Trends and challenges in chemoinformatics research in Latin America

Jazmín Miranda-Salas [a], Carlos Peña-Varas [b,c], Ignacio Valenzuela Martínez [b],
Dionisio A. Olmedo [d], William J. Zamora [e,f,i], Miguel Angel Chávez-Fumagalli [g],
Daniela Q. Azevedo [h], Rachel Oliveira Castilho [h], Vinicius G. Maltarollo [h], David Ramírez [b],
José L. Medina-Franco [a,*]

[a] DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, México 04510, Mexico
[b] Departamento de Farmacología, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción 4030000, Chile
[c] Departamento de Ciencias de la Computación, Facultad de Ingeniería, Universidad de Concepción, Concepción 4030000, Chile
[d] Centro de Investigaciones Farmacognósticas de la Flora Panameña (CIFLORPAN), Apartado 0824-00178, Facultad de Farmacia, Universidad de Panamá, Ciudad de Panamá, Panamá
[e] CBio3 Laboratory, School of Chemistry, University of Costa Rica, San Pedro, San José, Costa Rica
[f] Laboratory of Computational Toxicology and Biological Testing Laboratory (LEBi), University of Costa Rica, San Pedro, San José, Costa Rica
[g] Computational Biology and Chemistry Research Group, Vicerrectorado de Investigación, Universidad Católica de Santa María, Arequipa 04000, Perú
[h] Department of Pharmaceutical Products, Faculty of Pharmacy, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte, Minas Gerais 31270-901, Brazil
[i] Advanced Computing Lab (CNCA), National High Technology Center (CeNAT), Pavas, San José, Costa Rica

ARTICLE INFO

ABSTRACT

Chemoinformatics is an independent inter-discipline with a broad impact in drug design and discovery, medicinal chemistry, biochemistry, analytical and organic chemistry, natural products, and several other areas in chemistry. Through collaborations, scientific exchanges, and participation in international research networks, Latin American scientists have contributed to the development of this subject. The aim of this perspective is to discuss the status and progress of the chemoinformatic discipline in Latin America. We team up to provide an author´s perspective on the topics that have been investigated and published over the past twelve years, collaborations between Latin America researchers and others worldwide, contributions to open-access chemoinformatic tools such as web servers, and educational-related resources and events, such as scientific conferences. We conclude that linking and fostering collaboration within each nation as well as among other Latin American nations and globally is made possible by open science and the democratization of science. We also outline strategic actions that can boost the development and practice of chemoinformatic in the region and enhance the interaction between Latin American countries and the rest of the world.

1. Introduction

Chemoinformatics is an independent discipline that has a broad range of applications in chemistry [1,2]. It has several formal definitions, for example, "*All concepts and methods that are designed to interface theoretical and experimental efforts involving small molecules*" [3] or "*Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information*" [4]. Other definitions are collectively discussed elsewhere [1]. It should be noted that several terms published in the literature refer to the same discipline and its use sometimes depends on the geographical region. To illustrate this fact, Fig. 1 shows a parallel coordinate plot with the number of papers in PubMed in the period 2010–2022 with the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics." The figure shows an overall increase in the number of publications since 2010 with a slight decrease in 2022. Among the two frequently used terms in conferences and scientific journals, "chemoinformatics" shows a slightly larger number of publications than those containing the term "cheminformatics". It should be emphasized that the community has not yet agreed on a single title, unlike the closely related discipline "bioinformatics," whose term has mostly been ap-
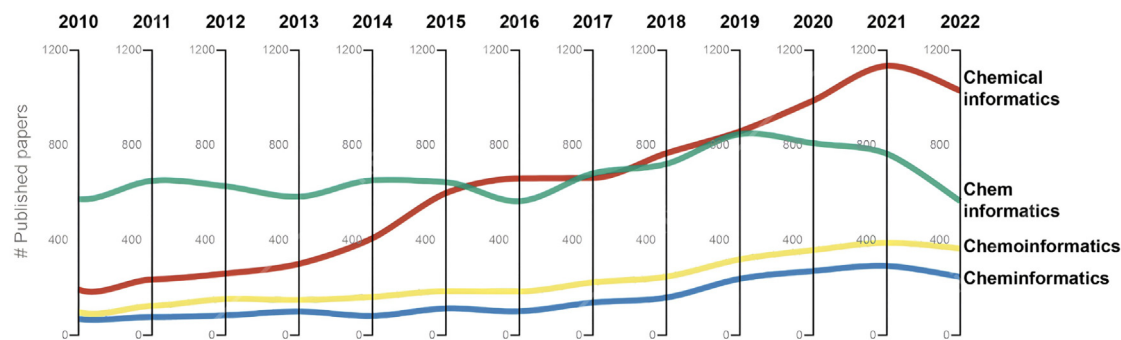
Fig. 1. Number of papers in PubMed in the period 2010–2022 with the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics." The figure shows the trend in usage of the most common terms that refers to the same discipline.
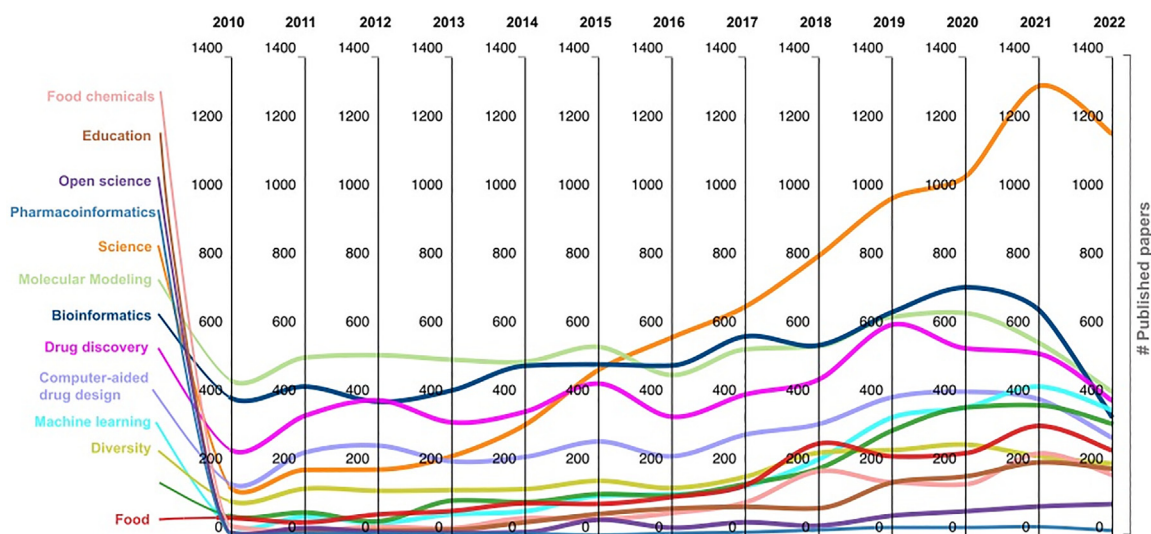

Fig. 2. The number of papers in PubMed with the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics" plus another relevant keyword by year.

proved [2]. For the remainder of this manuscript, we will use the term "chemoinformatics."

Regardless of the specific term, since the first applications of chemoinformatics in the '50s, the discipline has been evolving and expanding rapidly with the most recent and large advances in artificial intelligence (AI) [5]. Chemoinformatics has a broad range of applications across different research areas, including drug design, chemogenomics, and systems pharmacology, as well as medicinal, analytical, organic, cosmetic, food, and inorganic chemistry [6–8], and natural product research [6,9], among others. To illustrate the applications of chemoinformatics in other areas, Fig. 2 shows the number of publications in PubMed with the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics" plus another relevant keyword by year. The analysis reveals that whereas molecular modeling was the primary topic connected to chemoinformatics at the beginning of the examined period (2010–2022), other fields have been involved with time, with education being one of them. This has allowed us to glimpse the tremendous growth of this discipline and its significant application in related areas as the years pass by. In the same period (2010–2022), the most frequent areas published along with chemoinformatics (as measured by the number of publications with concurrent keywords) were mainly molecular modeling and drug discovery with 6935 and 5459 papers, respectively. Other frequent areas published along with chemoinformatics were bioinformatics (6689), computer-aided drug de-

sign (3764), and machine learning (2402) (see also Fig. S1 in the Supplementary material).

The principle of open science has been used to build many of the current chemoinformatics applications and developments, including teaching aids and research tools. Undoubtedly, the democratization of science would increase global connections between research teams and students of all academic levels. It would be easier to make connections between various geographical zones when you are aware of the study that has been done in various worldwide places. To this goal, in 2018 we presented a viewpoint on chemical informatics from a Latin American group [10].

