciones Biológicas (CIB, CSIC). Ramiro de Maeztu, 9, 28040-Madrid, Spain
[2]Department of Chemistry and Biochemistry. University of Bern, Switzerland
[§]Theses authors have contributed equally to this work