When fragments link: A bibliometric perspective on the development of fragment-based drug discovery

Romasanta, A.K.S.; van der Sijde, P.; Hellsten, I.; Hubbard, R.E.; Keseru, G.M.; van Muilwijk-Koezen, J.; de Esch, I.J.P.

DOI
10.1016/j.drudis.2018.05.004

Publication date
2018

Document Version
Final published version

Published in
Drug discovery today

ELSEVIER

Teaser We study the organizational aspects of the development of fragment-based drug discovery (FBDD), using tools from bibliometrics.

When fragments link: a bibliometric perspective on the development of fragment-based drug discovery

Angelo K.S. Romasanta[1,2], Peter van der Sijde[1], Iina Hellsten[3], Roderick E. Hubbard[4,5], Gyorgy M. Keseru[6], Jacqueline van Muijlwijk-Koezen[2] and Iwan J.P. de Esch[2]

[1] Department of Science, Business & Innovation, Faculty of Science, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[2] Amsterdam Institute of Molecules, Medicines and Systems (AIMMS), Division of Medicinal Chemistry, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands
[3] Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
[4] Vernalis Research, Granta Park, Abington, Cambridge CB21 6GB, UK
[5] York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK
[6] Research Centre for Natural Sciences, Hungarian Academy of Sciences, 1117 Budapest, Magyar Tudósok Körútja 2, P.O. Box 17, Budapest 1525, Hungary

Fragment-based drug discovery (FBDD) is a highly interdisciplinary field, rich in ideas integrated from pharmaceutical sciences, chemistry, biology, and physics, among others. To enrich our understanding of the development of the field, we used bibliometric techniques to analyze 3642 publications in FBDD, complementing accounts by key practitioners. Mapping its core papers, we found the transfer of knowledge from academia to industry. Co-authorship analysis showed that university–industry collaboration has grown over time. Moreover, we show how ideas from other scientific disciplines have been integrated into the FBDD paradigm. Keyword analysis showed that the field is organized into four interconnected practices: library design, fragment screening, computational methods, and optimization. This study highlights the importance of interactions among various individuals and institutions from diverse disciplines in newly emerging scientific fields.

Introduction

FBDD is a widely adopted approach to lead discovery [1,2]. The origin of the field can be traced back to its first demonstration 20 years ago at Abbott Laboratories by Shuker *et al.* [3]. The historical development of FBDD has been discussed as anecdotes, for example during lectures at various conferences [4] and in scientific publications [5,6]. The technical aspects of the approach

Corresponding author: de Esch, Iwan J.P. (i.de.esch@vu.nl)

Angelo K.S. Romasanta is an early-stage researcher at the Marie Curie ITN FragNet based at the Chemistry & Pharmaceutical Sciences Department at VU University Amsterdam. Within the division of Science, Business and Innovation, he is studying how companies in the pharmaceutical industry absorb and apply external knowledge from academia and other firms. He is a graduate of the Erasmus Mundus Master in Chemical Innovation and Regulation Program under the consortium of the University of Barcelona, University of Algarve, and University of Bologna.

Peter van der Sijde is a professor of organization, entrepreneurship & technology in the Department of Science, Business & Innovation at VU University Amsterdam, and has a background in social sciences. His research and teaching is focused on (academic) entrepreneurship and technology transfer.

Iina Hellsten is an associate professor in social sciences at the Amsterdam School of Communication Research (ASCoR) at the University of Amsterdam. She has expertise in communication networks, science and technology studies (STS), and scientometrics.

Jacqueline E. van Muijlwijk-Koezen is a professor in innovation in human health and life sciences at VU University Amsterdam. Her group aims to apply the theory of science education within the context of human health and life sciences. Her research focuses on innovations and didactics in science and education, with special emphasis in pharmaceutical sciences and drug discovery research as embedded within the Amsterdam Institute for Molecules, Medicines and Systems. Research on new teaching concepts and innovative learning approaches lead to new insights that are implemented in the various study programs of the Faculty of Science.

Iwan J P. de Esch is a professor in medicinal chemistry at VU University Amsterdam and head of the Chemistry & Pharmaceutical Sciences Department. His work focuses on two research lines: G-protein-coupled receptors and fragment-based drug discovery (FBDD). He has co-founded several academic spin-out companies, including De Novo Pharmaceuticals, Griffin Discoveries, and IOTA Pharmaceuticals.

https://doi.org/10.1016/j.drudis.2018.05.004

have also been described in key reviews [7–10]. Still, there are insights to be learned by systematically studying how the field has developed. In this paper, we look at the organizational and social aspects of the development of FBDD by analyzing scientific publications that describe new developments in the FBDD field and the references that are provided in those publications. To analyze these records, we used bibliometrics, an approach in information sciences to analyze the relationship among written publications.

Previously, technological breakthroughs resulted from scientists working together at the interface of diverse disciplines, recombining knowledge from various fields [11]. The emergence of FBDD can be seen as various scientific fields coming together, including computational methods, molecular biology, biophysics, and medicinal chemistry. With pharmaceutical sciences being more multidisciplinary and the pharmaceutical industry seeking more collaborations, especially in preclinical development [12–14], it is appealing to investigate the drivers that have made FBDD so successful. With the increasing interest in how organizational factors can enable drug discovery [15], we seek to understand the roles of various groups from industry and academia in the rise of FBDD. By tracing how each publication from academia and industry influenced the field, we can better understand the role of each institution in driving forward new innovations.

Finally, looking at the trends in keyword usage in the publications over time and identifying which keywords usually go together in these publications can lead to a better grasp of how the field is organized. More importantly, by looking at the trends in each keyword over time, we can get a sense of how the focus of FBDD has changed over time and its current direction.

