Check for updates

POLICY ARTICLE

INTELLECTUAL PROPERTY

Mapping the scope of drug patents

Narrowing patents that claim more than they disclose would open space for rival innovators, not generics

Janet Freilich,[1] Kasper Dinkla,[2] Gerhard Ingmar Meijer,[2] Lisa Larrimore Ouellette[3]

Drug developers have historically sought broad "genus" patents covering millions or billions of chemical structures designed to keep competitors far away from a protected drug. This may no longer be possible. In the United States, many cases over the past decade—including the 2023 Supreme Court decision *Amgen Inc.* v. *Sanofi*—have invalidated broad patents because they fail to disclose sufficient information about the many compounds they cover (*1*). Our analysis of a genus patent dataset indicates that large portions of claimed chemical space are far enough from any disclosed example that standard cheminformatics tools would not predict shared biological activity, suggesting that these patents claim beyond what their data credibly support. The approved drug is structurally central among disclosed examples, suggesting that broad patents function as strategic "moats" around identified leads rather than as placeholders for exploration. Narrowing patents' scope may open space for competing innovators without chilling early-stage research.

Pharmaceutical firms typically protect their investments with multiple patents, each containing distinct legal "claims" that define the boundaries of the patent. The first, foundational patent often includes broad claims in Markush format—named after chemist Eugene Markush who pioneered the technique in 1923—which specify a core molecular scaffold with variable substituent groups at defined positions, allowing a single claim to encompass numerous distinct chemicals (*2*). This patent, filed years before a drug enters clinical trials, creates a protection perimeter that may provide flexibility for lead optimization or block structurally proximate competitors.

In the US and other countries, a patent must disclose sufficient information for a skilled person in the field to make and use the full scope of the claim (*1*). Unlike drug approval, which requires extensive human clinical trials, patent protection requires only a credible basis for asserting that a drug will work, such as in vitro assays or animal studies. Satisfying even this low bar for a large genus can be challenging, as jurisdictions such as the US (the world's largest pharmaceutical market) and Europe require sufficient disclosure over the whole range of a patent claim.

To evaluate how these legal requirements could affect drug development and patenting strategy, we mapped the chemical space defined by Markush claims, the structures specifically disclosed and tested in patents, and the structures of the approved drugs. Our results reveal a striking gap between claimed scope and evidentiary support, with implications for pharmaceutical patents, generic entry, innovator competition, and drug development in the post-*Amgen* era.

MAPPING DRUG PATENTS

To construct our dataset of genus patents associated with approved drugs, we began with patents listed in the US Food and Drug Administration's (FDA's) *Approved Drug Products with Therapeutic Equivalence Evaluations*, where brand-name firms are required to list patents covering a drug under the Hatch-Waxman framework that governs generic entry. From this list of 12,516 patents, we filtered to patents designated as "drug substance" (covering the active ingredient) with priority filing dates on or after 1 January 2000, selecting the earliest priority-date patent for each drug. We manually reviewed the remaining 318 patents (covering 415 drugs), to retain only Markush-format genus compound claims where all disclosed molecules were depicted as chemical structures, yielding 73 patents, each covering a distinct drug.

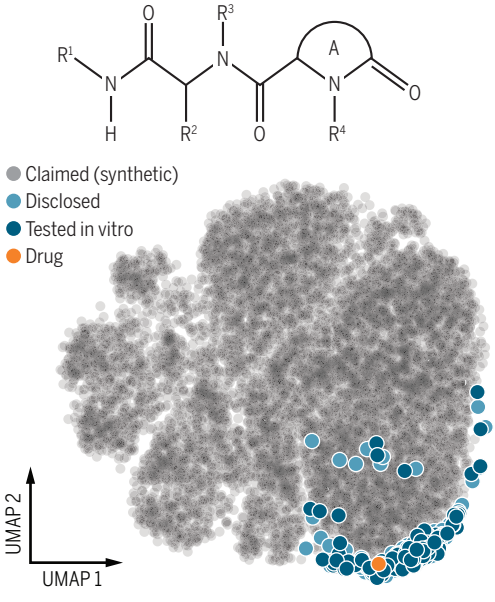
Drug patents often cover a genus of millions or billions of related structures and specifically disclose a small subset as exemplars. We extracted all structures specifically disclosed in the 73 patents as SMILES (Simplified Molecular Input Line Entry System, a text-based notation representing molecular structures) using the PatCID (Patent-extracted Chemical-structure Images database for Discovery) open-access database of chemical structures in patent documents (*3*). We manually labeled the evidentiary support in the patent for each molecule as (i) disclosed only (listed without additional information); (ii) synthesized; (iii) tested in vitro; or (iv) tested in vivo. We excluded structures appearing solely as comparators, background information, intermediates, or reactants, leaving 13,005 disclosed structures across the 73 patents (median 100 per patent, interquartile range 45 to 234), of which 11,209 were synthesized (median 79 per patent, interquartile range 36 to 189), 5985 were tested in vitro (median 53 per patent, interquartile range 12 to 158, for the 61 patents that included in vitro tests), and 114 were tested in vivo (median 3 per patent, interquartile range 1 to 6, for the 24 patents that included in vivo tests).