The goal of this manuscript is to discuss the status of Chemoinformatics in Latin America with input from expert research groups working on the region. We also present an author's perspective on the future directions of chemoinformatics, emphasizing the need to strengthen research networks between Latin American scientists as well as with other in different countries, making it possible to increase funding and the multidisciplinary formation of advanced human capital. The paper is organized into six main sections. After this Introduction, Section 2 discusses the status of the research in Latin America. Section 3 presents exemplary research outcomes through representative chemoinformatics-related applications and tools recently developed in Latin America, emphasizing open-access resources. The next Section discusses exemplary contributions to education, training, and scientific dissemination. Section 5 elab-
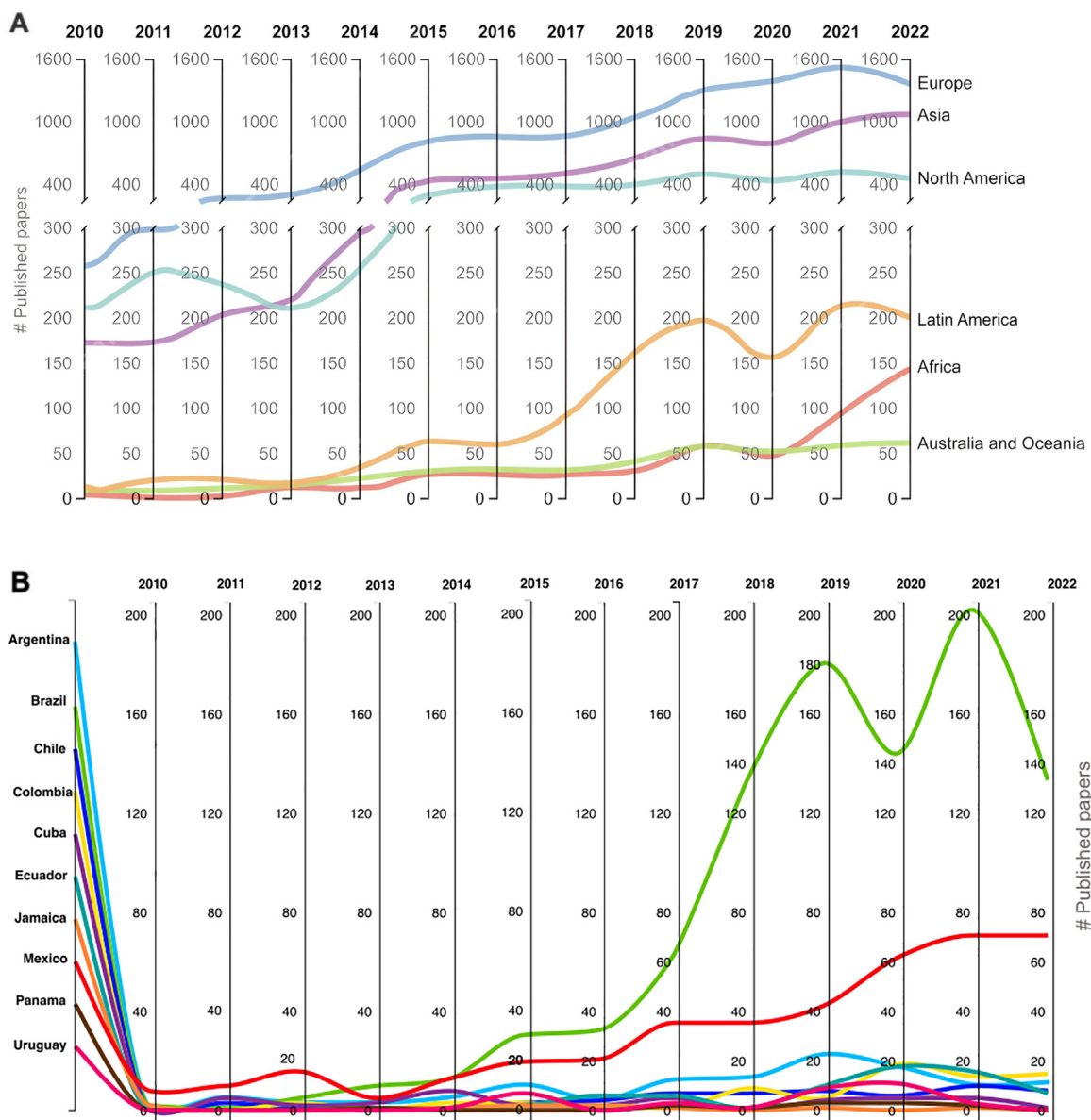
Fig. 3. Number of papers in PubMed in the period 2010–2022 with the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics" **A.** Analyzed by continent. America is separated as North and Latin America. **B.** Analyzed by each Latin American country.

orates on future directions, challenges, and perspectives to strengthen chemoinformatics in Latin America. Section 6 presents summary conclusions.

2. Research: publications, applications, and collaborations

2.1. Publications and main applications

In this section, we discuss contributions by research groups in Latin America towards the development and application of chemoinformatics as analyzed through peer-reviewed publications between 2010 and 2022. Fig. 3 depicts the tremendous expansion of chemoinformatics in Latin America during the previous 12 years. Fig. 3A underlines that, during the previous five years, Latin America's publication rate has significantly increased, even though Europe, Asia, and North America have published many more papers than Latin America overall. Fig. 3B indicates that, as the years pass by, more Latin American countries have been joining the development of this discipline, contributing to more published peer-reviewed papers. By putting together the studies of all Latin American countries per year, it is remarkable the growth that chemoinformatics had during the last coronavirus pandemic (2020–2022). Indeed, almost all the countries that belong to this region have at least one published paper related to this discipline, as it is explicit in Fig. 3B. It can be highlighting that Brazil, Mexico, Argentina, Colombia, and Ecuador are within the top-5 Latin American countries with the largest number of papers, as found in PubMed during 2010–2022. This leads to the conclusion that, thus far, these countries are leading the development of chemoinformatics in Latin America. Given its geographical location throughout the Latin American territory, they may be motivating the rest of the countries, especially those belonging to Central America and the Caribbean, to develop this discipline since it is also relevant to the growing frequency that the rest of the Latin American countries have been generating.

To elucidate which are the main applications that Latin American countries are giving to cheminformatics and to contrast the information obtained until now in PubMed, we made the same research in the same period but in Web of Science. The data was obtained using the keywords "Chemoinformatics OR Cheminformatics OR Chem informat-
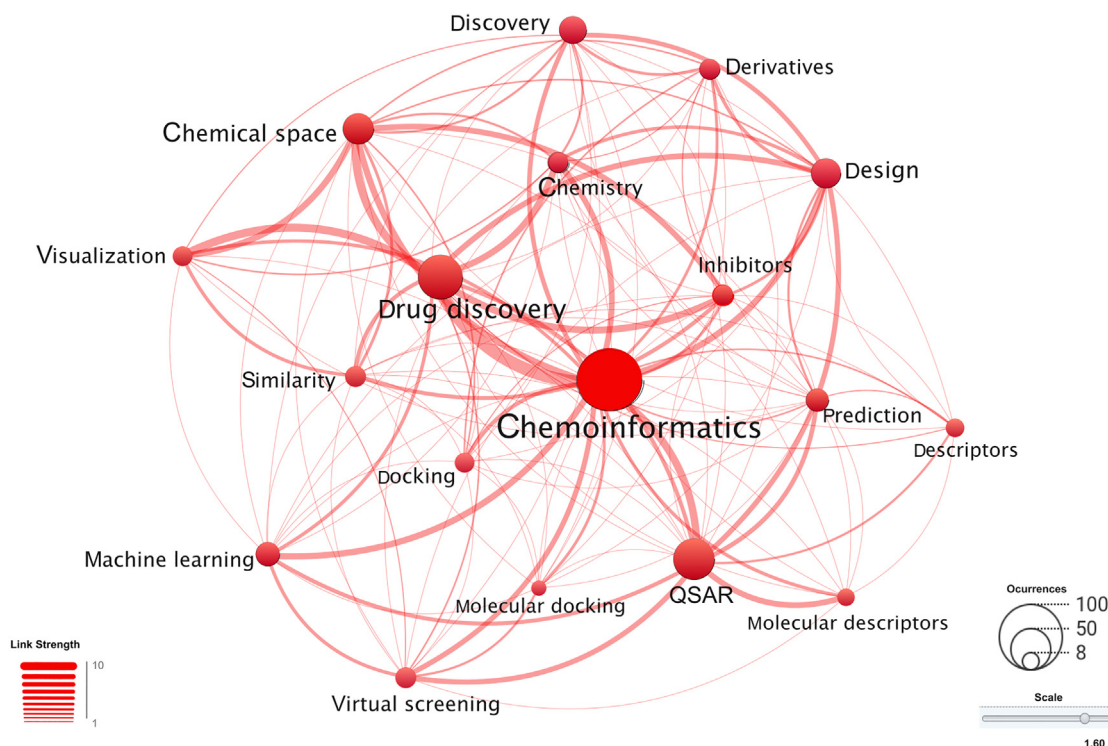
J. Miranda-Salas, C. Peña-Varas, I. Valenzuela Martínez et al.

Artificial Intelligence in the Life Sciences 3 (2023) 100077

Fig. 4. Bibliometric map based on Web of Science data based on keyword co-occurrence during the period 2010–2022. The data were obtained using the keywords "Chemical informatics", "Cheminformatics", "Chemoinformatics", and "Chem informatics" and each of the Latin American countries with published papers in 2010–2022. A total of 474 publications were identified. The relative size of the nodes is related to the number of publications containing the keyword.

ics OR Chemical informatics" for each of the Latin American countries. Surprisingly, there was a significant difference between the data obtained, finding more total publications obtained in Web of Science (536) in contrast with PubMed (451), which also were reflected in fewer peer-reviewed publications found per country as it is depicted in Fig. S5. Fig. 4 shows a network with the key concepts related to chemoinformatics in each Latin American country as found in the Web of Science. The network was created with Visualization of Similarity (VOS) viewer desktop version 1.6.19 (0) (https://www.vosviewer.com/) [11] based on keyword concurrent during the period 2010–2022. This figure depicts the main concepts involved in chemoinformatics. Each keyword and its entanglement are given by the usage of publications, and it is also notable the thickness of the edges in the graph, where chemoinformatics is strongly associated with drug discovery and chemical space. The network shows that drug discovery, quantitative structure-activity relationships (QSAR), and chemical space are amongst the most frequent con-current words and applications published so far by Latin American scientists.

Fig. S2 (Supplementary data) highlights the principal areas of application in which the publications of each Latin American country are focused according to the Web of Science categories. The analysis indicates that the main categories are "Pharmacology Pharmacy", "Chemistry Medicinal", "Chemistry Multidisciplinary", "Computer Science Information Systems", "Computer Science Interdisciplinary Applications", Biochemistry Molecular Biology", "Computer Science Information Systems", "Neurosciences", "Biochemistry Molecular Biology", and Microbiology". The category "Chemistry Multidisciplinary" predominates in Brazil, Mexico, Colombia, Ecuador, and Cuba. For Chile and Saint Lucia "Pharmacology Pharmacy" is the major category. In Argentina, "Chemistry Medicinal" and "Computer Science Information Systems" are, equally, the most frequent categories. Greater variability of focus areas is evident with respect to the previously mentioned countries, and they experience differentiation by the percentages assigned to each focus area. Likewise, the recent growth of focus areas such as

Microbiology and Neurosciences is evident, where the latter has shown greater growth in Chile, while "Microbiology" has greater growth in Saint Lucia, combining percentage points with "Medicinal Chemistry."

2.2. Journals: where research has been published

According to the Web of Science Core Collection, the 536 peer-reviewed articles published by the Latin American countries between 2010 and 2022 appeared in 252 journals. Fig. 5 shows the ten peer-reviewed journals with major usage by Latin American countries, evidencing that the journal with the highest occurrence is the Journal of Cheminformatics, which at least eight Latin American countries have published there. Molecular Informatics is the journal with the largest number of publications, although only seven countries published in it. Furthermore, it is possible to distinguish in which of the journals the highest number of publications by country is found; Brazil, Mexico, and Cuba, have the largest number of its articles published in Molecular Informatics. Colombia, the fourth one according to the Web of Science, has published the most in the Journal of Cheminformatics, which is in turn the second most popular journal.

