China's rise in the chemical space and the decline of US influence

Marisol Bermúdez-Montaña,[1,2] Angel Garcia-Chung,[2,3] Peter F. Stadler,[2,4-9] Jürgen Jost[2,7,9] and Guillermo Restrepo[2,4,10]

[1]Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Av. Lago de Guadalupe, 52926, Estado de México, México
[2]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[3]Departamento de Química, Universidad Autónoma Metropolitana, San Rafael Atlixco No. 186, Iztapalapa, Ciudad de México 09340, México
[4]Interdisciplinary Center for Bioinformatics, Universität Leipzig, Leipzig, Germany
[5]Bioinformatics Group, Department of Computer Science, Universität Leipzig, Leipzig, Germany
[6]Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria
[7]The Santa Fe Institute, Santa Fe, NM, USA
[8]Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia
[9]Center for Scalable Data Analytics and Artificial Intelligence, Universität Leipzig, Leipzig, Germany
[10]School of Applied Sciences and Engineering, EAFIT University, Medellín, Colombia

Abstract

At the turn of the 21st century, China has achieved a spectacular surge in the scientific arena. However, little is known about its role in the material core of chemistry, encoded in the chemical space, which spans the discovery of chemicals. Here we show, by analysing the chemical space between 1996 and 2022, that its expansion has been dominated by China ever since 2013. Chinese dominance occurs at different levels, from organic to rare-earth chemistry, but is less dramatic in organometallic chemistry. We also found that Chinese dominance is mainly the product of the country's own efforts, rather than the result of international collaboration. China's surge mainly comes at the expense of the contribution of the US. Interestingly, the US share of the chemical space is more dependent on international collaboration, which mainly occurs with China. We also observe the emergent role of India. We believe these results provide a contemporary account of the geopolitics of the chemical space, which may constitute the basis for future national and international science policies, as well as research and development agendas.

Main

China's remarkable rise in the scientific and technological arena of the 21st century is well-documented, along with the factors contributing to this growth.[1,2] However, limited attention has been given to China's influence on the recent development of chemistry—the most prolific natural science in terms of publications and arguably the most significant in its economic and societal impact.[3] In particular, little is known about China's role in the expansion of the chemical space (CS), which spans all chemicals and reactions reported in the literature.[3] Here we quantify China's contribution to the CS and contrast it with contributions of other leading countries, especially the US. Likewise, we quantify the role of international collaboration upon the expansion of the CS, which has grown at a rate of a bit more than 4% for more than two centuries—a pace maintained in spite of varying economic or social conditions,[4] which include dramatic shifts of the centres of research among countries. After being led by France, Germany and the US,[5] with no effect on its stable growth rate, CS is currently undergoing another shift, from the US to China, while keeping its growth rate.[4] Consequently, the increasing contributions of China to the CS come at the expense of other countries, as we report in this contribution. In particular, it is the US whose share is dramatically decreasing. Meanwhile, the relative contributions of other countries remain relatively steady, although India's share is gradually rising, and Japan's share is decreasing, albeit less dramatically than the shifts observed for China and the US. Interestingly, US contributions to the CS increasingly rely on international collaboration, especially with China, while the Asian country's contributions are mainly the result of its own national efforts. We also found that China is today the leading contributor to the organic chemistry and rare-earth regions of the CS. The former spans the vast majority of the CS and the rare-earth region is a key driver of modern technologies.

The study is based on information retrieved from Reaxys®,[1] Dimensions and OpenAlex databases for the time period 1996-2022 (Methods, Figure S1), which shows China's predominant role in the CS as it covers 41% of the current CS, dwarfing the 11% of the US (Figure 1a). While China grew exponentially during this period, US participation modestly grew until 2007 and afterwards rapidly dropped (Figure 1a). In 2013, China became the overall leading contributor. The only other country whose share was systematically increasing is India, but its growth is modest compared with that of China (Figure 1a).