This approach has limitations. We do not capture patents with molecules in forms other than drawn structures, including proteins. We also do not capture patents outside of the US FDA's registry, including drugs in clinical trials that are not approved.

We then turn to the scope of the patent's Markush claim, which is legal language that defines the possible structures covered by the patent—often many more structures than those disclosed. As an

Visualizing disclosed structures

Markush structure from Claim 1 of US Patent No. 9,474,779 (top). The drug ivosidenib and the patent's disclosed structures, including those tested in vitro, are mapped using uniform manifold approximation and projection (UMAP) onto the chemical space defined by synthetic structures from that claim (bottom).

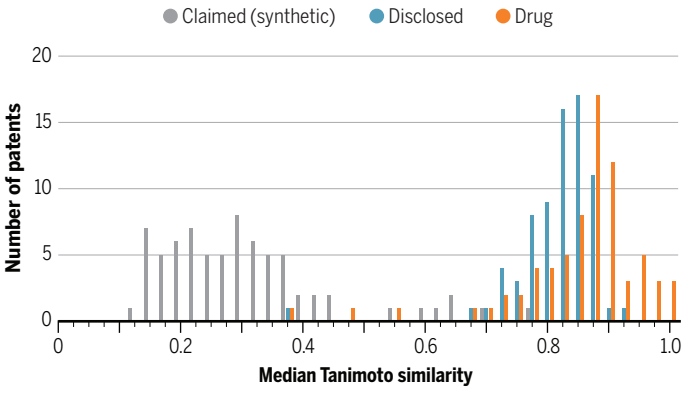


GRAPHIC: C. BICKEL/*SCIENCE*

Downloaded from https://www.science.org at Boston University on October 29, 2026

The gap between claims and disclosures

Histograms show the median (by patent) similarity between each synthetic claimed structure and the nearest disclosed structure [Claimed (synthetic)]; each disclosed structure and its nearest neighbor (Disclosed); and the approved drug and the nearest nondrug structure (Drug).



example, a patent might claim all structures with a specified core molecular scaffold and variable substituent groups (so-called R-groups). For example, a patent might claim "a compound of formula (I) wherein R_1 is C_1–C_6 alkyl or C_3–C_6 cycloalkyl; and R_2 is hydrogen or halogen." The patents in our dataset often define dozens of R-groups with hundreds of members each, creating claims that can encompass billions of structures.

We cannot enumerate billions of structures, so we approximated the chemical space claimed by these patents by using a subset of 10,000 synthetic (computer-generated) structures that are covered by each patent claim. To do this, we manually extracted the scaffold and R-group definitions from Claim 1 of each patent. Using Anthropic's large language model Claude Sonnet 4.5, we generated the 10 simplest members of each R-group [see supplementary materials (SM)]. We then randomly selected a member from each R-group pool, attached these to the scaffold to generate a synthetic structure, and repeated this process until reaching 10,000 structures. Manual review by a chemistry PhD of a random subset of R-groups found errors in 1.1% of R-group members (2 of 187 R-group members across 12 patents).