2.3. Regional and global collaboration networks between Latin American groups

As discussed in Section 2.1. undoubtedly, Latin America has had significant growth in the discipline (i.e., Fig. 3A), although there is still a comparable difference with other regions of the world. This growth is largely implicit thanks to the collaborations that have been generated within Latin America and other geographical regions. These collaborations are represented in Fig. S3, where a research collaboration network is presented. Fig. S3A, C and D identifies the top-10 countries that have the greatest number of collaborations with Latin American countries, highlighting those with a larger number of joint publications through the edge thickness and total number of publications with the node size.
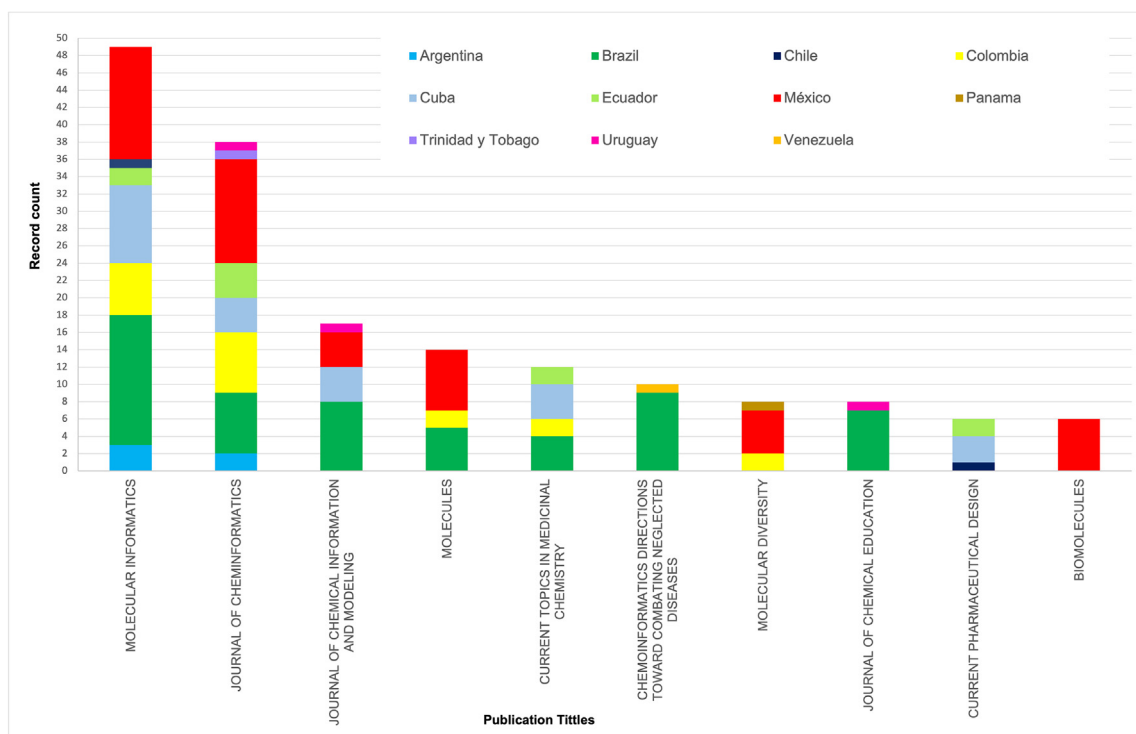
Fig. 5. Top ten peer-reviewed journals publishing chemoinformatics papers by Latin American countries during 2010–2022 according to the Web of Science Core Collection.

We can find a greater collaboration with respect to the number of articles with North America, but Europe also generates a relevant collaboration with Latin America by involving a larger number of countries implicated in said collaborations (six out of ten countries). Finally, Asia and Oceania, although they are only shown with one country within the top ten collaborators for each of these two continents, they illustrate a relevant connection with Latin American countries.

Regional collaborations are shown in Fig. S3B, where we visualize the important support network that involves the entire region, and that has allowed Latin America to grow in the development of chemoinformatics.

Figs. S3C and S3D show the interactions using an alternative graphic method, to further highlight each collaboration between Latin American countries and other regions of the world. Fig. S3C distinguishes Brazil as the Latin American scientific superpower that has the highest normalized collaboration with the different countries described, foregrounding among them USA, Germany, Sweden, and Japan. Mexico is the second country with the highest number of collaborations between the same countries and only compares Brazil with Spain regarding the number of collaborations.

Since data normalization with respect to the number of collaborations from the two Latin American countries with the highest number of publications makes it difficult to visualize this last graph for the rest of the Latin American countries, Fig. S3D was generated. Fig. S3D allows us to observe in greater detail the differentiation by each Latin American country with respect to its collaboration with the rest of the ten countries with the greatest collaboration, visualizing that Argentina, Chile, Colombia, and Peru in general have very similar trends, differing in that Peru has a very marked and greater collaboration with Italy while Argentina, Chile and Colombia have it with Spain in different proportions.

Similar trends are perceived with the rest of the Latin American countries, where Ecuador, like Brazil, has a greater collaboration with the USA. Jamaica, Uruguay, Panama, Cuba, Venezuela, and Costa Rica remain in a limited interval of collaborations. Finally, Nicaragua, Barba-

dos, El Salvador, and Trinidad and Tobago turn out to be the countries with the least collaborations found.

Although we have reduced our analysis to a specific number of top-10 non-Latin American countries with which we collaborate, there is a greater number of collaborations with the rest of the world that is illustrated in Fig. S4A (Supplementary material) which shows the collaborations between Latin America and the rest of the world, visualizing a great interaction with these countries. Fig. S4B illustrates the collaboration between non-Latin American countries, and although a greater number of collaborations are visualized, this Fig. S4B allows to infer the relevance of the Latin American collaborations.

Based on the collaborations and analysis, it is necessary to underline the vast effort that is being made at the regional and global level to contribute to the development of this discipline, finding diverse collaborations between different geographical areas that must undoubtedly be maintained or continue to increase.

3. Resources and applications for research

Over the past few years academic groups in Latin America have been developing open-source and freely available resources for research focused on chemoinformatic and related disciplines. These tools form part of the increasingly open chemoinformatic resources developed and used worldwide [12–14]. Part of such contributions are the compound databases. This section summarizes examples of chemoinformatics-related tools developed by Latin American groups in the past twelve years, with emphasis on open-access resources.

Table 1 summarizes web servers, standalone tools, and other resources recently developed by Latin American countries in the public domain. All the tools listed in the table are characterized by having intuitive use with a user-friendly interface accompanied by their respective tutorials. The tools also provide at least one means of communication with the users for questions, feedback, or comments in general, which further facilitates interactions between Latin American scientists. Several of the platforms listed in the table offer various types of com-

Table 1
Exemplary web servers, standalone tools, and other resources developed in Latin America.

Country	Resource	Application - use	Ref., URL
Argentina	LIDeB Tools	Publicly available open-source customizable cheminformatics tools to be used in computer-assisted drug discovery.	[34], https://lideb.biol.unlp.edu.ar/?p=1237
Brazil	UniCamp	Molecular Dynamics Simulations and custom software tools for analyzing the structure and thermodynamics of solutions using simulation data.	http://leandro.iqm.unicamp.br/m3g/main/home.shtml
Chile	Pharmacoinformatics and drug design	Computational polypharmacology with wet lab analyses to strengthen the drug design and development processes.	https://ramirezlab.github.io/5.1_resources
Chile	CBSM Center of Bioinformatics and Molecular Modeling	More than resources done and under development for modeling and analyzing protein sequences and structures.	https://cbsm.utalca.cl/?page_id=2073
Colombia	MPTG-CBP	Computer simulations at different levels of resolution and complexity to decipher how biological macromolecules interact with each other to accomplish their function.	https://mptg-cbp.github.io/
Costa Rica	Cbio3 Laboratory	Codes in GitHub repositories for the calculation of useful descriptors in chemoinformatics: i) pH-dependent lipophilic descriptors for amino acids ii) descriptors based on the count of n conjugated systems with $n = 2$ ($n = \pi$ electrons/2) relevant for the prediction of partition coefficients in an aromatic solvent such as toluene.	https://cbio3group.netlify.app/
Mexico	DIFACQUIM, D-Tools	Five free web services to analyze chemical diversity and chemical space, generate *in silico* libraries of peptides, and predict bioactivity profiles.	https://www.difacquim.com/d-tools/
Uruguay	Institut Pasteur de Montevideo	Dedicated to the development of methods that allow advanced simulations to be carried out but with low computational cost, to improve the comparability of theoretical studies with biochemical, biophysical, or molecular biology experiments.	https://pasteur.uy/laboratorios/simulaciones-biomoleculares/

putational studies beyond chemoinformatics e.g., tools associated with pharmacoinformatics in general, molecular dynamics, and biophysics. Within chemoinformatics, the web servers and tools are mostly tuned for drug discovery and development of lead candidates but also can be used for other research areas, as discussed in Section 2.1.

3.1. Compound databases

Concerning the development of compound databases, perhaps one of the most notable contributions in the region is the generation, curation, and update of natural products databases freely accessible. For more than ten years, Brazil has been developing an extensive database of natural products, and several other Latin American countries have followed the initiative. Table 2 summarizes exemplary natural product databases in the public domain. The status of the development of natural product databases in Latin America has been reviewed recently. Notably, there is a multi-country effort to put together a unified Latin American Natural Product Database – LANaPDB [15,16].

3.2. Blind challenges for validating chemoinformatics methods

As mentioned in the previous section, the Latin American region, through its research groups, has generated knowledge regarding the development of databases, web servers, and chemoinformatics tools. In a complementary way, the blind challenges known as Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL Challenges [17]) funded by the USA National Institutes of Health has provided an excellent opportunity for Latin Americans to validate the impact of their tools and methodologies through the participation of researchers in these challenges. The section on the prediction of physicochemical properties in SAMPL Challenges has included predictions of solvation energies [18,19], acidity constant [20,21] and partition coefficients [21,22] in various biphasic systems (cyclohexane/water and *n*-octanol/water) relevant for rational drug design.

Table 3 shows the participation of Latin American research groups in predicting physicochemical properties of drug-like molecules in the period analyzed (2010–2022). It can be noted that our participation in the SAMPL-3, SAMPL-4, and SAMPL-5 blind challenges was null, however,

from the opening of the SAMPL-6 blind challenge (2017–2018) the incursion of Latin American groups from Argentina, Chile, and Costa Rica took place.

Despite the reduced representation in the SAMPL challenges, it is worth noting that valuable performance was obtained in the submissions where the participation of groups from Latin America was included. Fig. 6 shows that two of four submissions for predicting the *n*-octanol/water log*P* for eleven fragment-like small molecules endowed with kinase inhibitory bioactivity were in the top 50 submissions, however, considering the submissions by categories based on the computational methods used, the results of the groups belonging to our region were in the top 10.

In the case of the SAMPL-7 blind challenge, it was planned in three parts. First, the scientific community was proposed to predict the acidity constant of 22 N-sulfonamides compounds, then the *n*-octanol/water partition coefficient for the same compounds, and finally for the submissions with the first two parts completed, the distribution coefficient at pH 7.4 was predicted.

Fig. 7 represents the scarce participation for the pK_a prediction call, which was not surprising considering that the organizers did not request the macroscopic pK_a of compounds, but rather the pK_a of various microstates to predict the experimental observable and hence the predictive techniques based on in machine learning approaches did not participate due to the impossibility of having experimental values of microstates. In fact, the submissions received were all from the physical category based on first-principles methods, even though some of these submissions did the conformational analysis using the open chemoinformatics tool such as Open Babel [23]. Interestingly, the participation of Latin American groups obtained excellent performances, among which was obtained the second best-ranked submission.