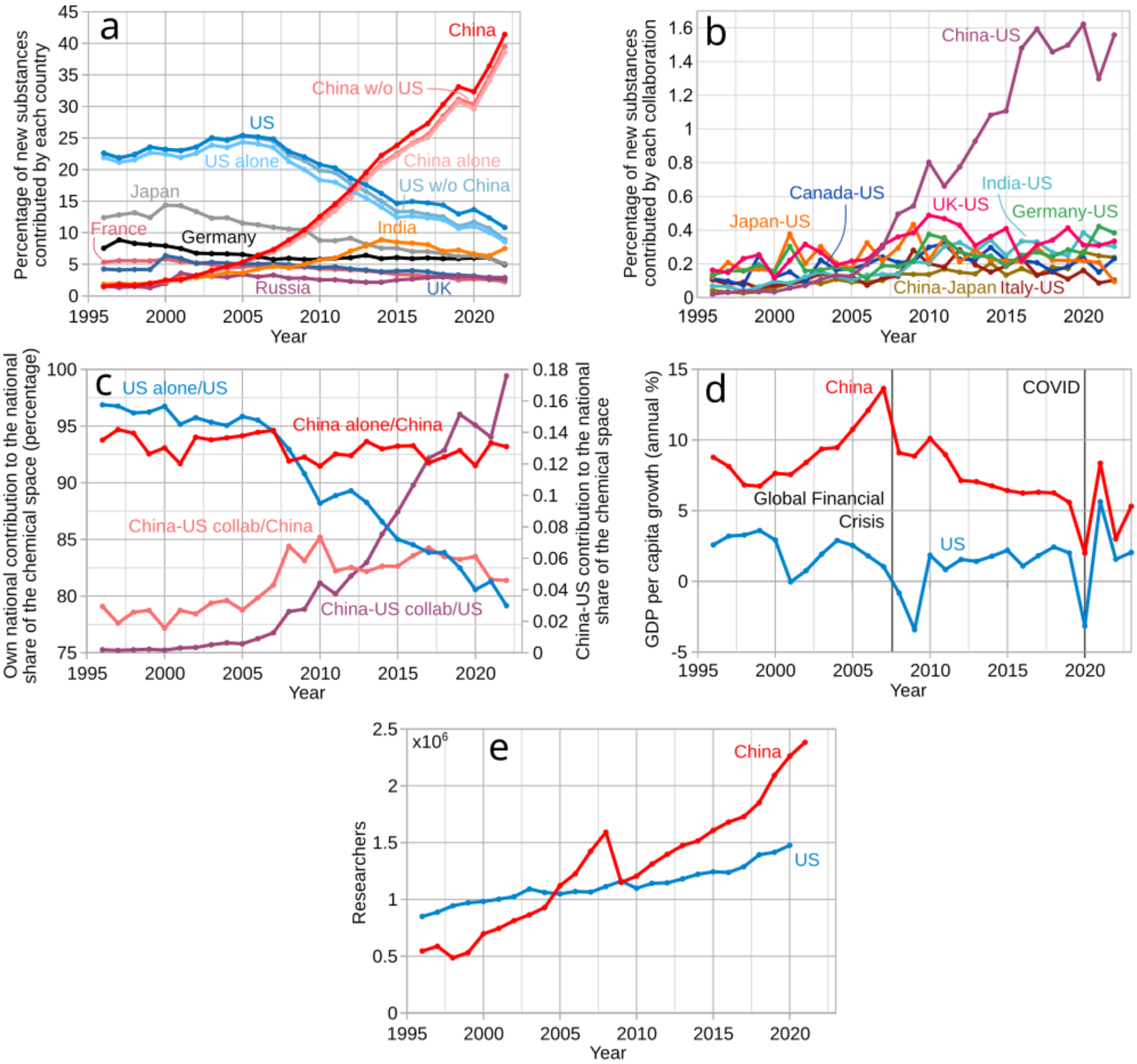

Figure 1. Countrywise expansion of the chemical space (CS). a) Country participation in the growth of the CS. Only the top eight countries are shown (further information in Methods). b) Eight most relevant international collaborations in the CS (Methods). c) Percentage of new substances with participation of country c (c is either China or the US) resulting from China-US collaboration [right axis]. Note that China-US collaboration may include collaborations with other countries (Methods). Percentage of new substances with participation of country c that are reported in papers with no international collaboration [left axis] (Methods). d) Percentage of the annual growth rate of the gross domestic product (GDP) per capita.[6] Two important economic events are highlighted: the 2007/2008 Global Financial Crisis and the COVID pandemic (2020). e) Number of researchers in research and development activities (SI).[6]

National and collaborative efforts to expand the chemical space

The role of international collaboration in the 21st-century scientific and technological Chinese emergence has been documented, especially with the US.[7–9] We found that this collaboration dominates the CS. Nonetheless, its overall contribution is modest, representing less than 2% of all discoveries of chemicals (Figure 1b). China-US collaboration grew over time, peaked in 2017 and ever since has been volatile (Figure 1b). This partnership increased hand in hand with the increment of bilateral agreements enacted during the Obama administration (2009-2017).[10] Likewise, the post-2017 perturbation is presumably related to the 2018 China initiative, championed by the US government.[9]

Notwithstanding the China-US collaboration, over 90% of China's contribution to the CS was achieved by domestic teams operating independently. In contrast, the US has seen a decline in solo contributions, which exceeded 95% between 1996 and 2006 but sharply decreased to less than 80% of its participation in the CS expansion (Figure 1c). The drop of US solo contributions has been counterbalanced by an increase in international collaboration (Figure 1b), in particular with China.

Although the US invests more of its GDP on research and development than China (Figure S2), the big gap on the expansion of the CS between the two countries is expected to widen. China's economy continues to thrive (Figure 1d), since 2005 the Asian country boasts a considerably larger number of researchers than the US (Figure 1e) and the debt is much larger for the US than for China (Figure S3).

2008 and 2020 local and global drops on the expanding chemical space

We observe effects, with certain delay, of the 2007/2008 Global Financial Crisis for the US contribution to the CS and of the COVID-19 pandemic for China (Figure 1a). In the long run, the 2007/2008 crisis did not reduce the expansion of the space, whereas the pandemic did (Figures 2a, S4). Although China experienced a drop of researchers by 2007/2008 (Figure 1e), this did not affect its increasing contributions to the CS, which counterbalanced the declining US contributions (Figure 1a). The pandemic and the rapid lockdowns in China[11] reduced the contributions of the Asian country and, therefore, of the whole CS (Figure 2a). In contrast, the more relaxed US lockdowns[12] did not largely affect the US contributions to the space (Figure 1a).