Bibliometric methods

The papers analyzed in this study were downloaded from Thomson-Reuter's Web of Science (WOS). To collect an initial set of papers in the field of FBDD, keywords (Fig. 1) were used. The keyword search generated 3208 papers. To ensure that the keyword 'fragment' was used to refer to the field, we looked at the abstract, title, and keyword fields of the publications and tallied the phrases that co-occurred the most with 'fragment.' We removed the combinations that were unrelated to the field, resulting in a data set of 2781 papers. To verify whether these papers were representative of FBDD, we inspected the data set and found that key publications in the field were not captured by the keywords used in the preliminary search. Examples include Hopkins's paper on ligand efficiency [16] and Hann's paper on molecular complexity [17], because these do not mention any of the keywords used (Fig. 1). Thus, an additional data collection step was performed. Using the first set of papers, we checked for their most-cited references. Analyzing the references, we identified 861 additional publications that were cited at least ten times. This list contained some publications that might not be directly related to FBDD development but nevertheless helped to shape the field. An example is the many references to Berman's publication describing the Protein Data Bank (PDB), which marks the pivotal role of protein structural information in FBDD [18]. Merging these publication lists resulted in a total of 3642 publications that span the years from 1953 to 2016.

To understand the development of FBDD, we set the hallmark publication of Shuker *et al.* [3] in 1996 as the starting point of our analysis. We analyzed papers in the data set that were published

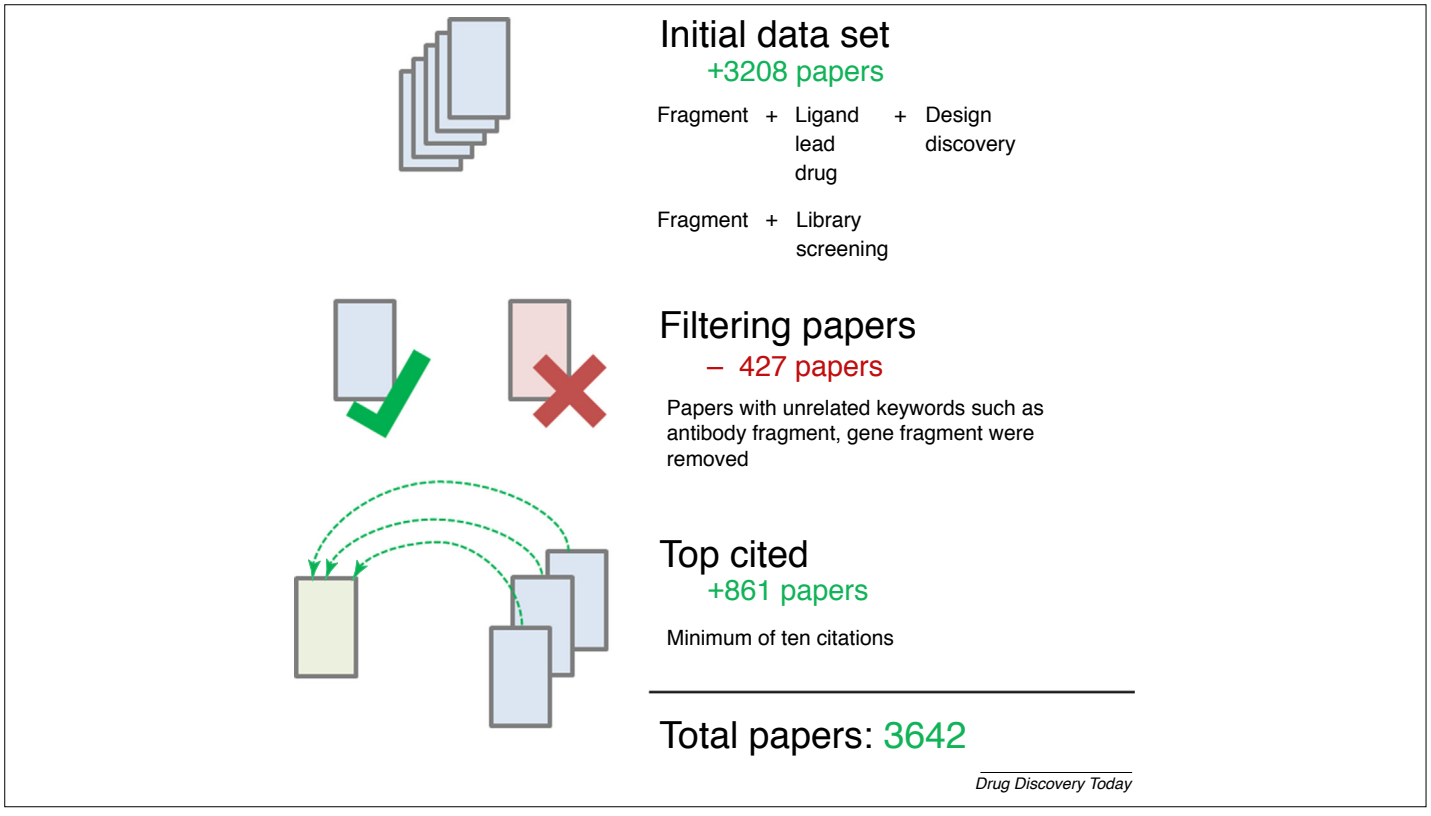

FIGURE 1

Data collection for fragment-based drug discovery (FBDD) publications.

from 1996 to 2016 in 5-year intervals. Various analyses were done to show the role of prior scientific knowledge in adjacent fields and of university–industry collaborations in the development of the field. First, the most-cited articles in our data set of FBDD articles were identified to find the core papers in FBDD. For further analysis, we used the software CitNetExplorer [19] to map the top 100 cited papers, showing the citation relationship among them, allowing us to trace the evolution of knowledge. To study how collaboration between academia and industry has evolved over the past 20 years, we generated co-authorship network maps using the software VosViewer [20]. To uncover the scientific roots of FBDD, we also analyzed the scientific field that the FBDD articles belonged to. Moreover, cluster analysis of keywords was performed. By plotting a network map of keywords that co-occur in publications, we were able to identify the disciplines that researchers study.