On the other hand, Fig. 8 depicts 17 submissions to the SAMPL7 n-octanol/water log P prediction challenge. This second part of the challenge was encouraging for the Latin American region. One of the participating groups (University of Costa Rica) using a classical machine learning model such as multiple linear regression and by calculating molecular descriptors using chemoinformatics tools accomplished the highest accuracy, among empirical methods and also in all submissions based on the ranked ones. In addition, the collaboration between the University
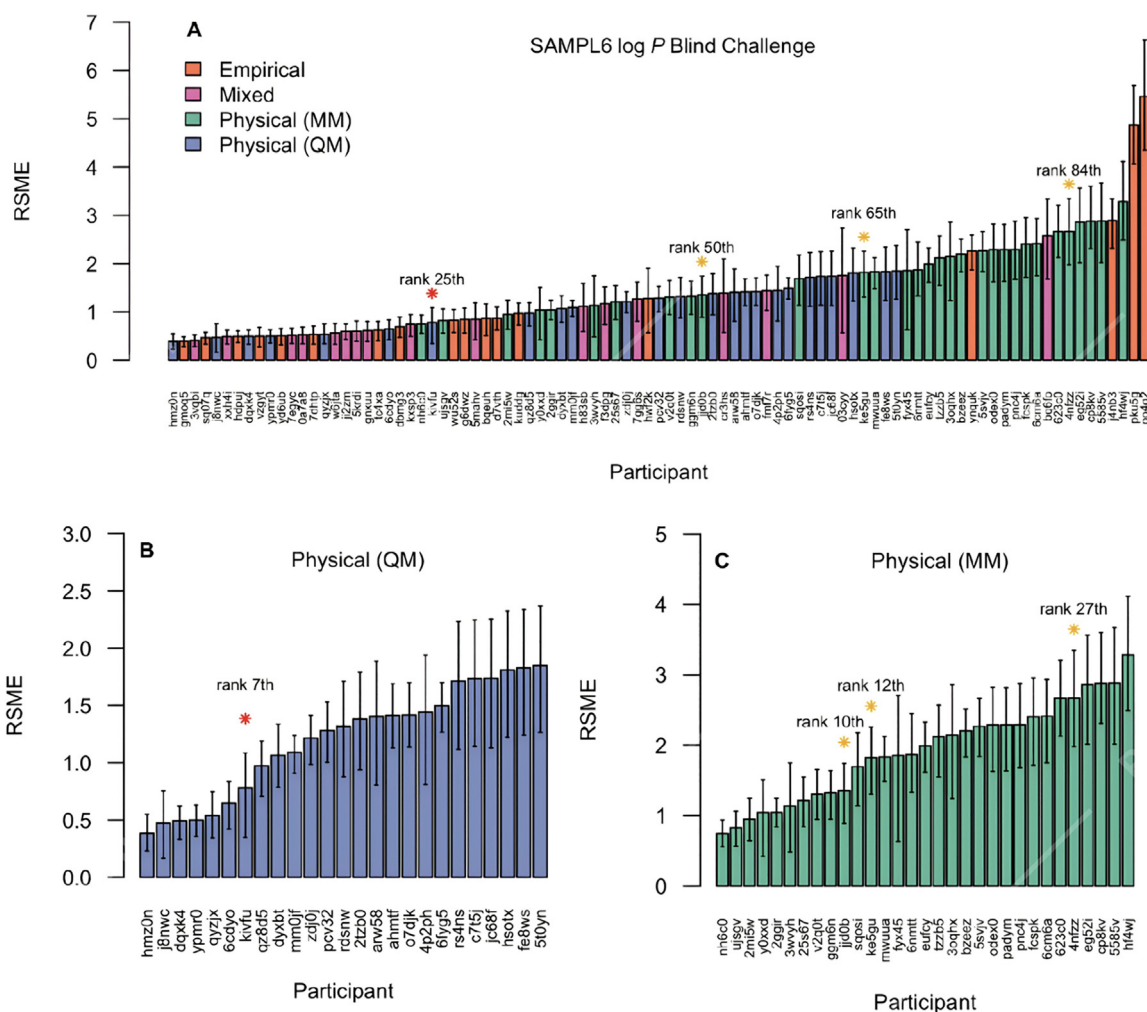
Fig. 6. Representation of the root-mean square error (RMSE) of **A.** the 91 submissions to the SAMPL6 *n*-octanol/water log *P* prediction challenge (the participation of groups from Latin America are marked with stars where the red star represent to the submission of the University of Costa Rica in collaboration with the University of Barcelona and yellow stars the submissions of the Austral Universidad of Chile in collaboration with the University of Concepción), **B.** Twenty five submissions provided by physical methods using quantum mechanical approaches, and **C.** Thirty one submissions provided by physical methods using molecular mechanics approaches. The rank obtained by each Latin American submission both globally and by category is represented at the top of each mark. Details of the methods and participants can be found in Table S1 (Supplementary material) and https://github.com/samplchallenges/SAMPL6/tree/master/physical_properties.
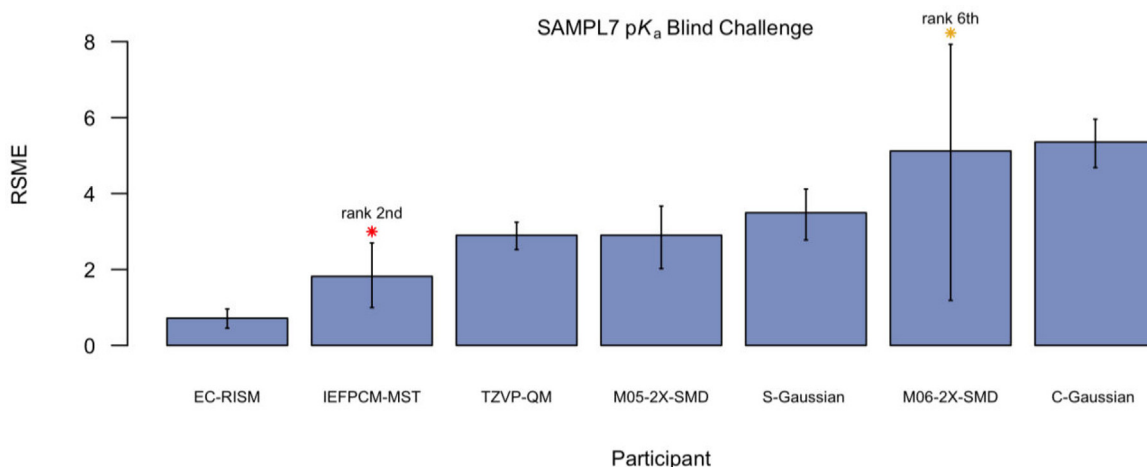

Fig. 7. Representation of the root-mean square error (RMSE) of seven submissions to the SAMPL7 to the pK_a challenge. The participation of groups from Latin America are marked with stars where the red star represents the submission of the University of Costa Rica in collaboration with the University of Barcelona and the orange star the submissions of the National University of Santiago del Estero, Argentina in collaboration with the Miami University. The rank obtained by each Latin American submission is represented next to each mark. Details of the methods and participants can be found in Table S2 (Supplementary material) and https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/pKa.

J. Miranda-Salas, C. Peña-Varas, I. Valenzuela Martínez et al.

Artificial Intelligence in the Life Sciences 3 (2023) 100077

Table 2

Exemplary compound databases of natural products generated in Latin America.

Database	Country	General description (link)[a]	Refs.
NaturAr	Argentina	Natura products from Argentina. https://naturar.quimica.unlp.edu.ar/es/	[35]
NuBBE_DB	Brazil	Natural products of Brazilian biodiversity. Developed by the São Paulo State University and the University of São Paulo. Includes natural products from plants, microorganisms, terrestrial animals, and marine animals. http://nubbe.iq.unesp.br/portal/nubbe-search.html	[36,37]
SistematX	Brazil	Database composed of secondary metabolites from plants, developed at the Federal University of Paraiba. https://sistematx.ufpb.br/	[38,39]
UEFS	Brazil	Natural products from plants that have been published separately. Developed at the State University of Feira de Santana. http://zinc12.docking.org/catalogs/uefsnp	[40]
BRACOLI	Brazil	Brazilian Compound Library is a new manually curated virtual substance library developed by Brazilian research groups to support computer-aided drug *design* work. The first version of BraCoLi has 1176 substances. Contains biological and chemical information on synthetic, semi-synthetic and natural substances.	
avMpNp	Brazil	Antiviral Medicinal Plants and Natural Products is a DB developed in Brazil and contains bioactive substances from biodiversity with antiviral activity.	
NAPRORE-CR	Costa Rica	Natural products isolated from plants and microorganisms. The database is being developed by the Cbio[3] Laboratory, School of Chemistry at the University of Costa Rica.	[41]
LAIPNUDELSAV	El Salvador	Developed by the Research Laboratory in Natural Products of the University of El Salvador.	
UNIIQUIM	Mexico	Natural products from plants mainly isolated and characterized at the Institute of Chemistry of the National Autonomous University of Mexico. https://uniiquim.iquimica.unam.mx/	[42]
BIOFACQUIM	Mexico	Natural products isolated and characterized in Mexico at the School of Chemistry of the National Autonomous University of Mexico and other Mexican institutions. Includes compounds from plants, fungi, propolis, and marine animals. version 2 https://figshare.com/articles/dataset/BIOFAQUIM_V2_sdf/11312702	[43,44]
CIFPMA	Panama	Natural products from plants that have been tested in over 25 in vitro and in vivo bioassays for different therapeutic targets. Developed at the University of Panama.	[45]
PeruNPDB	Peru	Database created at the Catholic University of Santa Maria with natural products from animals and plants. https://perunpdb.com.pe/	[46]
PSC-db	Chile	Free accessible database containing the 3D structures of several plant secondary compounds along with their physicochemical and pharmaceutical properties. http://pscdb.appsbio.utalca.cl/viewSearch/index.php	[47]

[a] When available.

Table 3

Participation of Latin American research groups in global challenges for the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL Challenges [17]).

