

COMMENT

Open Access



Chemical space as a unifying theme for chemistry

Jean-Louis Reymond^{1*}

Abstract

Chemistry has diversified from a basic understanding of the elements to studying millions of highly diverse molecules and materials, which together are conceptualized as the chemical space. A map of this chemical space where distances represent similarities between compounds can represent the mutual relationships between different subfields of chemistry and help the discipline to be viewed and understood globally.

Aiming to understand our world, natural sciences constantly expand at the endless frontier of knowledge and become increasingly diverse. For chemistry and against Occam's razor, matter is not simply earth, water, air and fire, let alone the hundred or so elements of the periodic table, nor is carbon the essence of the *vis vitalis*. Our field has developed a broad array of experimental methods leading to the discovery and understanding of a very large number of compositional matters ranging from materials and polymers to biomolecules and drugs, accompanied by the creation of many subfields and their specific languages [1].

Cheminformatics arose from the need to enable access to and exploitation of the chemical knowledge accumulating in the scientific and patent literature. Tools were invented to create identifiers for chemical compounds for the purpose of classification and to describe chemical structures in data formats suitable to train statistical models rationalizing the properties of known compounds and possibly predicting new ones [2, 3]. However, cheminformatics remained for many years a hidden tool supporting commercial databases, and most chemists were unaware of its potential value to guide experiments.

Considering that chance favors the prepared mind, combinatorial chemistry was invented with the idea that trial and error should succeed even for difficult cases given enough trials [4]. Methods were developed to synthesize and test as many compounds as possible focusing on numbers and miniaturization [5–9]. This high-throughput screening approach for discovery, although only partly successful, popularized the evidence that discoveries in chemistry can benefit from exploiting very large datasets. In the area of medicinal chemistry, this triggered insights such as Lipinski's rule of five [10], the assembly of open access repositories for compounds [11], and the development of molecule collections for screening [12, 13].

Screening collections were obviously commented as being “astronomically” large, suggesting using the words “chemical space” to describe the ensemble of all chemical matter, known or unknown [14–17]. Thanks to the methods developed in cheminformatics, one can formulate chemical space as a mathematical and usually high-dimensional space where distances represent similarities between molecules or materials [18, 19], and which can be represented in the form of chemical space maps by applying various dimensionality reduction methods [20–26]. In this manner, collections of molecules or materials are conceptualized as lands of opportunities to be explored by informed searches, rather than as haystacks in which to blindly search for needles. Such informed searches can greatly improve the efficiency of new

*Correspondence:

Jean-Louis Reymond
jean-louis.reymond@unibe.ch

¹ Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

discoveries in various chemistry fields such as drug discovery [27–29], chemical synthesis [30, 31], asymmetric catalysis [32, 33], materials [34–39], quantum property predictions [40], or toxicology [41].

When looking across the chemical sciences, the idea of chemical space has recently gained popularity in a very simple sense of using “a chemical space” to refer to a precise subfield of investigation such as a compound series, while ignoring the rest, which is a bit unfortunate. I would argue here that “chemical space” as a concept has the potential to do much better, specifically to unify all chemical sciences under a common roof. This would facilitate communication and the identification of cross-disciplinary opportunities and help chemistry to be viewed and understood globally. To achieve this goal will require to draft a map of chemical space representing all subfields of chemistry and their mutual relationships, not an easy task for which multiple approaches to molecular representation including artificial intelligence might be required [42–48].

Author contribution

JLR conceived and wrote the paper.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 9 December 2024 Accepted: 9 January 2025