Although this approach captures only a fraction of possible structures and is biased toward simpler options, it provides a tractable sample for comparing claimed versus disclosed chemical space (see the first figure; and see fig. S4 for a contrasting example). This bias likely even overstates similarity between the sampled claimed chemical space and disclosed structures.

Because molecular structure is linked to function, medicinal chemists routinely compute structural similarity between compounds to predict whether untested compounds will share biological activity with known molecules. For robustness, we use four commonly used similarity metrics: Tanimoto similarity coefficients between molecules encoded using Morgan, MACCS (Molecular Access System), and Daylight fingerprints, and maximum common substructure. Unless otherwise specified, Tanimoto similarity coefficients reported here use 2048-bit Morgan (radius 2) fingerprints and were computed using RDKit (see SM for other similarity measures).

Tanimoto similarity coefficients run from 0 (no similarity) to 1 (largely identical). For Morgan fingerprints, Tanimoto similarity coefficients above 0.3 can be used as a rough permissive relevance cutoff, indicating enough structural overlap that it is reasonable to examine whether compounds might share activity (a conservative benchmark—many studies use thresholds of 0.55 or higher) (*4*–*6*). Although structural similarity does not guarantee functional similarity, and fingerprints are imperfect representations, these metrics are standard tools in drug development.

GAP BETWEEN PATENT CLAIMS AND DISCLOSURES

For each patent, we mapped each synthetic claimed structure to its most similar disclosed structure. Across our sample, the median Tanimoto similarity coefficient between a synthetic claimed structure and its most similar disclosed structure is 0.29 (interquartile range 0.21 to 0.36), below the conservative 0.3 threshold (see the second figure). In other words, even under conservative assumptions about both claim scope and threshold for shared bioactivity, much of the claimed chemical space is so distant from any disclosed structure that shared biological activity would not be predicted. This reinforces the Supreme Court's concern that disclosure of some species may not adequately teach others how to make and use the invention.

By contrast, the disclosed structures form a close cluster: The median Tanimoto similarity coefficient between a disclosed structure and its nearest neighbor is 0.83 (interquartile range 0.79 to 0.85) (see the second figure). The approved drug covered by the patent is part of an even tighter cluster of disclosed structures: The median Tanimoto similarity coefficient between the approved drug and its nearest neighbor is 0.87 (interquartile range 0.82 to 0.90).

Disclosure requirements are legal standards, not quantitative ones; however our data suggest that many of these patents may be vulnerable to invalidation because disclosure requirements are not met. First, the patents disclose only a fraction of the chemical space they claim. The US Federal Circuit has analogized overly broad genus claims to "a plot of land [where] the disclosed species only abide in a corner of the genus" [(*7*), page 1300]. Our quantitative analysis suggests that many drug patent claims are consistent with this metaphor, with disclosed structures representing only a small, structurally similar, subset of the vast chemical territory claimed. Second, even among disclosed structures, many lack experimental evidence of biological activity. Scholars have questioned whether disclosing a chemical structure alone—without testing—satisfies patent law's requirement that applicants provide credible evidence that their invention works (*8, 9*). Finally, claims must also be "definite:" A claim so broad and structurally diverse that a skilled person in the field cannot understand its boundaries is invalid.

GENERIC ENTRY REMAINS BLOCKED

A common assumption in policy debates is that invalidating broad genus claims on the basis of precedents like *Amgen* v. *Sanofi* will accelerate generic entry. Our data suggest otherwise. Pharmaceutical patents often include both the broad genus claims described above and narrower claims to just one molecule that is discussed in depth in the patent. Even if broad claims are invalidated, if an approved drug itself is adequately disclosed in a patent, a narrow claim to that compound blocks generic entry and is less vulnerable to invalidation.

Although a patent can claim a drug without disclosing it, in our sample, 66 of the 73 patents (90%) explicitly disclosed the approved drug. Each of the seven undisclosed drugs is covered by later-priority patents that disclose the drug and also create barriers to generic entry. Additionally, the median Tanimoto similarity coefficient between these undisclosed drugs and the nearest disclosed structure was 0.68 (interquartile range 0.51 to 0.77). Although the *Amgen* decision will not broadly accelerate generic entry, specific patent types may be more at risk. It is possible that patents that do not disclose the drug are filed earlier in the drug development process (*10*); suggestively, none of the seven patents have in vivo experiments, compared to 24 (36%) of the 66 patents that do disclose the approved drug.