Challenge	Physicochemical blind challenge	Participants	Latin American participants in special issue publications	Information	Refs.
SAMPL3 (2011–2012)	Prediction of 36 solvation free energies for three series of chlorinated compounds	21 submissions 8 different groups/participants	0	https://github.com/samplchallenges/SAMPL3/tree/main/solvation_energy	[18]
SAMPL4 (2013–2014)	Prediction of hydration free energies for a series of 47 small molecules.	49 submissions 19 different groups/participants	0	https://github.com/samplchallenges/SAMPL4/tree/main/hydration_free_energy	[19]
SAMPL5 (2015–2016)	Prediction of distribution coefficient predictions were introduced for the first time for a set of 53 small molecules	76 submissions 18 different groups/participants	0	https://github.com/samplchallenges/SAMPL5/tree/main/distribution_coefficients	[50]
SAMPL6 (2017–2018)	Prediction of microscopic and macroscopic pKas of 24 small organic molecules protein kinase inhibitors	37 submissions 11 different groups/participants	0	https://github.com/samplchallenges/SAMPL6/tree/master/physical_properties/pKa	[20]
SAMPL6 (2017–2018)	Prediction of the octanol/water partition coefficients (log P) of 11 small molecules protein kinase inhibitors	91 submissions 27 different groups/participants	4 submissions 2 groups (Chile and Costa Rica/Spain)	https://github.com/samplchallenges/SAMPL6/tree/master/physical_properties/logP	[22,51,52]
SAMPL7 (2019–2020)	Prediction of the pKa for 22 N-sulfonamides	9 submissions 7 different groups/participants	2 ranked submissions 2 groups (Argentina and Costa Rica/Spain)	https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/pKa	[21,53,54,55]
SAMPL7 (2019–2020)	Prediction of the octanol-water partition coefficients for 22 N-sulfonamides	33 submissions 17 different groups/participants	3 groups (Argentina/USA, Costa Rica/Spain, and Costa Rica)	https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/logP	
SAMPL7 (2019–2020)	Prediction of the octanol-water distribution coefficients for 22 N-sulfonamides combining the pKa and logP predictions	6 submissions 6 different groups/participants	2 ranked submissions 2 groups (Argentina/USA and Costa Rica/Spain)	https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/logD	

of Costa Rica and the University of Barcelona using the IEFPCM-MST model presented very good results with both globally (rank 8th) and physical (rank 2nd) methods. Considering the difficulty of these challenges, the collaboration between the National University of Santiago del Estero, Argentina with the Miami University had also good results (rank 15th).

Fig. 9 shows the performance based on the root-mean square error of six submissions to the SAMPL7 *n*-octanol/water log *D* prediction challenge. Only quantum chemistry-based methods considered the effect of solvent. The conformational space of the 22 molecules was explored using cheminformatics tools. As expected, due to the outstanding pK_a

and log*P* prediction results of the collaboration between the University of Costa Rica and the University of Barcelona using the IEFPCM-MST model, the best ranked method was obtained to replicate the experimental results of pH-dependent distribution coefficients for the 22 sulfonamides that the challenge proposed.

In summary, Latin American groups have achieved important results in the blind challenges mentioned in this work, positioning the region as a benchmark for the reliable prediction of physicochemical properties of drug-like compounds. The experimental determination of these properties requires a high investment of time and resources, so having expert groups in the region on these issues may be attractive for future collab-
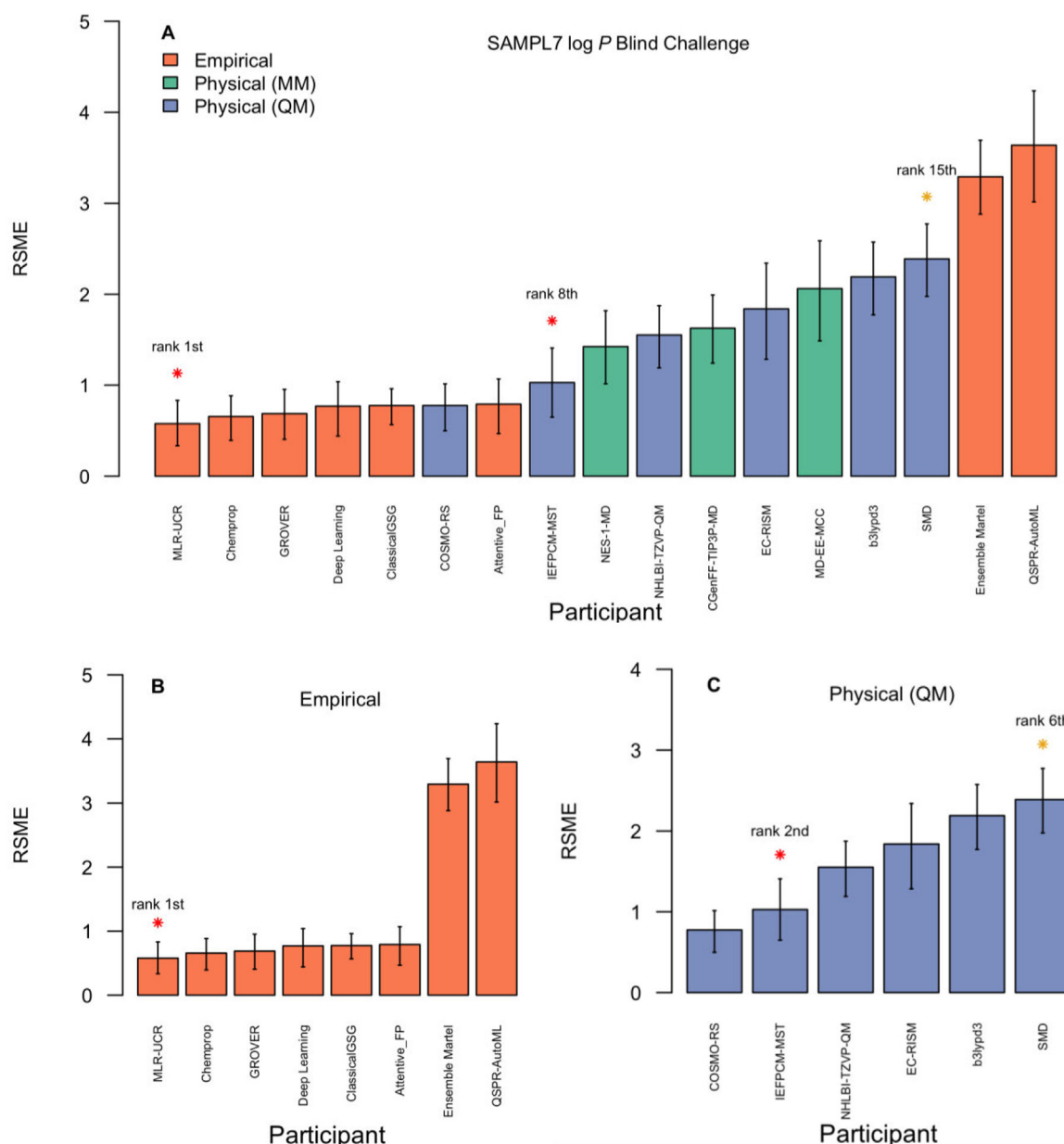
Fig. 8. Representation of the root-mean square error (RMSE) of **A.** the 17 submissions to the SAMPL7 *n*-octanol/water log *P* prediction challenge (the participation of groups from Latin America are marked with stars where the red star represent to the submission of the University of Costa Rica (rank 1st), University of Costa Rica in collaboration with the University of Barcelona (rank 8th) and orange stars the submissions of the National University of Santiago del Estero, Argentina in collaboration with the Miami University (rank 15th)), **B.** Eight submissions provided by empirical methods using chemoinformatics and machine learning approaches, and **C.** Six submissions provided by physical methods using quantum mechanics approaches. Details of the methods and participants can be found in Table S3 (Supplementary material) and https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/logP.

orations between academic groups, not only with the pharmaceutical industry, but also from the food and chemistry industry, and in environmental policies where these properties are used to determine risk factors for exposure to chemicals.

4. Education, training, and dissemination of science

In contrast to bioinformatics where it is common to find specialized courses, textbooks, and educational resources, it is still rare to find courses, and even more, graduate programs focused on chemoinformatics. This is not only a particular case of Latin America but worldwide. Indeed, as commented elsewhere [2], chemistry departments and the schools of pharmacies in universities are conservative in implement-

ing computational chemical education and cheminformatics education in their undergraduate and graduate programs. Indeed, much of the chemoinformatics discipline, as we perceive it today, initiated at the pharmaceutical industry instead of academic settings, which also explains the strong orientations of chemoinformatics toward drug design and discovery. In any case, it is highly relevant to contribute to the education and training of highly qualified researchers to enter into the increasingly demanding job market requesting professionals trained in chemoinformatics.

Historically, Brazil has researchers and professors who have actively worked in the field of chemoinformatics and, therefore, have considerable scientific dissemination in the area. This fact is observed by the prominence of this country related to the number of publications asso-
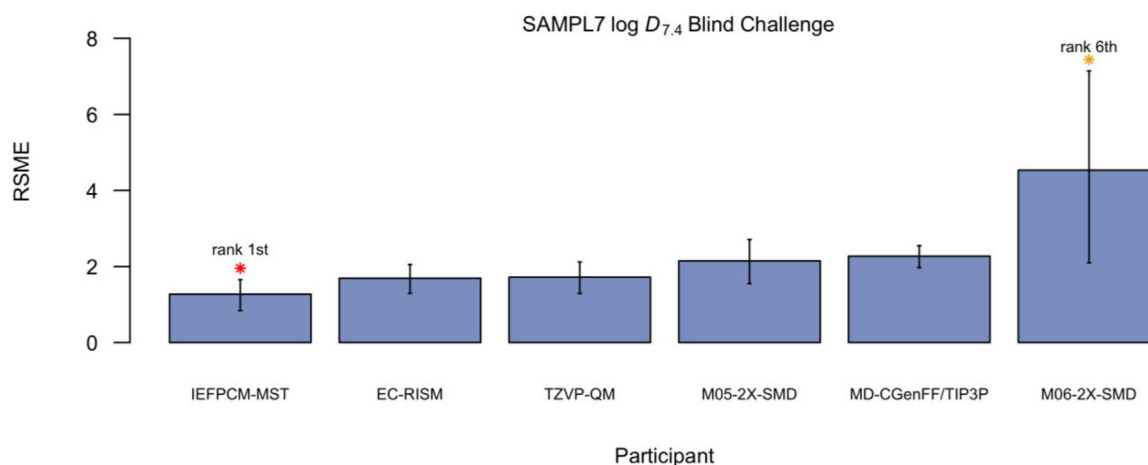
Fig. 9. Representation of the root-mean square error (RMSE) of six submissions to the SAMPL7 *n*-octanol/water log *D* prediction challenge. The participation of groups from Latin America are marked with stars where the red star represents the submission of the University of Costa Rica in collaboration with the University of Barcelona and the orange star the submissions of the National University of Santiago del Estero, Argentina in collaboration with the Miami University. The rank obtained by each Latin American submission is represented next to each mark. Details of the methods and participants can be found in Table S4 (Supplementary material) and https://github.com/samplchallenges/SAMPL7/tree/master/physical_property/logD.

ciated with the theme (Fig. 3B and 3C). In an evaluation of the last 20 years, these publications emphasized the Brazilian panorama regarding medicinal chemistry, natural products chemistry and molecular biology, addressing their advances and challenges, especially in the sphere of education and research. Several relevant topics and historical landmarks were contextualized and discussed throughout this period, such as the role of pharmaceutical and chemistry educational institutions in the development of the area, the integration of multidisciplinary knowledge required for its expansion, and the evolution in the area for application of systems and computational tools designed to extract knowledge from a huge set of information [24,25].