Chemical space, its subspaces and national contributions to their expansions

We divided the space into its organic, organometallic and rare-earth (REE) subspaces (Figure 2a, Methods). We found that the exponential expansion of the space, as observed since 1800,[4] is largely dominated by the discovery of organic chemicals (Figure 2a). In contrast to the swift growth of the organic subspace, the organometallic and REE subspaces exhibit a distinct trend (Figure 2a). The former slightly increased its expansion until 2013 and ever since has slowed down its unfolding. REE chemical space, in turn, grows slowly.

Organic space: since organic chemistry encompasses the majority of the CS, the overall growth of CS mirrors the expansion of organic chemistry, with China currently contributing about 40% of the organic CS (Figure S5) and China-US collaboration accounting for about 2% of this subspace (Figure S6). Likewise, the 2007/2008 crisis did not affect the organic CS, while the 2020 crisis did (Figures 2a, S5).

Organometallic space: unlike the organic CS, the Chinese leading role in the organometallic CS is very recent and much less pronounced (Figure 2b). While the contributions of the US and Japan decreased rapidly, not only China, but also India and Germany increased their share (Figure 2b). At the collaborative level, China-US partnership remarkably increased in 2008 and kept growing until 2010, when it began to oscillate (Figure 2c). The pandemic marked the change of dominant country on the expansion of the organometallic space, with the US stepping aside and giving place to China as the leading discoverer of organometallics. In fact, it was a dramatic change, as right after 2020 US contributions plummeted and Chinese ones drastically surged (Figure 2b). The pandemic also brought a drop in the China-US collaboration (Figure 2c), which suggests that China's post-pandemic rise was driven by its own national efforts.

REE space: as REEs become technologically and hence economically ever more relevant,[15] its extraction and commercialisation is dominated by only a few countries, with the leading role of China.[16] This has led to international tensions[17] and to set up research programmes motivating innovation in REE chemistry (SI).[18] Our results show that China not only controls the extraction and trade of REEs but also leads the exploration of their chemical space (Figure 2d). Nonetheless, REE space grows slowly (Figure 2a). While earlier the shares were distributed among several countries, China became dominant already by 2004, with a current share of the REE space of about 40%, while most of the top contributors slowed down their discoveries (Figure 2d). The drop is particularly dramatic for Japan. India, in contrast, grew from 5% in 1996 to 10% today. Russia also grew, but less steadily. Currently, however, China's contribution is four times that of the second largest producers of REE compounds: India and Russia (Figure 2d).

China-US collaboration has grown from less than 1% of the share of the REE space before 2008 to 2% in 2020. However, over the following two years, presumably as the result of a blend of pandemic and political factors that include, for instance the American China initiative,[19] the collaboration declined, and it currently represents 0.86% of the REE CS.