Results and discussion

Emergence as 'fragment-based drug discovery'

The fragment-based approach to drug development is widely recognized to have started in 1996,with its first demonstration at Abbott Laboratories [3]. This seminal paper referred to the approach as 'structure–activity relationship by nuclear magnetic resonance' (SAR by NMR), for the first time demonstrating the detection, ranking, and progressing of small and weak-affinity binders.

In our analysis, the FBDD publications in the first 5 years mostly operated under the general umbrella of drug discovery instead of distinguishing themselves as a particular discipline. However, traces of the keywords related to FBDD were present as early as the 1990s, for example in the computational work of Moon and Howe [21] at Upjohn; Rotstein and Murcko [22] at Vertex; and Böhm [23] at BASF. Synonyms, such as 'needles' and 'needle screening', used to describe early applications by Böhm and coworkers, now at F. Hoffmann-La Roche [24], were not adopted by the scientific community because these keywords were used in fewer than five publications in any year. As shown by Fig. 2, It would take a few more years before research in the field would come together in a term such as 'fragment-based drug discovery', which first appeared in the abstract of the 2002 paper by Murray and Verdonk [25]. Even then, the field swopped between the keywords 'lead' and 'drug'. The term 'lead discovery' dominate during the early years, stimulated by influential reviews from researchers at Astex [26–28] during the mid-2000s. Differentiating between the two, the term 'lead' emphasizes the early stage wherein fragments are used (e.g., before pharmacokinetic properties are being considered). By contrast, the term 'drug' can be helpful in that it contextualizes the ultimate goal that fragments aim to achieve, which is to develop drugs.

By 2009, the term 'fragment-based drug discovery' had finally become the top keyword that researchers used to identify the field, whereas 'lead discovery' had lost favor from its peak in 2005, as shown in Fig. 2. As it currently stands, the field is still divided between 'drug discovery' and 'drug design'. Discovery refers more to the finding of a new drug or drug candidate, whereas drug design puts more emphasis on the rational approaches to build the new drug (candidate). As it is, the abbreviation FBDD now appears

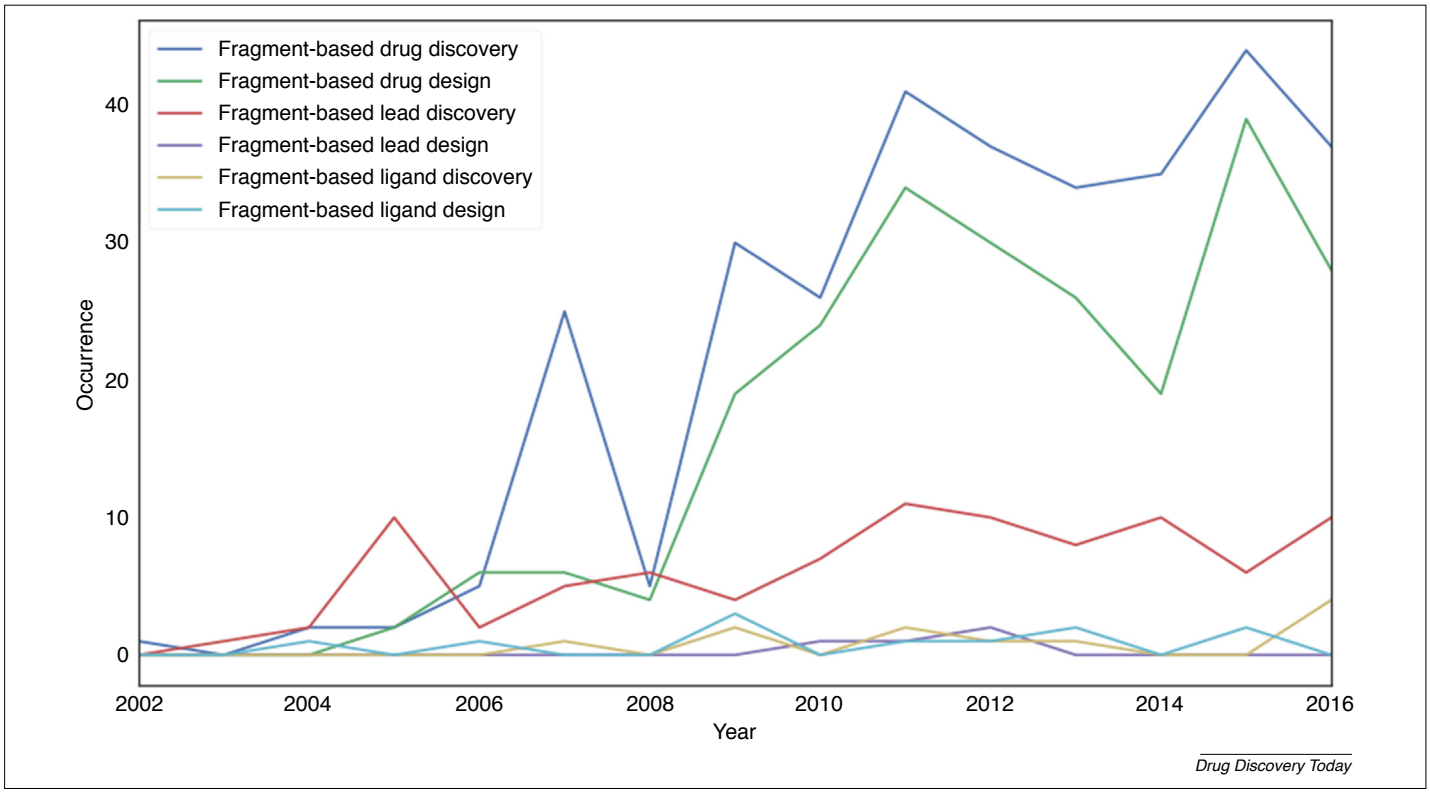

FIGURE 2

Occurrence of fragment-based drug discovery (FBDD) umbrella keywords in the literature. These keywords were chosen because they were the terms used to refer to the field in various important reviews.