In the 1980s, the sector was driven by the Human Genome Project, with great progress in the areas of genomics and molecular biology, whose essential technology was bioinformatics for increasing analysis of data generated in DNA, RNA and protein sequences [26]. Therefore, by the late 1990s there were a significant number of biomolecules in GenBank. Also, in the 90's the use of QSAR spread worldwide. In Brazil due to international collaborations, especially with Professor Hugo Kubinyi, from the University of Heidelberg, Germany, and, consecutively, courses and workshops given in the country. One of the political consequences of this event was the creation of the Chemical Structure and Biological Activity Division in the Brazilian Chemical Society, forerunner of the Medicinal Chemistry Division, the latter formed in 1998 [27,28].

Already in the 2000s, the discussions that permeated chemoinformatics in the scientific and educational context were based on the standardization of terms and methods, such as the elaboration of a glossary of terms used in drug planning and the improvement of tools, particularly docking and molecular dynamics [29–31]. The classical versus industrial paradigm in the process of drug discovery was also a debated issue, as well as the availability of a large virtual database of small molecules and proteins (PDB) of natural and synthetic origin, drove the development of software for *in silico* screening, applied to the approach known as virtual screening [32].

On this occasion, the published books authored by distinguished Brazilian professors and researchers, such as "Medicinal Chemistry: the molecular basis of drug action" by Barreiro & Fraga, "Methods of Theoretical and Molecular Chemistry" by Morgon, and "Bioinformatics: from biology to molecular flexibility" by Verli, became indispensable bibliographic references in disciplines focused on medicinal chemistry in undergraduate and graduate courses throughout Brazil. Furthermore, the development of new, safer and more effective drugs for many diseases has been demonstrated, as well as the influence of educational institu-

tions in this process, and an example of relevance was the development of captopril. By applying the "structure-based drug discovery" strategy, captopril was the first antihypertensive drug elucidated, whose mechanism of action is through the inhibition of the angiotensin-converting enzyme (ACE). Its discovery benefited from pharmacological studies on the effects of the venom of the Brazilian snake *Bothrops jararaca*, carried out by Rocha e Silva's group, and later by Sérgio H. Ferreira from the University of São Paulo Medical School, demonstrating its vasodilator action, dependent on its ability to potentiate the activity of bradykinin [56].

In parallel, the understanding of the molecular aspects directly involved in the emergence of diseases, led to the expansion of technological resources available for pharmaceutical research in the search for new drugs of natural and synthetic origin, as well as the insertion of new disciplines in undergraduate and graduate courses related to the molecular bases and interactions between the metabolism and drugs. In addition, there was an impact on the automation of processes and data analysis, with automated screening on a large scale, Automated high-throughput screening, coupled with the development of new methods of organic synthesis, such as combinatorial chemistry and the analysis of medicinal plants and natural products by dereplication and studies of its action by metabolomics studies. In general, education and research in the various areas of science that use chemoinformatics tools has increased exponentially in the last two decades, the research groups are better distributed throughout Latin America and their scientific contributions have significant impact in the science, with consequent publications in the main journals of the area as shown in Fig. 5.

Toward this end, this section summarizes representative educational resources developed in Latin America e.g., web-based applications and tools developed for educational and research purposes. We also talk about frequent regional and worldwide conferences, seminars, and schools that are entirely or mostly organized by scientists from Latin America and where chemoinformatics (or a directed discipline) is the main topic or a significant portion of the scientific agenda.

Table 4 summarizes the event, main purpose, and target audience. In several cases, the conferences, webinars, and educational/tutorial material have been made available on demand (when applicable, the URL is included in the table). Throughout the table, different types of materials are displayed in the area, where many of them have educational approaches with which this discipline can increase in use. For example, some websites are included that contain tutorials on the use of chemoinformatics tools and that allow the introduction or specification of greater

Table 4

Examples of recent conferences, workshops, and educational resources generated in Latin America with a focus on chemoinformatics or directly associated disciplines.

Type	Event or resource	Description (target audience)	Access/Refs.
Meeting online	Colloquium: Chemoinformatics and Artificial Intelligence	Online meeting for the general public with an emphasis on undergraduate and graduate students.	https://www.difacquim.com/english/events/2022-colloquium/ [48]
Meeting	LatinXChem	A virtual forum through which the community of Latin American chemists located anywhere in the world can share and discuss their research results. It includes a section dedicated to computational chemistry.	https://www.latinxchem.org/lxchemcomp
Workshop	CABANA Workshop: Chemoinformatics in Drug Discovery	The aim was to introduce researchers to Chemoinformatics, especially in structure-based and ligand-based drug design. Topics covered: the use of protein, ligand and drug databases, protein modeling, molecular docking and virtual screening for drug discovery applications; managing and analyzing virtual chemical libraries; machine learning algorithms in drug discovery.	https://www.ebi.ac.uk/training/events/cabana-workshop-chemoinformatics-drug-discovery/
Workshop	Computational Chemistry in Drug Discovery and Development. From Academia to Industry. Computer-Aided Drug Design	The objective was to introduce the concepts of Chemoinformatics and its applications in the academic environment. In addition, show national researchers topics related to cheminformatics, data mining, database representations in 2D and 3D.	Information in the office of Sponsor SENACYT.
Conference	Integrated Chemoinformatics approaches to virtual screening in the search of new lead compounds against Leishmania	This conference provided information on chemotherapy against Leishmania where the development of CYP51 inhibitor drugs is sought as a new purpose, so in general terms the conference focused on the use of chemoinformatics as a complement to virtual detection approaches (VS) based on ligands and structures.	https://www.researchgate.net/publication/235412450_Integrated_chemoinformatics_approaches_to_virtual_screening_in_the_search_of_new_lead_compounds_against_Leishmania
Conference	Informatics method to bridge gap between experimental results and simulation for carbon nanotube reinforced composites	It included applications of computational chemistry to simulation for carbon nanotube reinforced composites.	https://repositorioslatinoamericanos.uchile.cl/handle/2250/926817
Conference	Latin American Chemistry and Catalysis.	It included applications of computational chemistry to catalysis.	https://www.youtube.com/watch?v=wlRbuIf2Hbw
Conference	Women in Bioinformatics & Data Science-Latin America.	Recurrent meeting bringing together Latin American researchers who work in the areas of systems biology, omics technologies, artificial intelligence, machine learning, data science, data mining, and high-performance computing with applications in biology.	https://wbds.la/ [49]
Conference	School of chemoinformatics	Online event discussing developments and applications and chemoinformatics, including chemical space, data analysis, natural products, virtual screening, toxicity prediction, and drug discovery for neglected diseases. It included a hands-on session to use the ChEMBL. Editors of chemoinformatics journals discussed points related to scientific publishing.	https://www.youtube.com/@SchoolChemInfLA/videos
Conference	Digital Technologies, Innovation and Health in the Covid-19 Pandemic | PROSIC - UCR	Online event to publicize the uses and practical applications of information and communications technologies during the pandemic caused by the Covid-19 virus, which included a talk on the role of Chemoinformatics and Computational Biology during the pandemic.	https://www.youtube.com/watch?v=T49FvEk6aCg™t=2196s
Educational / tutorial	RamirezLab Github	Website with tutorials, script library and FAQ for pharmacoinformatics and drug design.	https://github.com/ramirezlab/WIKI
Educational / Course	Practical Aspects of Drug Discovery: At the Interface of Biology, Chemistry and Pharmacology - Latin America and the Caribbean	Course open to applicants from institutes based in Latin America and the Caribbean. The program was aimed at researchers with a background knowledge of any discipline related to drug discovery, including biology, chemistry, pharmacology, computational chemistry and informatics.	https://coursesandconferences.wellcomeconnectingscience.org/event/practical-aspects-of-drug-discovery-at-the-interface-of-biology-chemistry-and-pharmacology-latin-america-and-the-caribbean-20221113/
Educational	Online Manual of Chemoinformatics (in Spanish).	This manual contributes to the academic formation of the students, strengthening in students the understanding of basic concepts of Chemoinformatics and seeks to allow them to be able to handle and interpret computational techniques associated with this scientific discipline, frequently used in the discovery, design and development of bioactive compounds.	https://difacquim.gitbook.io/quimioinformatica/
Educational and Research	NAPROC13	The 13C NMR Database of Natural Products is a web application that provides different search tools to know the structure of a compound. In addition, it allows the use of chemoinformatics applications based in JSME-X-2020 by Peter Ertl and Bruno Beinfait, (Novartis), in the design and development of undergraduate student theses in Pharmacy.	https://c13.materia-medica.net/
Meeting online	10th School of molecular modeling in biological systems	Meeting of researchers and students interested in acquiring knowledge in the following topics: protein structure prediction, machine learning methods, molecular modeling applied to drug development, receptor-ligand docking methodologies, molecular dynamics, monte carlo, quantum electronic structure calculations, hybrid quantum/classical calculations, and bioinformatics.	http://www.emmsb.lncc.br/

Table 5

Representative strategic actions to strengthen chemoinformatics in Latin America.