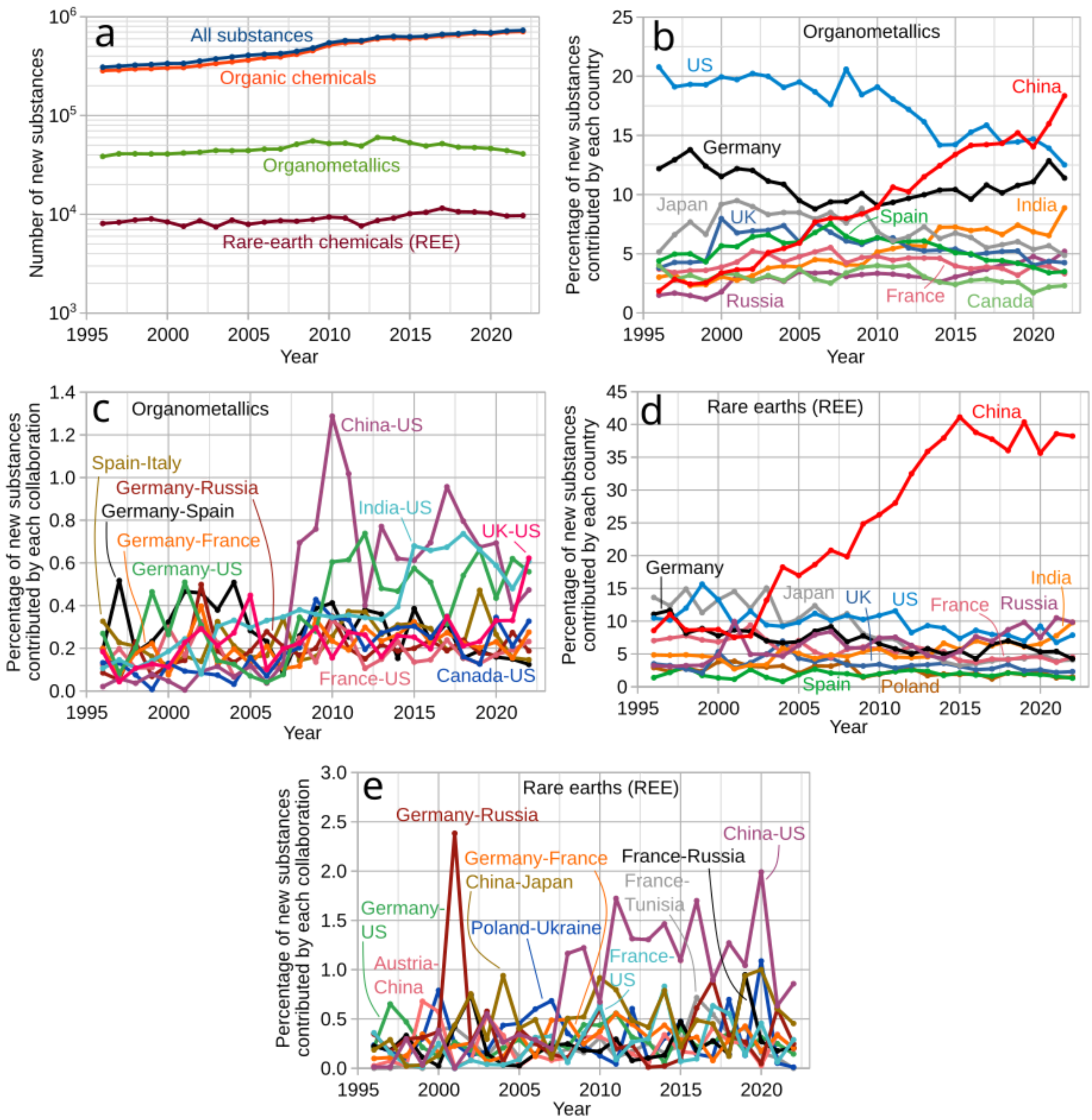

Figure 2. Countries' share on the expansion of subregions of the chemical space (CS). a) Recent expansion of the CS and of three of its subspaces (Methods). b) Countries' participation in the expansion of the organometallic CS and its c) top-10 international collaborations. d) Countries' participation in the expansion of the REE CS and its e) top-10 international collaborations. Criteria for the selection of the substances belonging into these spaces is found in Methods.

Despite the geopolitic dynamics of the CS, the chemical nature of the CS has remained untouched, as the chemical composition of substances of each one of the three analysed chemical subspaces has not been affected (Figure S7).

Methods

Data

As chemical space (CS) entered into a new regime of its expansion by 1980,[4] we retrieved information from Reaxy on substances having molecular formula and publication year between 1980 and 2022 (data dump on 27th July 2023), which led to 17,621,956 substances. To complete the bibliographic information associated with the publications reporting these substances, we retrieved information from Dimensions and OpenAlex (bibliographic databases). By using the DOI of each publication, when available from Reaxys®, we pulled the available bibliographic data from the bibliographic databases. When DOIs were not available in Reaxys®, a combination of the bibliographic information on Reaxys® about authors, publication titles, years, pages and ISSNs, was used as the key to pull data from the bibliographic databases. As Reaxys® data match with bibliographic information was over 80% for the period 1996-2022 (Figure S1), this was the selected period for the current study, which involves 15,725,488 chemicals with author's affiliation countries. In case of an author having multiple affiliations, for the sake of simplicity only the first affiliation was selected. Reaxys® data used in this work is property of RELX Intellectual Properties SA. The code supporting all calculations is available at GitHub and the processed data with the participation of each country in the CS is provided in the SI.

Country assignment to substances

Based on bibliographic information, the annual contribution of each country to the expansion of CS was calculated by considering the share of the number of substances reported by each country in the respective year. For example, if a paper authored by three scientists reports one new substance and two of the authors have affiliation to a Chinese institution and the third one to an American institution, we assign 2/3 of substances to China and 1/3 substances to the US. Thus, if for a given year t, there are n_t new reported chemicals, country i contribution to the CS of year t is given by

$$C_{i,t} = \frac{1}{n_t} \sum_s \frac{n_{i,s,t}}{n_{a,s,t}},$$

where $n_{i,s,t}$ represents the number of authors of country i that published substance s for the first time in year t and $n_{a,s,t}$ accounts for the number of authors participating in the publication of substance s on year t. The complete list of $C_{i,t}$ values is shown in file data/raw/Country_weighted_allSubs.tsv (SI).

Top countries in the expansion of the chemical space

For every year t, countries were ranked based on their $C_{i,t}$ and an average rank was calculated for each country. Top-10 countries contributing more to the expansion of the CS were selected and plotted in Figure 1a.

Regions of the chemical space

Classification of substances in different regions of the CS was based on the presence of particular elements in the formulas of the studied compounds.

Organic: We say that a chemical is an organic substance if carbon is part of its composition.
Organometallic: All chemicals containing carbon and metals. If a compound is a salt, at least one of the ions must contain C and a metal. As metals, we considered the following elements:
Li,Be,Na,Mg,Al,K,Ca,Sc,Ti,V,Cr,Mn,Fe,Co,Ni,Cu,Zn,Ga,Rb,Sr,Y,Zr,Nb,Mo,Tc,Ru,Rh,Pd,Ag,Cd,In,Sn,Cs,Ba,La,Ce,Pr,Nd,Pm,Sm,Eu,Gd,Tb,Dy,Ho,Er,Tm,Yb,Lu,Hf,Ta,W,Re,Os,Ir,Pt,Au,Hg,Tl,Pb,Bi,Po,Ra,Ac,Th,Pa,U,Np,Pu,Am,Cm,Bk,Cf,Es,At,Fr,Fm,Md,No,Lr.
REE: all chemicals containing at least one of the following elements Y, Sc, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu.