TABLE 1

Summary of the FBDD data set from 1996 to 2016

Feature	Timeframe			
	1996–2000	2001–2005	2006–2010	2011–2016
No. of publications	277	496	939	1709
No. of journals	95	143	220	363
No. of researchers (with a minimum of five publications)[a]	102	190	343	389
No. of organizations (with a minimum of ten publications)[a]				
No. of academic institutions	1	4	15	53
No. of SMEs	0	3	3	6
No. of big pharma companies	1	7	7	7

[a] This threshold needed to be set because some firms and researchers co-author publications but do not necessarily practice FBDD.

to be favored over 'fragment-based drug design', being used as much as three times more in 2016 according to the Web of Science, although the different words appear to be used as synonyms.

Aside from the more extensive keyword use, the growth of the field is shown by looking at the increase in number of publications (Table 1). From an initial number of 277 publications in the first 5 years, this increased sixfold to 1709 publications from 2011 to 2016. There has also been an increase in the number of unique institutions, authors, and countries associated with the field, clearly indicating that the approach is being adopted by an increasing number of scientists.

From ideas to application: the role of industry

Clearly, industry has had a pivotal role in developing FBDD. Although the approach was first demonstrated at Abbott Laboratories [3], other organizations in the private sector were instrumental in subsequent FBDD development, in particular by improving emerging technologies and approaches to allow their application in drug discovery. During the first few years of the field, most articles were published by industry researchers. This is noteworthy because an inherent bias towards universities is expected when focusing on scientific publications, given the incentive of academics to publish. Considering that the industry has the opposite incentive of withholding information for competitive advantage [29,30], it emphasizes how influential the industry was in the development of FBDD.

This is also supported by looking at the top institutes in terms of scientific impact, as measured by citations. As seen in Table 2, especially for the first years of FBDD, the industry clearly led the field. Abbott Laboratories dominated during the late 1990s. Astex (founded in 1999 by University of Cambridge professors Tom Blundell and Chris Abell and former head of structural biology and bioinformatics of GlaxoWellcome, Harren Jhoti) led during the following decade. Only in the most-recent 5 years has there be a surge in publications from academics in the top-ten list. Table 2 also shows that biotech companies, such as Astex, Vertex, and Sunesis, have had an important role in establishing the field. However, some prominent biotechs and pharmaceutical companies in FBDD do not show up in this particular analysis because they might have placed less emphasis on authoring scientific publications.

The important role of the private sector in FBDD innovation is also apparent when looking at the top-ten cited papers from our collec-

TABLE 2

Top institutional publishers and their total citations in the field of FBDD over time[a]

Rank	Timeframe							
	1996–2001		2001–2006		2006–2011		2011–2016	
	Institution	No. of citations	Institution	No. of citations	Institution	No. of citations	Institution	No. of citations
1	Abbott Labs	154	Astex	368	Astex	595	Oxford University	368
2	Vertex	77	Abbott Labs	221	Abbott	320	University of Cambridge	348
3	University of California, San Francisco	52	Sunesis	187	University of California, San Francisco	261	GlaxoSmithKline	304
4	Roche	49	Novartis	163	AstraZeneca	249	Astex	232
5	Novartis	43	Pfizer	139	University of Cambridge	216	University of California, San Francisco	156
6	University of Sheffield	35	Scripps Institute	112	Novartis	188	AstraZeneca	139
7	BASF	34	AstraZeneca	93	Scripps Institute	187	Heptares	120
8	University of California, San Diego	29	GlaxoSmithKline	90	GlaxoSmithKline	184	Pfizer	110
9	University of Marburg	28	Sanford Burnham	87	Pfizer	155	Cancer Research UK	105
10	CCDC	26	University of California, San Francisco	85	Vernalis	135	University of Dundee	104

[a] Academic groups are in red.

TABLE 3

The ten most cited papers in the data set of FBDD articles[a]

Rank	Authors	Title	Journal	Affiliation	Year of publication	No. of citations
1	Shuker, S.B., Hajduk, P.J., Meadows, R.P., Fesik, S.W.	Discovering high-affinity ligands for proteins: SAR by NMR	*Science*	Abbott Labs	1996	454
2	Hopkins, A.L., Groom, C.R., Alex, A.	Ligand efficiency: a useful metric for lead selection	*Drug Discovery Today*	Pfizer	2004	403
3	Hajduk, P.J., Greer, J.	A decade of fragment-based drug design: strategic advances and lessons learned	*Nature Reviews Drug Discovery*	Abbott Labs	2007	353
4	Congreve, M., Carr, R., Murray, C., Jhoti, H.	A rule of three for fragment-based lead discovery?	*Drug Discovery Today*	Astex	2003	342
5	Congreve, M., Chessari, G., Tisi, D., Woodhead, A.J.	Recent developments in fragment-based drug discovery	*Journal of Medicinal Chemistry*	Astex	2008	290
6	Hann, M.M., Leach, A.R., Harper, G.	Molecular complexity and its impact on the probability of finding leads for drug discovery	*Journal of Chemical Information and Computer Science*	GlaxoSmithKline	2001	287
7	Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J.	Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings	*Advanced Drug Delivery Reviews*	Pfizer	1997	286
8	Rees, D.C., Congreve, M., Murray, CW., Carr, R	Fragment-based lead discovery	*Nature Reviews Drug Discovery*	Astex	2004	275
9	Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.	The Protein Data Bank	*Nucleic Acids Research*	Rutgers University, National Institute of Standards and Technology, Burnham Institute, University of California, San Diego	2000	257
10	Erlanson, D.A., McDowell, R.S., O'Brien, T.	Fragment-based drug discovery	*Journal of Medicinal Chemistry*	Sunesis	2004	219