Action	Time frame [a]	Expected contribution or impact
Research: enhance research collaborations within Latin America and between Latin America and other countries.	Medium/long	Advance significantly in research projects (basic and applied research), joining efforts with multi-sites; increase the diversity of points of view.
Research: continue working on topics relevant to geographical regions and global issues. Foresee upcoming challenges, emerging needs, areas of opportunity, and trends.	Short/medium	Working on top of regional and global issues and challenges will contribute to maintaining Latin American countries on top.
Research: Partner with industry-driven foundations and research institutes (e.g., tres cantos Open Lab foundation).	Medium/long	Get involved in basic research projects that can be translated into practical applications.
Research: Partner with software companies (e.g.,Pharmacelera).	Short/medium	Collaborating in the development of chemoinformatics tools such as molecular docking programs
Education: Promote the university research seedbeds in undergraduate programs at Latin American universities.	Short/medium	This will enhance the interest in research in undergraduate students, allowing them to generate skills in the use of computational tools in the faculties of computer science, biology, chemistry, and pharmacy. This would foster their interest in furthering studies in basic chemoinformatics and its applications in academia, industry, and emerging companies in AI.
Education: Establish and sustain postgraduate programs (e.g., Masters's) in Latin American universities.	Long	It will help to train highly educated specialists in chemoinformatics that will join the industry, academia, and start-up companies.
Education: continue and increase the number of quality conferences, seminars, and schools. These could be in-person, hybrid, or virtual events.	Short/medium	Scientific conferences help increase the interactions among students and research scientists in the area and the world; will contribute to being up to date with recent developments and identifying future needs.
Funding: attract attention, collaboration, and partnerships with industries (pharmaceutical, food, materials, etc.) to fund research programs with translational applications.	Medium/long	Expand funding sources beyond traditional government support will help to increase the research funding itself and will promote work on applied projects. Students participating in research projects will gain experience working on industry-driven projects and help them to work on industry positions.
Funding: continue and increase the connections with foundations, not-for-profit institutions, publishers, and industries in the region and worldwide that help support educational and research programs.	Medium	*Idem* (previous entry)
General: Create a Latin American chemoinformatics society that guides and oversees the development.	Long	Help to sustain the collective efforts in terms of research, education, and funding.
Dissemination and collaborations: continue to enhance the presence and participation of Latin American Scientists in international conferences organized in other parts of the world.	Short/medium	Active participation of Latin American students and scientists will favor the generation and/or consolidation of collaborative networks; will also help the awareness of the scientific community worldwide of the research done in Latin America.
Dissemination and collaborations: Continue and enhance the participation of Latin American scientists in specialized international committees and editorial boards of peer-reviewed journals.	Medium	Strengthen the connection and networks between Latin America with other countries, be "ambassadors" that share the needs and progress in the Latin American region.
Chemoinformatics community-wide blind challenges.	Long	With the progress in the generation of databases of natural products, biological activities and physicochemical properties of compounds in the region, it is intended to democratize reliable data so that the scientific community around the world can validate computational tools, propose new predictive models that can be critically evaluated through blind challenges.

[a] Approximate timeframe to start the implementation and development of the task. In all cases, it is expected that the action sustains for several years. Short-term: 1 year; medium-term: 1–5 years; long-term: 5–10 years.

interrelationships. A manual of the discipline is even presented, with which an integration of a greater number of scientists who approach this discipline is generated. Conferences and webinars are very illustrative of the applicability that is being given to chemoinformatics, allowing interactions with experts, who can guide towards the areas of focus and interest of the viewer. Since many webinars were recorded and are freely available, they are also useful resources for teaching and training. At the same time, they are excellent references to witness and keep track of the evolution of chemoinformatics during the next few years. Although a few examples are listed, the conferences, workshops and resources in Table 4 point to the efforts of several countries to develop chemoinformatics in Latin America.

Due to the recent COVID-19 pandemic, the online and hybrid formats could represent an advantage to increase the number of national, regional, and/or international conferences. Of course, such online or hybrid conferences would be added to the full in-person meetings that also provide key advantages through the direct interaction between scientists and students.

5. Future directions

In this section, we summarize the author's perspective on the expected progress and evolution of chemoinformatics in Latin America during the next few years. Based on the discussion of the preceding Sec-

tions and the authors' perspective, we outline a concise action plan to strengthen the development of chemoinformatics in this geographical region and worldwide. Table 5 summarizes the action, the expected period required to implement it, and its anticipated contribution or impact. The list of tasks is not exhaustive but aims to stimulate the analysis of these and other strategic actions that can take place. It is expected that the action tasks can be sustained for several years.

From the action tasks summarized in Table 5, we would like to point out the need to strengthen the education of undergrad and graduate students so that highly trained professionals in Latin America join start-ups, pharmaceutical and other companies, academic groups in universities or start and develop their own companies. Start-up (AI) companies based in Latin America could stablish solid collaborations with medium-to-large industries following the example of successful partnerships happening in other countries, for example, Exscientia (https://www.exscientia.ai/)-Sanofi; Recursion (https://www.recursion.com/)-Roche and Genentech; Benevolent AI (https://www.benevolent.com/)-AstraZeneca [33]. Notably, there are scientists who graduated from Latin American universities currently employed by Recursion and Benevolent AI.

Scientific meetings and conferences favor the exchange of knowledge and technology transfer based on computational tools that can be incorporated into universities and institutes, allowing for the following in the academic environment: the participation of undergraduate students in scientific initiation sessions and their incorporation into uni-

versity research hotbeds with groups are already well established. In any case, face-to-face (or hybrid) meetings that are also beneficial can be conducted and financed through collaborations within Latin America. These interactions lead us to the generation of conglomerates that can develop multi-center projects in the Latin American region.

This will favor the interests of financing agencies (public or private) in financing collaborative projects. The above, together with the advantages of online and live conferences and webinars, make it increasingly easier to hold meetings or conferences that have the advantage of saving time and reducing the cost of organization and attendance. It is also beneficial to have the continued convenience of being involved in editorial boards of peer-reviewed publications. Currently, several Latin Americans are taking an active part in the editorial boards of journals such as the Journal of Chemoinformatics, Journal of Chemical Information and Modeling, eLife, to name a few examples. Also, a current trend of some scientific publishers is to include early career researchers in the review board, e.g., Protein Science, which opens opportunities for scientists in Latin America to take part in editorial boards where our representation is underrepresented.

In several of the action points proposed in Table 5, it is crucial to expedite the paperwork needed to formalize confidential agreements with industry. The overabundance of bureaucracy frequently saps industry and other institutions' motivation to collaborate with academic organizations on scientific achievements. Of course, these points require a greater effort that includes the participation of high management and governments. In all, research, education, and funding are among the main pillars that will help to sustain and improve Chemoinformatics and related disciplines in Latin America. The research includes dissemination through peer-reviewed publications, participation in international scientific conferences, and creation and strengthening of collaboration networks within Latin America and with other countries.

6. Summary conclusions

This is a collective effort to discuss the status, progress, and challenges of chemoinformatics in Latin America by experts working on this geographical region. Although there are several other Latin American research groups working on chemoinformatics that are not included in this manuscript, we consider that this is the first effort to team up and discuss the subject that is loosely based on the "four W´s": when the publications and applications have been disclosed (for this work we focused on the period 2010–2022); who (e.g., Latin American countries working on chemoinformatics along with network collaborations within the continent and the rest of the world); what applications of chemoinformatics are being published; and where (we analyzed the journals, online databases, scientific conferences, and forums where the research is being disseminated). We concluded that like "bioinformatics," there is a need to establish a unique term for a well-defined and independent discipline. It is also concluded that, overall, Latin America is the fourth geographical region that contributes to chemoinformatics worldwide after Europe, Asia, and North America. By far, Brazil is the Latin American country with the strongest contributions, followed by Mexico. In the last twelve years, the most common cheminformatics applications with other disciplines have been molecular modeling and drug discovery that, in general are common applications of chemoinformatic in other geographical regions. Also, QSAR and chemical space are common topics published along with chemoinformatics. Collectively, Molecular Informatics and Journal of Chemoinformatics, followed by the Journal of Chemical Information and Modeling, are the most frequent journals where Latin American groups publish their papers. The same conclusion can be drawn from the scientific journals where researchers from other countries publish their chemoinformatic research papers.

We also conclude that there is a strong collaboration network within Latin America, the strongest being between Brazil and Mexico this far.

Also, there are extensive collaboration networks between Latin American countries and other geographical regions. Of note, collaborations between Brazil and USA and Brazil and Germany. Other Latin American countries have frequent interactions with the USA, Germany, Spain, and Italy. While Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay are among the countries that have been developing important open-access resources for research, education, and training on chemoinformatics and closely related disciplines. It is also noticeable that more than ten compound databases have been developed with an emphasis on natural product collections. Development of compound databases, including natural products, is also a common trend in other countries outside Latin America.

Continuing doing research, consolidating current collaborations, and expanding interaction networks within research groups within Latin America and other countries is an important challenge and will have a strong impact on developing chemoinformatics in the region. A strong challenge in Latin America to make academic-industry partnerships is the paramount of paperwork and lengthy time required to sign and implement legal cooperation agreements. We think that speeding up the process will benefit both parties. The vast territory and richness of natural resources in Latin America offer an attractive opportunity to spread start-up and AI-focused companies that can foster the further development of chemoinformatics and related computational disciplines in the area. Democratization and open science would foster the connection of research groups and students at various levels within Latin America and across the world. Although open science favors science development worldwide, this is quite relevant in Latin America, where the funding level to acquire and maintain commercial software and computational resources is limited. Dissemination through peer-reviewed papers, active participation at scientific conferences (organized within Latin America and elsewhere), and training and education will play an important role to continue expanding chemoinformatics and related disciplines in Latin America. Indeed, formal and comprehensive courses on chemoinformatics and related disciplines have been introduced in Latin American universities in the past 3–5 years. This short time sharply contrasts with other geographical regions, such as Europe or the United States, where chemoinformatics courses and formal programs have been ongoing for several decades. It is anticipated that the continued and sustained growth of research will attract funding and additional partnership between Latin American universities with industry and other public or private organizations. Overall, we anticipate that chemoinformatics will continue growing in Latin America.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data collected in the paper is a bibliometric analysis, and it is described in the manuscript in detail.

Acknowledgments

Funding

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ailsci.2023.100077.

References

[1] Varnek A, Baskin II. Chemoinformatics as a theoretical chemistry discipline. Mol Inf 2011;30:20–32. doi:10.1002/minf.201000100.

[2] López-López E, Bajorath J, Medina-Franco JL. Informatics for chemistry, biology, and biomedical sciences. J Chem Inf Model 2021;61:26–35. doi:10.1021/acs.jcim.0c01301.

[3] Bajorath J. Understanding chemoinformatics: a unifying approach. Drug Discov Today 2004;9:13–14. doi:10.1016/s1359-6446(04)02916-2.

[4] Gasteiger J, Funatsu K. Chemoinformatics – An important scientific discipline. J Comput Chem 2006;5(2):53–8. doi:10.2477/jccj.5.53.

[5] Gasteiger J. Chemistry in times of artificial intelligence. Chemphyschem 2020;21(20):2233–42. doi:10.1002/cphc.202000518.

[6] Medina-Franco JL, Saldívar-González FI. Cheminformatics to characterize pharmacologically active natural products. Biomolecules 2020:1566. doi:10.3390/biom10111566.

[7] Gasteiger J. Chemoinformatics: achievements and challenges, a personal view. Molecules 2016;21(2):151. doi:10.3390/molecules21020151.

[8] Medina-Franco JL, et al. Bridging informatics and medicinal inorganic chemistry: toward a database of metallodrugs and metallodrug candidates. Drug Discov Today 2022;27(5):1420–30. doi:10.1016/j.drudis.2022.02.021.

[9] Kirchmair J. Molecular informatics in natural products research. Mol Inf 2020;39(11):e2000206. doi:10.1002/minf.202000206.

[10] Naveja JJ, Oviedo-Osornio CI, Trujillo-Minero NN, Medina-Franco JL. Chemoinformatics: a perspective from an academic setting in Latin America. Mol Diversity 2017;22(1):247–58. doi:10.1007/s11030-017-9802-3.