To determine the span of the regions of the CS by different chemical elements and compositions in any particular year *t*, for every new compound published in t we extracted its formula. Each formula led to a composition, that is the set of chemical elements arranged in lexicographic order that belong into the formula. Thus, for instance, for compound H_2SO_4, its composition is HOS. The percentage of CS of region x spanned by composition c in year t was determined by $\%CS_{x,t}(c) = ($ # new formulas with composition c reported in t/# new formulas in $t) \times 100$. Likewise, the percentage of CS of region x spanned by chemical element e in year t was given by $\%CS_{x,t}(e) = ($# new formulas containing element e reported in t/ # new formulas in $t) \times 100$.

Data availability

Chemistry data used in this work is property of RELX Intellectual Properties SA. Bibliographic data (in part) is property of Digital Science.

Code availability

The code supporting all the calculations is available as a GitHub repository: https://github.com/mbermudezmoTec/China-s-great-leap-forward-in-the-chemical-space/tree/main/scripts

References

(1) *Report Table of Contents - S&E Indicators 2018 | NSF - National Science Foundation*. https://www.nsf.gov/statistics/2018/nsb20181/report (accessed 2024-09-20).
(2) Woolston, C. What China's Leading Position in Natural Sciences Means for Global Research. *Nature* **2023**, *620* (7973), S2–S5. https://doi.org/10.1038/d41586-023-02159-7.
(3) Jost, J.; Restrepo, G. *The Evolution of Chemical Knowledge*; Springer, 2022.
(4) Llanos, E. J.; Leal, W.; Luu, D. H.; Jost, J.; Stadler, P. F.; Restrepo, G. Exploration of the Chemical Space and Its Three Historical Regimes. *Proc. Natl. Acad. Sci.* **2019**, *116* (26), 12660–12665. https://doi.org/10.1073/pnas.1816039116.
(5) Bensaude-Vincent, B.; Simon, J. *Chemistry: The Impure Science*; World Scientific, 2008.
(6) *World Bank Open Data*. World Bank Open Data. https://data.worldbank.org (accessed 2024-10-25).
(7) Hu, A. G. Z. Public Funding and the Ascent of Chinese Science: Evidence from the National Natural Science Foundation of China. *Res. Policy* **2020**, *49* (5), 103983. https://doi.org/10.1016/j.respol.2020.103983.
(8) Zhou, P.; Glänzel, W. In-Depth Analysis on China's International Cooperation in Science. *Scientometrics* **2010**, *82* (3), 597–612. https://doi.org/10.1007/s11192-010-0174-z.
(9) Van Noorden, R. The Number of Researchers with Dual US–China Affiliations Is Falling. *Nature* **2022**, *606* (7913), 235–236. https://doi.org/10.1038/d41586-022-01492-7.
(10) *Assessing U.S.-China relations under the Obama administration*. Brookings. https://www.brookings.edu/articles/assessing-u-s-china-relations-under-the-obama-administration/ (accessed 2024-10-25).
(11) C&EN, H. J., special to; C&EN, H. J., special to. *The outlook is cloudy for China's chemical industry*. https://cendigitalmagazine.acs.org/2024/01/19/the-outlook-is-cloudy-for-chinas-chemical-industry/content.html (accessed 2024-10-15).

(12) Pilkington, E.; McCarthy, T. The Missing Six Weeks: How Trump Failed the Biggest Test of His Life. *The Guardian*. March 28, 2020. https://www.theguardian.com/us-news/2020/mar/28/trump-coronavirus-politics-us-health-disaster (accessed 2024-10-30).
(13) Reisch, S.; Tullo, A. H. Year In Review. *Chemical & Engineering News*. December 22, 2008. https://cen.acs.org/articles/86/i51/Year-Review.html (accessed 2024-10-30).
(14) Tullo, A. H. Ineos to Boost Capacity for Critical Solvent. *Chemical & Engineering News*. April 14, 2022. https://cen.acs.org/business/specialty-chemicals/Ineos-boost-capacity-critical-solvent/100/i13 (accessed 2024-10-30).
(15) Balaram, V. Rare Earth Elements: A Review of Applications, Occurrence, Exploration, Analysis, Recycling, and Environmental Impact. *Geosci. Front.* **2019**, *10* (4), 1285–1303. https://doi.org/10.1016/j.gsf.2018.12.005.
(16) Ball, P. Beyond Rare Earths. *Nat. Mater.* **2024**, *23* (3), 303–303. https://doi.org/10.1038/s41563-024-01829-9.
(17) Kullik, J. The Strategic Significance of Rare Earths for the Economic and Military Security of the West. No. 13.
(18) Hurst, C. *China's Rare Earth Elements Industry: What Can the West Learn?:*; Defense Technical Information Center: Fort Belvoir, VA, 2010. https://doi.org/10.21236/ADA525378.
(19) Benner, K. Justice Dept. to End Trump-Era Initiative to Deter Chinese Threats. *The New York Times*. February 23, 2022. https://www.nytimes.com/2022/02/23/us/politics/china-trump-justice-department.html (accessed 2024-11-29).

Acknowledgements

We thank RELX Intellectual Properties SA for the reaction dataset. M.B.M. acknowledges support from the Alexander von Humboldt Foundation.

Author contributions

M.B.M. produced the code, analysed data and created the Interactive Information. A.G.C. analysed data. M.B.M., A.G.C., P.F.S., J.J. and G.R. discussed results. G.R. conceptualised the project and wrote the paper.