Published online: 16 January 2025

References

- Whitesides GM (2015) Reinventing chemistry. *Angew Chem Int Ed* 54(11):3196–3209. <https://doi.org/10.1002/anie.201410884>
- Willett P (2011) Chemoinformatics: a history. *WIREs Comput Mol Sci* 1(1):46–56. <https://doi.org/10.1002/wcms.1>
- Jablonka KM, Schwaller P, Ortega-Guerrero A, Smit B (2024) Leveraging large language models for predictive chemistry. *Nat Mach Intell* 6(2):161–169. <https://doi.org/10.1038/s42256-023-00788-1>
- Furka Á (2022) Forty years of combinatorial technology. *Drug Discov Today* 27(10):103308. <https://doi.org/10.1016/j.drudis.2022.06.008>
- Xiang X-D, Sun X, Briceño G, Lou Y, Wang K-A, Chang H, Wallace-Freedman WG, Chen S-W, Schultz PG (1995) A combinatorial approach to materials discovery. *Science* 268(5218):1738–1740. <https://doi.org/10.1126/science.268.5218.1738>
- Lam KS, Lebl M, Krchnák V (1997) The “one-bead-one-compound” combinatorial library method. *Chem Rev* 97(2):411–448. <https://doi.org/10.1021/cr9600114>
- Nefzi A, Ostresh JM, Houghten RA (1997) The current status of heterocyclic combinatorial libraries. *Chem Rev* 97(2):449–472. <https://doi.org/10.1021/cr960010b>
- Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003) Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* 2(5):369–378
- Peterson AA, Liu DR (2023) Small-molecule discovery through DNA-encoded libraries. *Nat Rev Drug Discov* 22:699–722. <https://doi.org/10.1038/s41573-023-00713-6>
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23(1):3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
- Williams AJ (2008) A perspective of publicly accessible/open-access chemistry databases. *Drug Discov Today* 13(11):495–501. <https://doi.org/10.1016/j.drudis.2008.03.017>
- Tingle BI, Tang KG, Castanon M, Gutierrez JJ, Khurelbaatar M, Dandar-chuluun C, Moroz YS, Irwin JJ (2023) ZINC-22—a free multi-billion-scale database of tangible compounds for ligand discovery. *J Chem Inf Model* 63(4):1166–1176. <https://doi.org/10.1021/acs.jcim.2c01253>
- Neumann A, Marrison L, Klein R (2023) Relevance of the trillion-sized chemical space “eXplore” as a source for drug discovery. *ACS Med Chem Lett* 14(4):466–472. <https://doi.org/10.1021/acsmchemlett.3c00021>
- Kirkpatrick P, Ellis C (2004) Chemical space. *Nature* 432(7019):823–823
- Reymond J-L (2015) The chemical space project. *Acc Chem Res* 48(3):722–730. <https://doi.org/10.1021/ar500432k>
- Warr WA, Nicklaus MC, Nicolaou CA, Rarey M (2022) Exploration of ultralarge compound collections for drug discovery. *J Chem Inf Model* 62(9):2021–2034. <https://doi.org/10.1021/acs.jcim.2c00224>
- Orsi M, Reymond J-L (2024) Navigating a 1E+60 chemical space of peptide/peptoid oligomers. *Mol Inform* e202400186. <https://doi.org/10.1002/minf.202400186>
- Scior T, Bender A, Tresadern G, Medina-Franco JL, Martinez-Mayorga K, Langer T, Cuanalo-Contreras K, Agrafiotis DK (2012) Recognizing pitfalls in virtual screening: a critical review. *J Chem Inf Model* 52(4):867–881. <https://doi.org/10.1021/ci200528d>
- López-Pérez K, Avellaneda-Tamayo JF, Chen L, López-López E, Juárez-Mercado KE, Medina-Franco JL, Miranda-Quintana RA (2024) Molecular similarity: theory, applications, and perspectives. *Artif Intell Chem* 2(2):100077. <https://doi.org/10.1016/j.aichem.2024.100077>
- Oprea TI, Gottfries J (2001) Chemography: the art of navigating in chemical space. *J Comb Chem* 3(2):157–166
- Schuffenhauer A, Ertl P, Roggo S, Wetzel S, Koch MA, Waldmann H (2007) The Scaffold tree—visualization of the Scaffold universe by hierarchical Scaffold classification. *J Chem Inf Model* 47(1):47–58. <https://doi.org/10.1021/ci600338x>
- van Deursen R, Blum LC, Reymond JL (2010) A searchable map of PubChem. *J Chem Inf Model* 50(11):1924–1934
- Awale M, Reymond JL (2015) Similarity mapplot: interactive visualization of the directory of useful decoys and ChEMBL in high dimensional chemical spaces. *J Chem Inf Model* 55(8):1509–1516. <https://doi.org/10.1021/acs.jcim.5b00182>
- Probst D, Reymond J-L (2020) Visualization of very large high-dimensional data sets as minimum spanning trees. *J Cheminformatics* 12(1):12. <https://doi.org/10.1186/s13321-020-0416-x>
- Orsi M, Probst D, Schwaller P, Reymond J-L (2023) Alchemical analysis of FDA approved drugs. *Digit Discov* 2(5):1289–1296. <https://doi.org/10.1039/D3DD00039G>
- Orlov AA, Akhmetshin TN, Horvath D, Marcou G, Varnek A (2024) From high dimensions to human insight: exploring dimensionality reduction for chemical space visualization. *Mol Inform* n/a(n/a):e202400265. <https://doi.org/10.1002/minf.202400265>
- Burgi JJ, Awale M, Boss SD, Schaer T, Marger F, Viveros-Paredes JM, Bertrand S, Gertsch J, Bertrand D, Reymond JL (2014) Discovery of potent positive allosteric modulators of the Alpha3beta2 nicotinic acetylcholine receptor by a chemical space walk in ChEMBL. *ACS Chem Neurosci* 5(5):346–359. <https://doi.org/10.1021/cn4002297>
- Young RJ, Flitsch SL, Grigalunas M, Leeson PD, Quinn RJ, Turner NJ, Waldmann H (2022) The time and place for nature in drug discovery. *JACS Au* 2(11):2400–2416. <https://doi.org/10.1021/jacsau.2c00415>
- Sadybekov AV, Katritch V (2023) Computational approaches streamlining drug discovery. *Nature* 616(7958):673–685. <https://doi.org/10.1038/s41586-023-05905-z>
- Coley CW (2021) Defining and exploring chemical spaces. *Trends Chem* 3(2):133–145. <https://doi.org/10.1016/j.trechm.2020.11.004>