NARROWING THE MOAT, NOT IMPEDING RESEARCH

If generic entry is largely unaffected by invalidation of broad genus claims, what is the impact of stricter disclosure standards? There are two competing narratives about the function of broad drug patent claims. One is that patents are filed at a stage when researchers have identified the general structure for their drug but not refined their lead candidate; broad patents stake out an area of chemical space for

Downloaded from https://www.science.org at Boston University on October 29, 2026

GRAPHIC: C. BICKEL/SCIENCE

exploratory work. The other narrative is that researchers have already identified potential leads when the patent is drafted but still want broad coverage—a "moat" around the drug—to prevent competitors from developing similar drugs. These accounts have different policy implications, but distinguishing between them requires quantitative evidence about where the drug sits within the patent's disclosed chemical space.

Our data suggest that the drug patents studied here claim broadly to build moats, not to preserve room for lead selection. Although our selected sample cannot causally distinguish between these possibilities nor reveal patterns for nonapproved drugs where patterns may differ, within our sample, the approved drug covered by the patent is central by several measures, even though the patents were filed a median of 11 years (interquartile range 8 to 14 years) before drug approval. First, as shown, 90% of the patents disclose the structure of the approved drug. Second, we can calculate the centrality of the approved drug among the other disclosed structures. If the patent covered an exploratory genus with lead selection ongoing, the eventually selected drug might be central or peripheral to the set of other disclosed structures. If the patent is intended as a moat, the drug should be central. We found the latter. In 12 of the 66 patents that disclose the drug (18%), that drug is the structural medoid (the disclosed structure whose average distance to all other structures is the smallest) relative to all disclosed structures; the chance of a random molecule being the medoid is 2% (the mean across patents of $1/N$, where N is the number of molecules in the patent). Third, the approved drug is far more likely to be selected for in vivo testing than other disclosed compounds: For the 24 patents with in vivo testing, 12 (50%) test the approved drug in vivo, compared to a 4% chance that any molecule from those patents is tested in vivo. This is consistent with a strategy of deliberately centering the drug, indicative of a research team that already had a strong lead candidate. This pattern is difficult to reconcile with the exploratory account: If broad claims were needed to preserve room for eventual lead selection, the approved drug would not so frequently be central.

IMPLICATIONS FOR COMPETITION

The US Supreme Court's decision in *Amgen* has heightened disclosure standards and reduced patent applicants' ability to obtain broad claims. This will have greater impact on competition among innovator companies developing distinct but related compounds. Broad claims have historically allowed first-movers to exclude competitors from related chemical space, even for compounds that the patent holder never made or tested. If the allowed patent moat contracts to disclosed structures plus modest structural extensions, competing innovators may be able to develop related drugs without infringement liability, potentially increasing later-in-class options available to patients. The structural territory at stake is not hypothetical. Many patent infringement cases involve later-in-class drugs that allegedly fall within an earlier patent claim (*1*), suggesting that the poorly supported regions of genus claims are attempting to reach into the space where therapeutically meaningful competitor drugs are developed.

The role of broad claims in blocking competing innovators rather than generics is further illustrated by the fact that US litigation over whether genus claims have adequate disclosure has almost entirely involved suits where the defendant was a brand-name pharmaceutical firm, not a generic, suggesting that innovators have more to gain by challenging the genus claim (*1*). And in some of these cases, the patent holder failed to develop a drug themselves. Given the breadth of many Markush claims, they may exclude not only "me-too" drugs with minor structural changes, but also those with more substantial differences. This may affect incentives for initial developers by increasing the likelihood of later innovator competition.

PATENTS AND POLICY: MOVING FORWARD

Our findings suggest that stricter enforcement of patent law's disclosure requirements against chemical genus claims will reshape pharmaceutical patent strategy, albeit not in the ways some have anticipated. Generic entry will remain largely blocked; the protective moat against other innovators will narrow.