Competing interest declaration

The authors declare no competing interests.

Supplementary information

Extended data

Data for each figure is found in: https://github.com/mbermudezmoTec/China-s-great-leap-forward-in-the-chemical-space/tree/main/data/processed

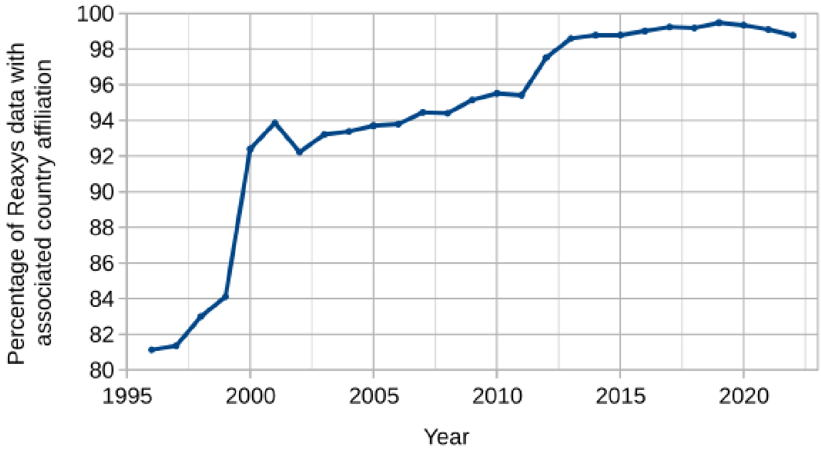

Figure S1. Percentage of Reaxys® data with country label after merging Reaxys® data with Dimensions and OpenAlex databases.

Data on the annual participation of each country to the chemical space (absolute figures)
https://github.com/mbermudezmoTec/China-s-great-leap-forward-in-the-chemical-space/blob/main/data/raw/Country_weighted_allSubs.tsv

Exponential growth of Chinese contribution to the chemical space
Fitting equation (linear regression) for the growth of the percentage of chemical space spanned by China (Figure 1a): $1.33\ e^{0.1372(t-1996)}$, $R^2 = 0.98$.

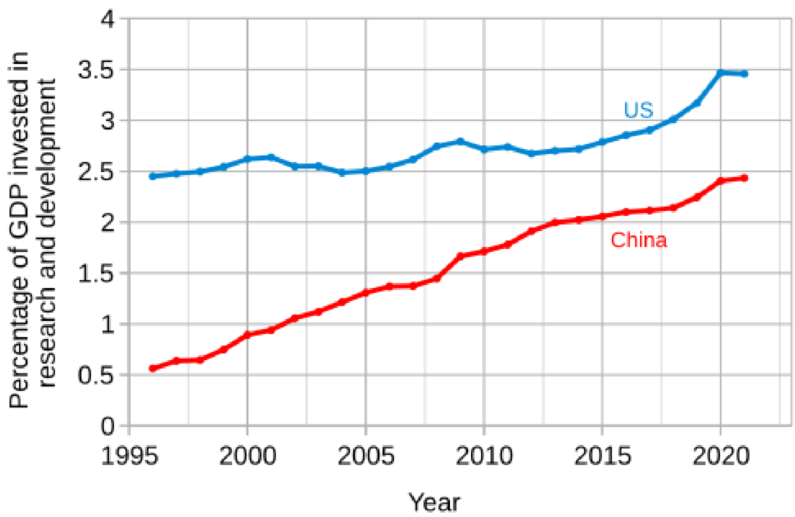

Figure S2. Research and development expenditure (% of GDP) for China and the US. Data from the World Bank.

Number of researchers in China and the US
Figure 1e was obtained by retrieving data from the World Bank on the number of researchers in research and development activities per million people and by multiplying the respective figure of each year by the corresponding country population (according to data from the World Bank).

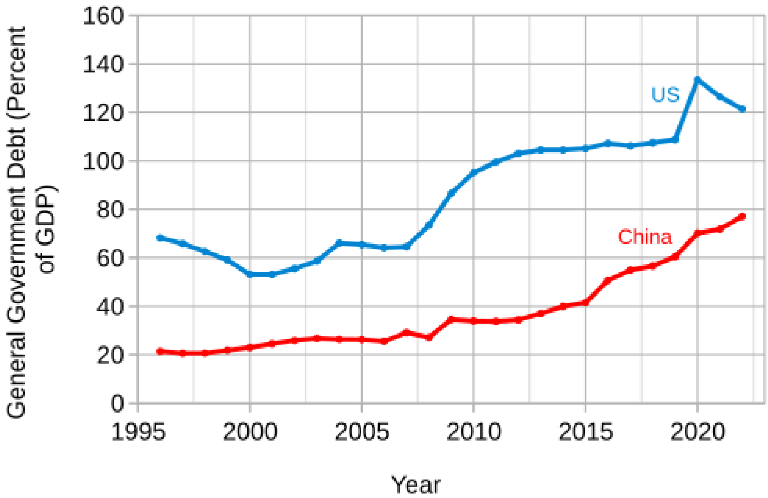

Figure S3. Government debt, as percentage of GDP. Data retrieved from the International Monetary Fund (https://www.imf.org/external/datamapper/GG_DEBT_GDP@GDD/CHN/USA). Retrieved on October 24 2024.