[a] Academic groups are in red.

tion of FBDD papers (Table 3). Nine of the top-ten publications were written by industry researchers. The only paper in the top ten by an academic is Berman's publication on the PDB [18], which does not strictly belong to FBDD but is a fundamental resource for drug discovery research in general and for FBDD in particular because many of the hit fragment optimization programs have been guided by protein structural data. Next to some influential reviews, including work from Hajduk (previously Abbvie/Abbott), Congreve (previously working for Astex), Rees (Astex) and Erlanson (at that time working for Sunesis Pharmaceuticals), the conceptual Ligand Efficiency (LE) work of Hopkins and co-workers (at that time working for Pfizer) has had an enormous impact (rank 2, Table 3). LE assesses the contribution of every atom to the affinity of the ligand and is used to select the most promising fragment hits and to guide the growing of fragments into bigger drug-like molecules. Also, the theoretical work of Hann and co-workers at GlaxoSmithKline (rank 6, Table 3) on understanding how molecular complexity impacts hit finding has been influential for FBDD. Among others, this work led to the realization that fragments should be simple and small molecules that can interrogate the binding sites with higher resolution. This also resulted in the guidelines captured in the 'Rule of Three', which define quality fragments. This popular mantra was attractively pitched by Congreve and co-workers (ranked 4) as a variation on Lipinski's Rule of Five (ranked 7, Table 3) that defines the properties of drug-like molecules, the ultimate goal of FBDD efforts.

However, if we look at the top-cited journals in recent years (Table 4), seven out of the ten most-cited publications were authored by academic from 2009. This adoption by academia is also validated by the increase in the share of publishing universities and research institutions in FBDD over the past 5 years. One of the reasons for the adoption by academia is the rise of academic medicinal chemistry and drug discovery groups [12,31]. We can also speculate on the mobility of researchers, including experts from industry who move towards university, setting up academic drug discovery research groups. Given the increase in interest in how researcher mobility affects innovation [32], the impact of this mobility and transfer of knowledge on the development of FBDD will be a topic of future research.

Knowledge transfer: the role of university–industry collaboration

We then explored the list of the top 100-cited articles in FBDD, representing the core papers of FBDD. By creating a citation map of these articles over time, we visualized the evolution in ideas within FBDD and the changing roles of industry and academia in shaping these ideas. Whereas Table 2 and Table 3 reveal the dominating role of the industry in establishing FBDD, the plot in Fig. 3 reveals that ideas and tools developed in academia provided groundwork for the field.

Most of the theoretical grounding of FBDD came with ideas from academia as early as the 1970s. This early influence by

TABLE 4

The ten most cited papers published from 2009 in the data set of FBDD articles[a]

Rank	Authors	Title	Journal	Affiliation	Year of publication	No. of citations
1	Murray, CW., Rees, D.C.	The rise of fragment-based drug discovery	*Nature Chemistry*	Astex	2009	141
2	Chessari, G., Woodhead, A.J.	From fragment to clinical candidate-a historical perspective	*Drug Discovery Today*	Astex	2009	82
3	Murray, C.W., Verdonk, M.L., Rees, D.C.	Experiences in fragment-based drug discovery	*Trends in Pharmacological Sciences*	Astex	2012	76
4	Scott, D.E., Coyne, A.G., Hudson, S.A., Abell, C.	Fragment-based approaches in drug discovery and chemical biology	*Biochemistry*	University of Cambridge	2012	75
5	Murray, C.W., Blundell, T.L.	Structural biology in fragment-based drug design	*Current Opinion In Structural Biology*	University of Cambridge, Astex	2010	70
6	de Kloe, G.E., Bailey, D., Leurs, R., de Esch, I.J.P.	Transforming fragments into candidates: small becomes big in medicinal chemistry	*Drug Discovery Today*	IOTA, Vrije University, Amsterdam	2009	69
7	Filippakopoulos, P., Bradner, J.E. *et al*.	Selective inhibition of BET bromodomains	*Nature*	Dana Farber Cancer Institute, Harvard University, University of Notre Dame, Oxford University	2010	68
8	Baker, M	Fragment-based lead discovery grows up	*Nature Reviews Drug Discovery*		2013	67
9	Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K.	Features and development of Coot	*Acta Crystallographica Section D Biological Crystallography*	Karolinska Institute, University of York, University of California, Santa Cruz, Oxford University	2010	67
10	Chen, Y., Shoichet, B.K.	Molecular docking and ligand specificity in fragment-based inhibitor discovery	*Nature Chemical Biology*	University of California San Francisco	2009	62

[a] Academic groups are in red.

academia can be seen explicitly with the paper of Jencks from Brandeis University [33]. In his paper on the additivity of binding energies, he suggests the idea that large molecules can be considered as a combination of fragments.

On the upper left side of the citation map, several papers authored by academics can also be seen. These are foundational publications about influential drug discovery tools, such as the PDB in 1977 [34], molecular docking approaches by Ferrin and co-workers in 1982 [35], the molecular modeling software CHARMM by Karplus and co-workers in 1983 [36], Goodford's computational procedure for determining energetically favorable binding sites in 1988 [37], and functionality maps of binding sites by Karplus *et al*. in 1991 [38]. Other computational chemistry efforts (e.g., Karplus, Schneider, and Hubbard) to develop *de novo* structure generation and molecular docking software have also made a tremendous impact. Frequently, the developed algorithms use fragment-based approaches as computational 'tricks' to dissect the complication of having to assess and weigh the various properties of bigger, drug-like compounds. During the early 1990s, the technologies and protocols used to determine fragment binding to proteins, using, for example, sensitive biophysical technologies, were not yet available. Computational approaches were also adopted by industry, for example by Schneider at Roche and both Klebe and Böhm at BASF. The latter scientist also contributed to the pioneering needle screening work at Roche that combines *in silico* approaches with biochemical and biophysical screening as an early example of fragment-based approaches in hit finding and lead development. The impact of Abbott Laboratories on developing the applications is not only apparent from the work of Fesik and co-workers with NMR technology, but also from the work of Greer and co-workers, which focuses on discovering ligands using X-ray crystallographic screening. Later, their crystallographic screening method, called CrystaLEAD, was further developed and exploited by influential scientists such as Hubbard (University of York, Vernalis), Rees and Jhoti (Astex), and Abell and Blundell (University of Cambridge, co-founders of Astex). These high-throughput X-ray crystallographic screening efforts were supported by academic activities, such as the development by Cowtan and co-workers of the software COOT, a program that is used to display electron density maps and atomic models.