[11] Perianes-Rodriguez A, et al. Constructing bibliometric networks: a comparison between full and fractional counting. J Informetr 2016;10(4):1178–95. doi:10.1016/j.joi.2016.10.006.

[12] González-Medina M, Jesús Naveja J, Sánchez-Cruz N, Medina-Franco JL. Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. RSC Adv 2017;7(85):54153–63. doi:10.1039/C7RA11831G.

[13] Cox PB, Gupta R. Contemporary computational applications and tools in drug discovery. ACS Med Chem Lett 2022;13(7):1016–29. doi:10.1021/acsmedchemlett.1c00662.

[14] Villoutreix BO, et al. One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. Drug Discov Today 2013;18(21–22):1081–9. doi:10.1016/j.drudis.2013.06.013.

[15] Gómez-García A, Medina-Franco JL. Progress and impact of Latin American natural product databases. Biomolecules 2022;12(9):1202. doi:10.3390/biom12091202.

[16] L Medina-Franco J. Towards a unified Latin American natural products database: lANaPD. Future Sci OA 2020;6(8):FSO468. doi:10.2144/fsoa-2020-0068.

[17] Mobley, D.L. The SAMPL challenges. SAMPL. 2023 https://samplchallenges.github.io/(accessed 2023-04-30).

[18] Geballe MT, Guthrie JP. The SAMPL3 blind prediction challenge: transfer energy overview. J Comput-Aided Mol Des 2012;26(5):489–96. doi:10.1007/s10822-012-9568-8.

[19] Mobley DL, Wymer KL, Lim NM, Guthrie JP. Blind prediction of solvation free energies from the SAMPL4 challenge. J Comput-Aided Mol Des 2014;28(3):135–50. doi:10.1007/s10822-014-9718-2.

[20] Işık M, Rustenburg AS, Rizzi A, Gunner MR, Mobley DL, Chodera JD. Overview of the SAMPL6 pK challenge: evaluating small molecule microscopic and macroscopic pK predictions. J Comput-Aided Mol Des 2021;35(2):131–66. doi:10.1007/s10822-020-00362-6.

[21] Bergazin TD, Tielker N, Zhang Y, Mao J, Gunner MR, Francisco K, Ballatore C, Kast SM, Mobley DL. Evaluation of log P, pK, and log D predictions from the SAMPL7 blind challenge. J Comput-Aided Mol Des 2021;35(7):771–802. doi:10.1007/s10822-021-00397-3.

[22] Işık M, Bergazin TD, Fox T, Rizzi A, Chodera JD, Mobley DL. Assessing the accuracy of octanol-water partition coefficient predictions in the SAMPL6 part II log P Challenge. J Comput-Aided Mol Des 2020;34(4):335–70. doi:10.1007/s10822-020-00295-0.

[23] O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: an open chemical toolbox. J Cheminf 2011;3:33. doi:10.1186/1758-2946-3-33.

[24] Barreiro EJ, Rodrigues CR, Albuquerque MG, Sant'Anna CMR, Alencastro RB. modelagem molecular: uma ferramenta para o planejamento racional de fármacos em química medicinal. Quím Nova 1997;20(3):300–10. doi:10.1590/S0100-40421997000300011.

[25] Montanari CA. A química medicinal na próxima década. Quím. Nova 2000;23(1):134–7. doi:10.1590/S0100-40422000000100024.

[26] Góes ACS, Oliveira BVX. Projeto genoma humano: um retrato da construção do conhecimento científico sob a ótica da revista ciência hoje. Ciênc educ 2014;20(3):561–77 (Bauru). doi:10.1590/1516-73132014000300004.

[27] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules 2015;20(7):13384–421. doi:10.3390/molecules200713384.

[28] Tavare, A.; Horta, C.; Kümmerlec, A.E.; Guido, R. A evolução da química medicinal no Brasil: avanços nos 40 anos da sociedade brasileira de química. 2017; 40:694–700 doi:10.21577/0100-4042.20170075.

[29] Amaral AT, Montanari CA. Química medicinal: 25 anos de planejamento racional de fármacos. Quím Nova 2002;25:39–44. doi:10.1590/S0100-40422002000800008.

[30] Sant'Anna CMR. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). Quím Nova 2002;25(3):505–12. doi:10.1590/S0100-40422002000300022.

[31] Verli H, Barreiro EJ. Um paradigma da química medicinal: a flexibilidade dos ligantes e receptores. Quím Nova 2005;28(1):95–102. doi:10.1590/S0100-40422005000100018.

[32] Lima LM. Química medicinal moderna: desafios e contribuição brasileira. Quím Nova 2007;30(6):1456–68. doi:10.1590/S0100-40422007000600015.

[33] King, A. Encoding creativity in drug discovery. Chemistry World. 2023 https://www.chemistryworld.com/news/encoding-creativity-in-drug-discovery/4016866.article (accessed 2023-04-13).

[34] Gori DNP, Alberca LN, Rodriguez S, Alice JI, Llanos MA, Bellera CL, Talevi A. LIDeB tools: a Latin American resource of freely available, open-source cheminformatics apps. Artif Intell Life Sci 2022;2:100049. doi:10.1016/j.ailsci.2022.100049.

[35] Website. 2023 https://naturar.quimica.unlp.edu.ar/es/.

[36] Valli M, dos Santos RN, Figueira LD, Nakajima CH, Castro-Gamboa I, Andricopulo AD, Bolzani VS. Development of a natural products database from the biodiversity of Brazil. J Nat Prod 2013;76(3):439–44. doi:10.1021/np3006875.

[37] Pilon AC, Valli M, Dametto AC, Pinto MEF, Freire RT, Castro-Gamboa I, Andricopulo AD, Bolzani VS. NuBBE: an updated database to uncover chemical and biological information from Brazilian biodiversity. Sci Rep 2017;7(1):7215. doi:10.1038/s41598-017-07451-x.

[38] Scotti MT, Herrera-Acevedo C, Oliveira TB, Costa RPO, Santos SYKO, Rodrigues RP, Scotti L, Da-Costa FB. SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. Molecules 2018;23(1). doi:10.3390/molecules23010103.

[39] Costa RPO, Lucena LF, Silva LMA, Zocolo GJ, Herrera-Acevedo C, Scotti L, Da-Costa FB, Ionov N, Poroikov V, Muratov EN, Scotti MT. The SistematX web portal of natural products: an update. J Chem Inf Model 2021;61(6):2516–22. doi:10.1021/acs.jcim.1c00083.

[40] ZINC catalog UEFS natural products. 2023 http://zinc12.docking.org/catalogs/uefsnp (accessed 2023-03-15.

[41] Zamora, W.J.; Pinheiro, S.; Acuña, D. NAPRORE-CR (NAtural PROducts REpository - Costa Rica), 2023. doi:10.5281/zenodo.7858102.

[42] Unidad de informática del Instituto de Química. 2023 https://uniiquim.iquimica.unam.mx/(accessed 2023-03-15).

[43] Pilón-Jiménez BA, Saldívar-González FI, Díaz-Eufracio BI, Medina-Franco JL. BIO-FACQUIM: a Mexican compound database of natural products. Biomolecules 2019;9(1). doi:10.3390/biom9010031.

[44] Sánchez-Cruz N, Pilón-Jiménez BA, Medina-Franco JL. Functional group and diversity analysis of BIOFACQUIM: a Mexican natural product database. F1000Res 2019;8. doi:10.12688/f1000research.21540.2.

[45] Olmedo DA, González-Medina M, Gupta MP, Medina-Franco JL. Cheminformatic characterization of natural products from Panama. Mol Diversity 2017;21(4):779–89. doi:10.1007/s11030-017-9781-4.

[46] Barazorda-Ccahuana, H.L., Ranilla, L.G., Candia-Puma, M.A., Cárcamo-Rodriguez, E.G., Centeno-Lopez, A.E., Del-Carpio, G.D., Medina-Franco, J.L., Chávez-Fumagalli, M.A. Peru NPDB: the Peruvian natural products database for in silico drug screening. Sci Rep 2023 13, 7577. https://doi.org/10.1038/s41598-023-34729-0

[47] Swamy MK, Kumar A. Phytochemical genomics: plant metabolomics and medicinal plant genomics. Springer Nature; 2023.

[48] Bajorath J, Chávez-Hernández AL, Duran-Frigola M, Fernández-de Gortari E, Gasteiger J, López-López E, Maggiora GM, Medina-Franco JL, Méndez-Lucio O, Mestres J, Miranda-Quintana RA, Oprea TI, Plisson F, Prieto-Martínez FD, Rodríguez-Pérez R, Rondón-Villarreal P, Saldívar-Gonzalez FI, Sánchez-Cruz N, Valli M. Chemoinformatics and artificial intelligence colloquium: progress and challenges in developing bioactive compounds. J Cheminf 2022;14(1):1–12. doi:10.1186/s13321-022-00661-0.

[49] Velez Rueda AJ, et al. Women in bioinformatics & data science - Latin America. MethodsX 2022;9:101907. doi:10.1016/j.mex.2022.101907.

[50] Bannan CC, Burley KH, Chiu M, Shirts MR, Gilson MK, Mobley DL. Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 Challenge. J Comput-Aided Mol Des 2016;30(11):927–44. doi:10.1007/s10822-016-9954-8.

[51] Riquelme M, Vöhringer-Martinez E. SAMPL6 octanol-water partition coefficients from alchemical free energy calculations with MBIS Atomic Charges. J Comput-Aided Mol Des 2020;34(4):327–34. doi:10.1007/s10822-020-00281-6.

[52] Zamora WJ, Pinheiro S, German K, Ràfols C, Curutchet C, Luque FJ. Prediction of the n-octanol/water partition coefficients in the SAMPL6 blind challenge from MST continuum solvation calculations. J Comput-Aided Mol Des 2020;34(4):443–51. doi:10.1007/s10822-019-00262-4.

[53] Viayna A, Pinheiro S, Curutchet C, Luque FJ, Zamora WJ. Prediction of n-octanol/water partition coefficients and acidity constants (pK) in the SAMPL7 blind challenge with the IEFPCM-MST model. J Comput-Aided Mol Des 2021;35(7):803–11. doi:10.1007/s10822-021-00394-6.

[54] Rodriguez SA, Tran JV, Sabatino SJ, Paluch AS. Predicting octanol/water partition coefficients and pKa for the SAMPL7 challenge using the SM12, SM8 and SMD solvation models. J Comput-Aided Mol Des 2022;36(9):687–705. doi:10.1007/s10822-022-00474-1.

[55] Lopez K, Pinheiro S, Zamora WJ. Multiple linear regression models for predicting the n-octanol/water partition coefficients in the SAMPL7 blind challenge. J Comput-Aided Mol Des 2021;35(8):923–31. doi:10.1007/s10822-021-00409-2.