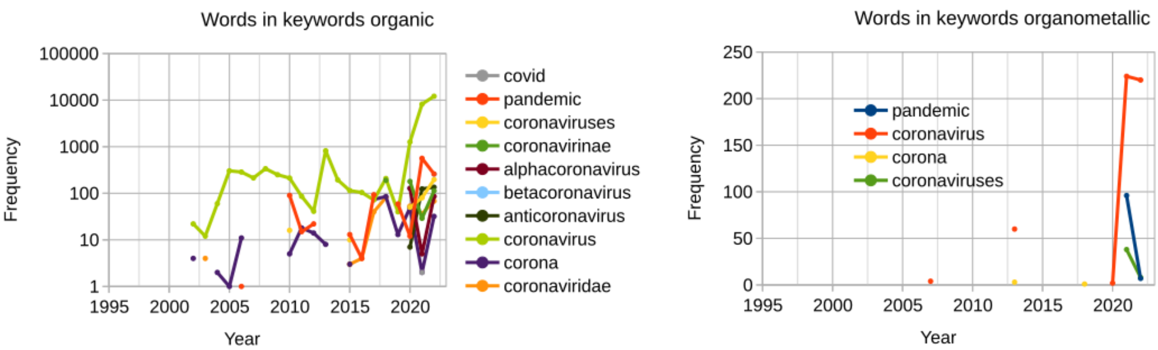

Figure S4. Frequency of COVID-19 related keywords in the chemical literature.

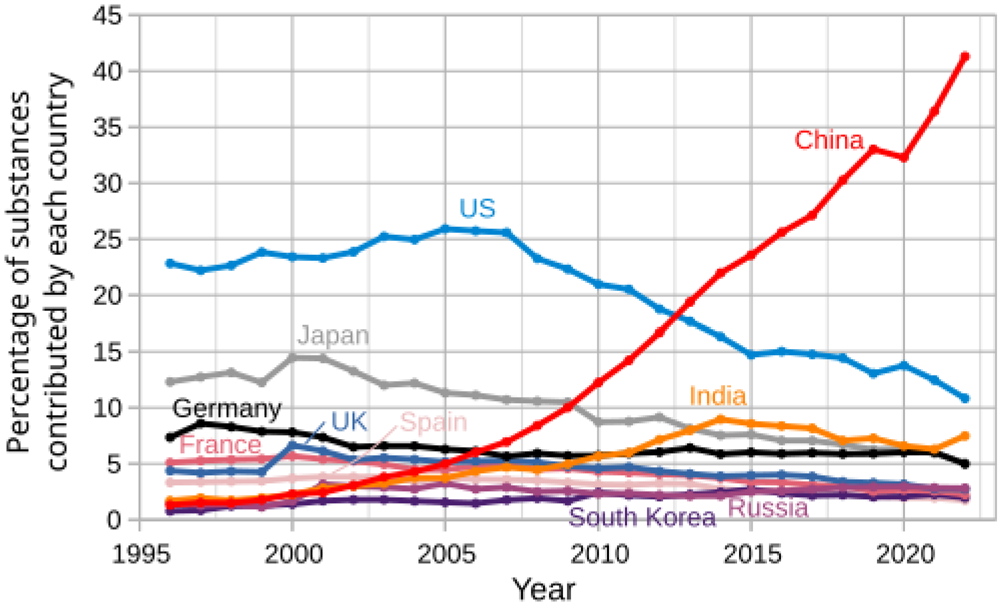

Figure S5. Countries' participation in the expansion of the organic space.

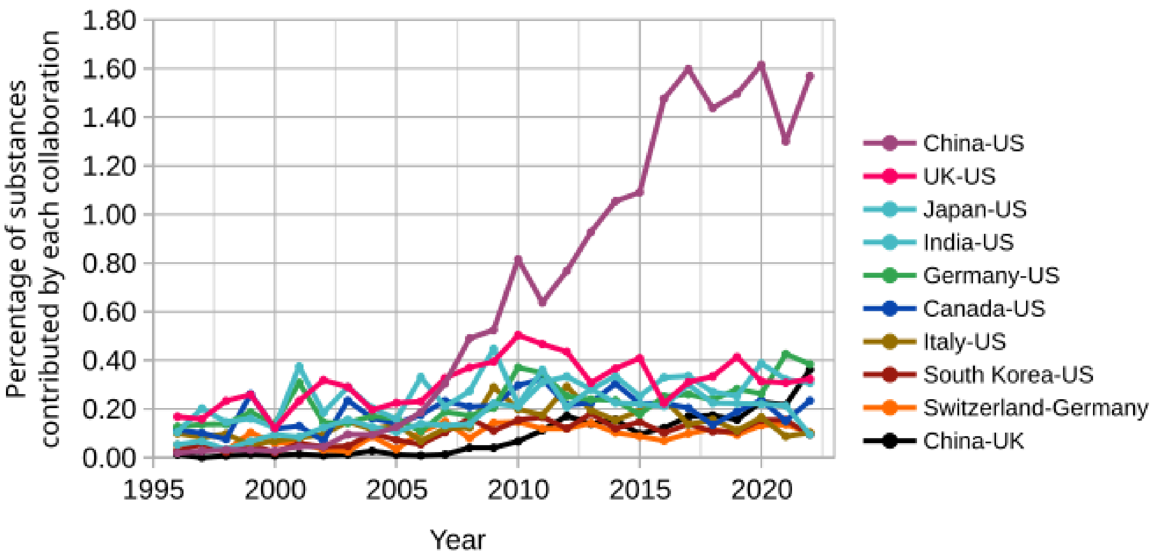

Figure S6. Top-10 international collaborations for expanding the organic chemical space.