With academia laying out the foundations of FBDD and Big Pharma first demonstrating the technique in 1996, the road was now ready for the valorization of the field. The next decade of key FBDD publications came almost exclusively from industry. Especially during the early 2000s, smaller structure-based drug discovery companies, such as Astex, Vertex, and Sunesis, come to have an important role. These biotechs specialized in specific FBDD technologies and approaches (e.g., crystal soaking, biochemical assays, and tethering) and perfected them for application in hit finding and lead generation. Fragments provided a way for these compa-

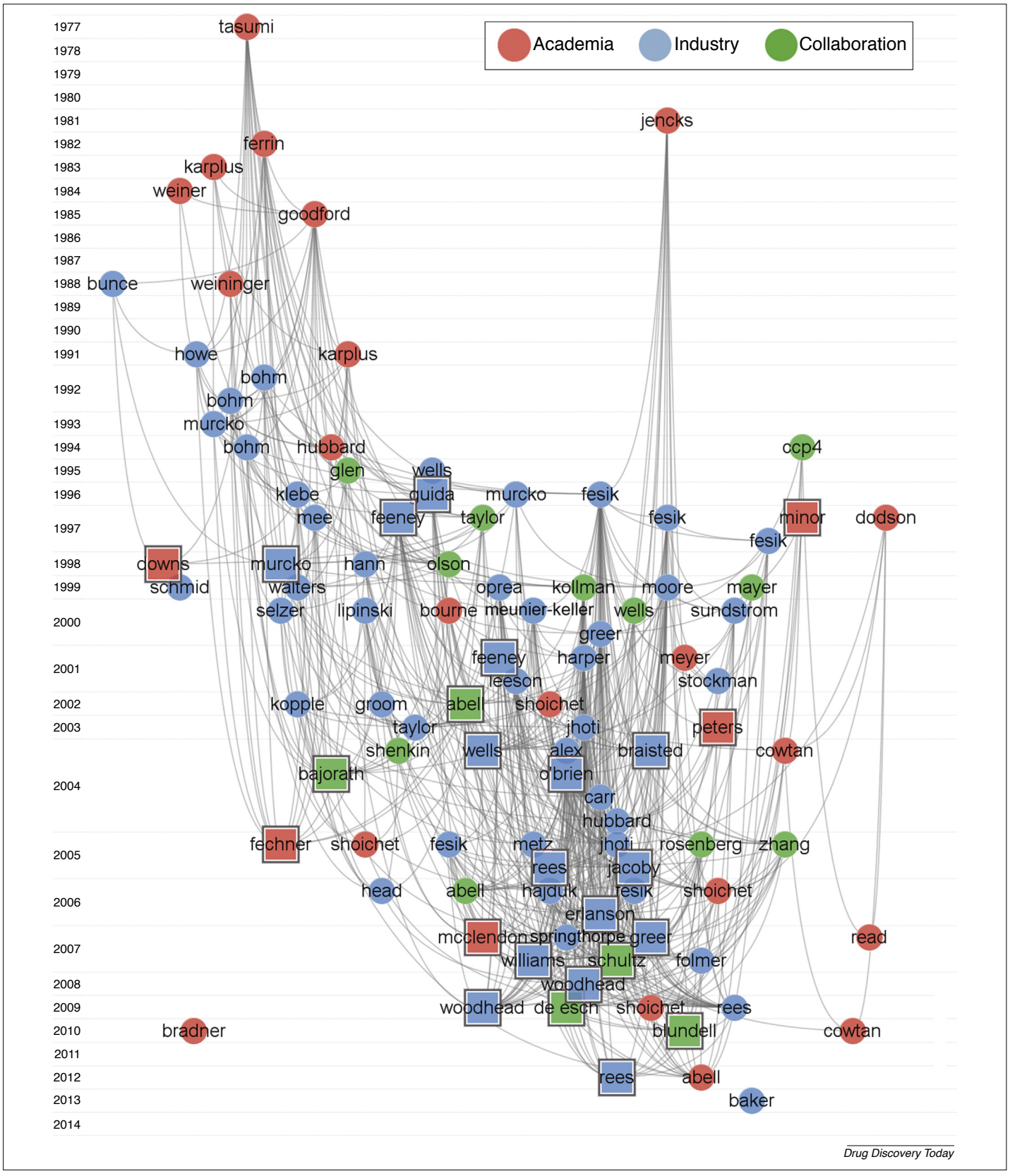

FIGURE 3

Citation map of 100 core papers in fragment-based drug discovery (FBDD). Each paper is labeled by its last author. Colors reflect the affiliation of the authors; squares highlight review articles.