31. Schwaller P, Vaucher AC, Laplaza R, Bunne C, Krause A, Corminboeuf C, Laino T (2022) Machine intelligence for chemical reaction space. *WIREs Comput Mol Sci* 12(5):e1604. <https://doi.org/10.1002/wcms.1604>
32. Wagen CC, McMinn SE, Kwan EE, Jacobsen EN (2022) Screening for generality in asymmetric catalysis. *Nature* 610(7933):680–686. <https://doi.org/10.1038/s41586-022-05263-2>
33. Olen CL, Zahrt AF, Reilly SW, Schultz D, Emerson K, Candito D, Wang X, Strotman NA, Denmark SE (2024) Chemoinformatic catalyst selection methods for the optimization of copper–bis(oxazoline)-mediated, asymmetric, vinylogous mukaiyama aldol reactions. *ACS Catal* 14(4):2642–2655. <https://doi.org/10.1021/acscatal.3c05903>
34. Gorai P, Parilla P, Toberer ES, Stevanović V (2015) Computational exploration of the binary A1B1 chemical space for thermoelectric performance. *Chem Mater* 27(18):6213–6221. <https://doi.org/10.1021/acs.chemmater.5b01179>
35. Cheng CY, Campbell JE, Day GM (2020) Evolutionary chemical space exploration for functional materials: computational organic semiconductor discovery. *Chem Sci* 11(19):4922–4933. <https://doi.org/10.1039/D0SC00554A>
36. Mroz AM, Posligua V, Tarzia A, Wolpert EH, Jelfs KE (2022) Into the unknown: how computation can help explore uncharted material space. *J Am Chem Soc* 144(41):18730–18743. <https://doi.org/10.1021/jacs.2c06833>
37. Tudi A, Li Z, Xie C, Baiheti T, Tikhonov E, Zhang F, Pan S, Yang Z (2024) Functional modules map of unexplored chemical space: guiding the discovery of giant birefringent materials. *Adv Funct Mater* 34(51):2409716. <https://doi.org/10.1002/adfm.202409716>
38. Park H, Onwuli A, Butler KT, Walsh A (2025) Mapping inorganic crystal chemical space. *Faraday Discuss*. <https://doi.org/10.1039/D4FD00063C>
39. Clymo J, Collins CM, Atkinson K, Dyer MS, Gaultois MW, Gusev VV, Rosseinsky MJ, Schewe S (2025) Exploration of chemical space through automated reasoning. *Angew Chem Int Ed* e202417657. <https://doi.org/10.1002/anie.202417657>
40. Huang B, von Lilienfeld OA (2021) Ab initio machine learning in chemical compound space. *Chem Rev* 121(16):10001–10036. <https://doi.org/10.1021/acs.chemrev.0c01303>
41. Samanipour S, Barron LP, van Herwerden D, Praetorius A, Thomas KV, O'Brien JW (2024) Exploring the Chemical space of the exposome: how far have we gone? *JACS Au* 4(7):2412–2425. <https://doi.org/10.1021/jacsau.4c00220>
42. Musil F, Grisafi A, Bartók AP, Ortner C, Csányi G, Ceriotti M (2021) Physics-inspired structural representations for molecules and materials. *Chem Rev* 121(16):9759–9815. <https://doi.org/10.1021/acs.chemrev.1c00021>
43. Wigh DS, Goodman JM, Lapkin AA (2022) A review of molecular representation in the age of machine learning. *WIREs Comput Mol Sci* 12(5):e1603. <https://doi.org/10.1002/wcms.1603>
44. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P (2022) Large-scale chemical language representations capture molecular structure and properties. *Nat Mach Intell* 4(12):1256–1264. <https://doi.org/10.1038/s42256-022-00580-7>
45. Medina-Franco JL, Chávez-Hernández AL, López-López E, Saldívar-González FI (2022) Chemical multiverse: an expanded view of chemical space. *Mol Inform* 41(11):2200116. <https://doi.org/10.1002/minf.202200116>
46. Krenn M, Ai Q, Barthel S, Carson N, Frei A, Frey NC, Friederich P, Gaudin T, Gayle AA, Jablonka KM, Lameiro RF, Lemm D, Lo A, Moosavi SM, Nápoles-Duarte JM, Nigam A, Pollice R, Rajan K, Schatzschneider U, Schwaller P, Skreta M, Smit B, Strieth-Kalthoff F, Sun C, Tom G, von Rudorff GF, Wang A, White AD, Young A, Yu R, Aspuru-Guzik A (2022) SELFIES and the future of molecular string representations. *Patterns* 3(10). <https://doi.org/10.1016/j.patter.2022.100588>
47. Wellawatte GP, Seshadri A, White AD (2022) Model agnostic generation of counterfactual explanations for molecules. *Chem Sci* 13(13):3697–3705. <https://doi.org/10.1039/D1SC05259D>
48. Anstine DM, Isayev O (2023) Generative models as an emerging paradigm in the chemical sciences. *J Am Chem Soc* 145(16):8736–8750. <https://doi.org/10.1021/jacs.2c13467>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.