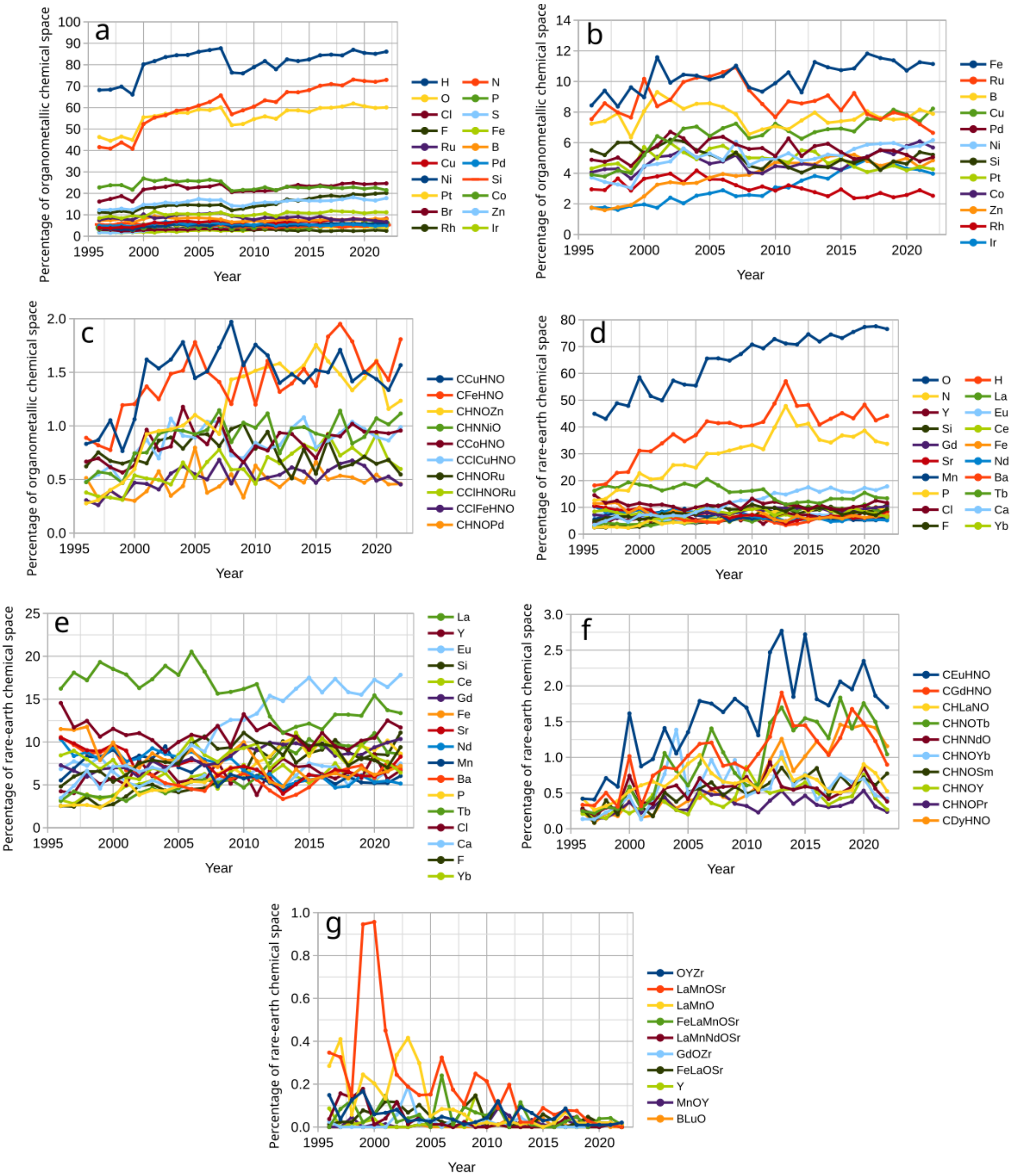

Figure S7. Elemental and compositional spans of the regions of the chemical space (CS). Note that in a report of 2019 [E.J. Llanos, W. Leal, D.H. Luu, J. Jost, P.F. Stadler, G. Restrepo, Exploration of the chemical space and its three historical regimes, *Proc. Natl. Acad. Sci. U.S.A.* 116 (26) 12660-12665, https://doi.org/10.1073/pnas.1816039116 (2019)], besides the ubiquitous presence of C in every organic compound, a few other organogenic elements, mainly H followed by N and O, constitute most of the organic compounds. In terms of compositions, most of the organic CS is concentrated on CHNO compounds. a) Organometallic CS spanned by elements, disregarding the ubiquitous share of C, b) particular detail of the non-organogenic elements that appear most in organometallic compounds. c) Organometallic compositions; a composition corresponds to the set of elements belonging to a substance, thus, HOS is the composition of H_2SO_4. d) REE CS spanned by elements. e) particular detail of the non-organogenic elements that appear most in REE compounds. f) REE organometallic compositions. g) REE inorganic compositions.