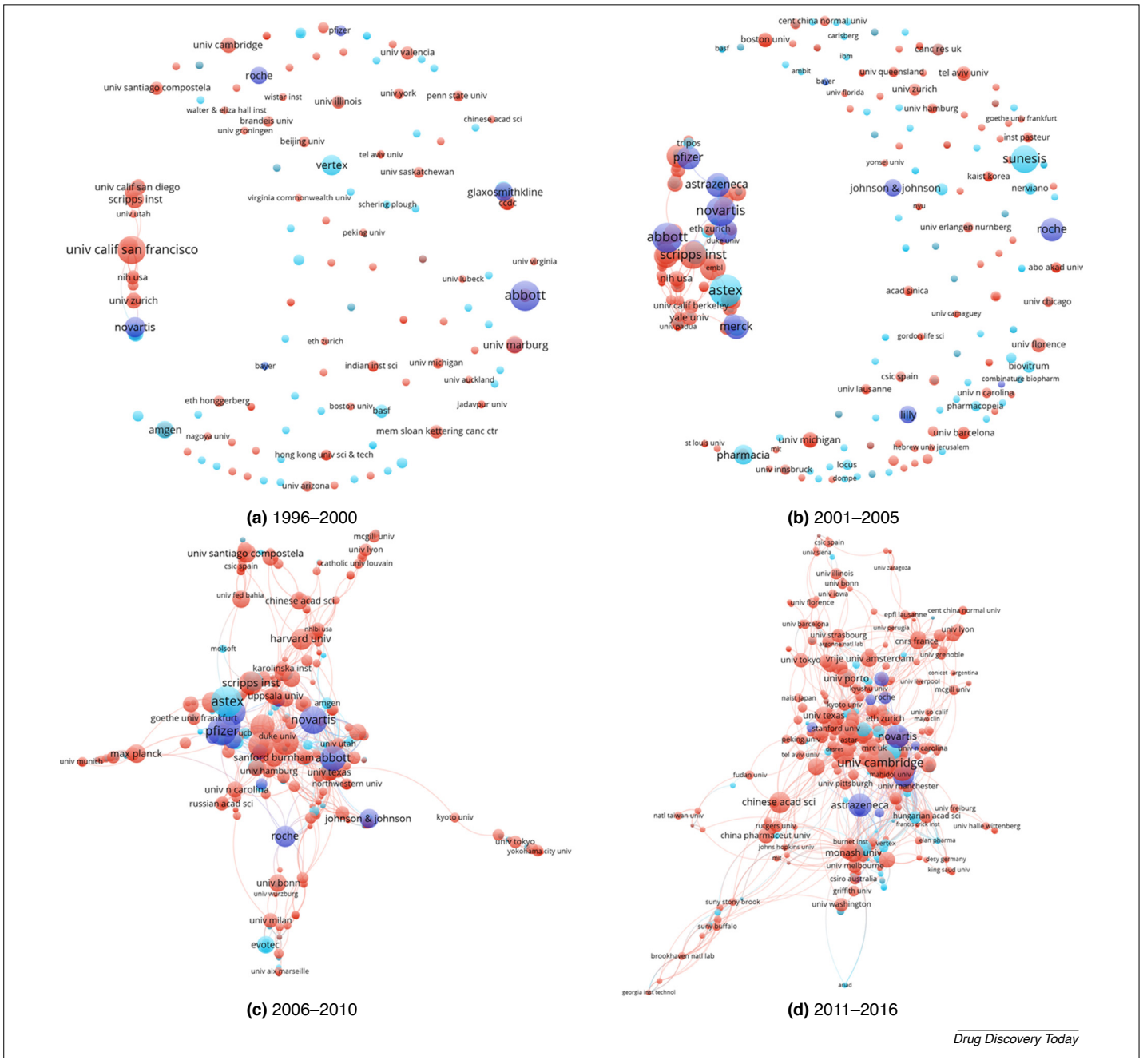

FIGURE 4

Collaboration network map of top 500 publishing institutions in fragment-based drug discovery (FBDD) grouped in 5-year periods from 1996 to 2016. Each node corresponds to an institution. The size reflects the number of publications. Red nodes are from academia, whereas blue nodes are from industry. Dark-blue nodes are from big pharma, whereas light-blue nodes are other industrial firms, including small biotechs and firms from adjacent industries. For the years 2006–2016, the biggest cluster is shown.

nies to obtain hits without the need to invest millions in compound libraries and robotics that are needed for typical high-throughput screening (HTS) approaches [6]. Not all known technologies and FBDD companies appear in this bibliometric analysis, possibly because of their restricted efforts to publish in scientific literature. It is interesting that those companies that do publish make a significant impact when considering collaborations that publish FBDD work (Fig. 4).

During the early years of FBDD, most institutions involved were carrying out research independently. During this period, only a small group of mostly academic institutions were collaborating with a few players in the industry (Fig. 4a). This is seen by the mostly fragmented nodes on the right side of the plot. However, by the early 2000s, a network of university–industry collaborations started to form (Fig. 4b). With the research in FBDD becoming more collaborative, institutions from big