One natural response will be for applicants to disclose more compounds. But as long as patent law permits "prophetic" examples—compounds described as if synthesized and tested but never actually made, and flagged only by a change of grammatical tense (*11*)—increasing disclosure quantity will not address the underlying problem that disclosures often reflect little meaningful innovation, particularly as generative AI makes it trivial to populate patents with thousands of plausible molecular structures (*10*).

The more fundamental challenge illustrated by our results is the mismatch between information available when drug patent applications are filed and more rigorous information available later in the development process. Patent applications cannot be filed later in the drug development process because of risks that the clinical trials will preclude patenting by destroying the invention's novelty (*12*). The tension between early patent application filing and meaningful disclosure is thus inherent in applying patent law to pharmaceuticals.

One response to circumvent this mismatch is to shift from patents toward regulatory exclusivity (*12*). Unlike patents, which must be filed before clinical value is known, regulatory exclusivity attaches at drug approval and blocks approval of generic versions of the same drug for a set period. Thus, the scope of exclusivity could be calibrated to information that developers actually provide through experimental results and clinical trials. Jurisdictions such as the US and European Union already provide some regulatory exclusivity for new drugs, and decade-old bipartisan US legislation proposed strengthening this framework, with patent holders waiving enforcement rights in exchange for guaranteed FDA-administered exclusivity (*12*). Such a system could provide more tailored incentives for the actual needs of pharmaceutical development, without requiring courts to police boundaries of chemical space that developers never explored.

Patenting a vast array of chemical structures on the basis of data about a small subset may no longer be viable. Our findings suggest that the industry can adapt. Patent holders will no longer be able to fence off vast, untested chemical territory, but will still have narrow claims that protect specific, experimentally supported molecules that they actually develop. For an industry long criticized for using broad patents to block competition and follow-on innovation, this recalibration may ultimately prove healthy. And to the extent the narrower moat proves inadequate, nonpatent frameworks such as regulatory exclusivity offer a more targeted alternative to fill the gap. □

REFERENCES AND NOTES

1. D. Karshtedt *et al.*, *Harvard J. Law Technol.* **35**, 1 (2021).
2. S. Wagner *et al.*, *Res. Policy* **51**, 104597 (2022).
3. L. Morin *et al.*, *Nat. Commun.* **15**, 6532 (2024).
4. S. Jasial *et al.*, *F1000Research*, 591 (2016).
5. D. Stumpfe, J. Bajorath, *J. Med. Chem.* **55**, 2932 (2012).
6. J. Krieger, D. Li, D. Papanikolaou, *Rev. Financ. Stud.* **35**, 636 (2022).
7. *AbbVie Deutschland GmbH & Co.* v. *Janssen Biotech, Inc.*, 759 F.3d 1285 (Fed. Cir. 2014).
8. L. L. Ouellette, *Vanderbilt Law Rev.* **69**, 1825 (2016).
9. J. Freilich, *UC Davis Law Rev.* **53**, 663 (2019).
10. J. Freilich, A. Rai, *Science* **388**, 924 (2025).
11. J. Freilich, L. L. Ouellette, *Science* **364**, 1036 (2019).
12. M. M. Durvasula, L. L. Ouellette, *Yale J. Regul.* **43**, 703 (2026).
13. J. Freilich, K. Dinkla, G. I. Meijer, L. L. Ouellette, Data for "Mapping the scope of drug patents," Zenodo (2026); https://doi.org/10.5281/zenodo.21995432.

ACKNOWLEDGMENTS

The authors thank D. Cunha, K. Zhang, V. Weber, N. Jameshenry, J. DeLuca, and K. Lin for research assistance. Data are available at (*13*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aeh0691

10.1126/science.aeh0691

[1]Boston University School of Law, Boston, MA, USA. [2]IBM Research–Zurich, Rüschlikon, Switzerland. [3]Stanford Law School, Stanford University, Stanford, CA, USA. Email: janetf@bu.edu

Downloaded from https://www.science.org at Boston University on October 29, 2026

Science
AAAS

Mapping the scope of drug patents

Janet Freilich, Kasper Dinkla, Gerhard Ingmar Meijer, and Lisa Larrimore Ouellette

Science **394** (6823), . DOI: 10.1126/science.aeh0691

View the article online
https://www.science.org/doi/10.1126/science.aeh0691
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at Boston University on October 29, 2026

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works