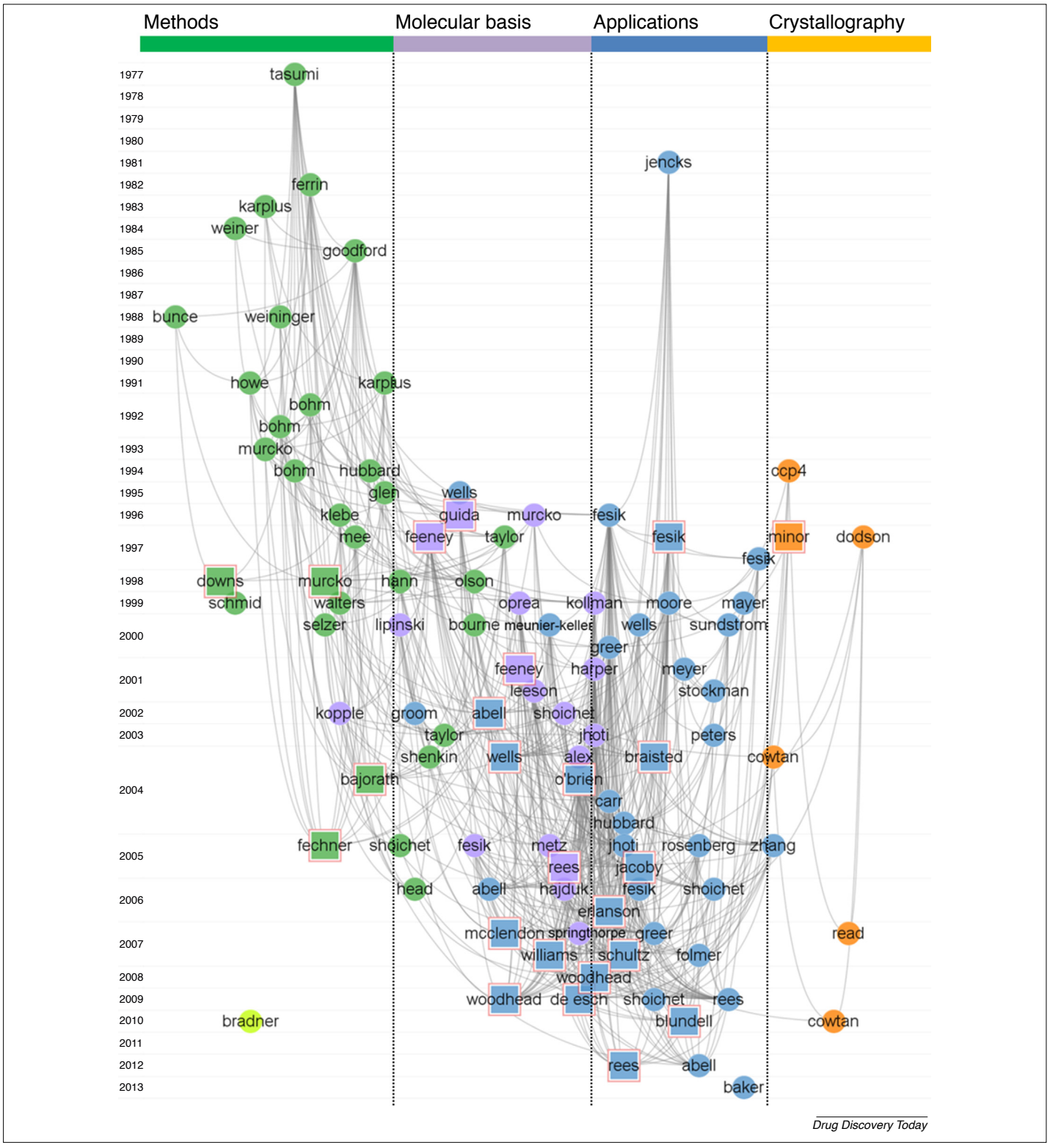

FIGURE 5

Citation map of 100 core papers in fragment-based drug discovery (FBDD). The colors show clustering of papers by similarity.

pharma, spinoffs, and academia co-authored an increasing number of articles. Especially from 2011 to 2016, a greater degree of integration among practicing institutions can be observed. The tight integration shows that FBDD is a high-tech and multidisciplinary research field in which specialists in various research areas collaborate in developing new pharmaceuticals. The development of this field also coincides with the transition of the pharmaceutical landscape in which the big companies outsource

more of their preclinical research [39,40], an important change that appears to have shaped the FBDD field.

Recombining knowledge from other scientific fields

To further understand how FBDD integrates knowledge from various scientific disciplines, we manually classified the previous core papers according to their content and discipline of origin, as shown in Fig. 5.

Before 1996, the scientific groundwork that would eventually be integrated in FBDD came from two separate fronts. As seen on the upper right side of Fig. 5, at one end, we have the work of Jencks, which provided the theoretical rationalization for fragments. At the other end (green cluster of Fig. 5), the previously discussed methodologies that are fundamental in FBDD research can be seen. These computational approaches form an independent branch that used fragment approaches in binding energy calculations and *de novo* structure generation software. As seen in Fig. 5, there is a clear separation between these two branches with no paper citing the two before Fesik's hallmark publication.

Thus, it shows how key the SAR by NMR Science paper by Fesik and co-workers was in jumpstarting the field. As shown in Fig. 5, the paper serves as a hub from which a dense amount of publications branch. The publication by Fesik brought the two separate branches together, explicitly referring to the paper of Jencks while also referring to Bohm's LUDI [23], Hubbard's HOOK [41], and Murcko's GroupBuild [22] at the same time. Thus, the theoretical considerations and the computer-aided drug design capabilities were combined, enabled by the emerging biophysical screening technologies (e.g., NMR) and combined with X-ray crystallography to measure and visualize, respectively low-affinity fragment binding.

We looked at the categories of the journal sources of FBDD papers. Doing so allowed us to see the disciplines that FBDD was building from. In Fig. 6, before 1995, FBDD literature cited articles from the fields of biophysics, biochemistry, molecular biology, and computer science. This signaled that advancements in knowledge in these various fronts was necessary for FBDD to expand. It also gave a clear indication that FBDD is entering mainstream with many publications now appearing in the more applied medicinal chemistry field, whereas during the early years, most papers were in the fields of biochemistry, molecular biology, biophysics, and computational chemistry.

Although this cluster includes the pre-1990s computational techniques described previously, the influence of this cluster extends into the early 2000s, including *de novo* structure generation and docking algorithms, such as Glide [42] in 2004 and the development of frequently used databases, such as ZINC [43] in 2005.

Referring back to Fig. 5, the blue cluster on the right side comprises what are considered to be integral FBDD publications. These include principles and demonstrations of how various biophysical techniques can be used in the paradigm of FBDD. Also included are applications of FBDD to various therapeutic targets (i.e., the actual use in drug discovery [44,45]). Moreover, it also includes 16 key reviews that summarize and integrate knowledge in the field.

We also see a violet cluster at the early stages of FBDD from 1996 to 2002, which describes concepts relating to the molecular basis of the approach. One way of interpreting this is that there are researchers (such as Fesik) who bridge the gap between a new field and established methods, in this case providing the molecular

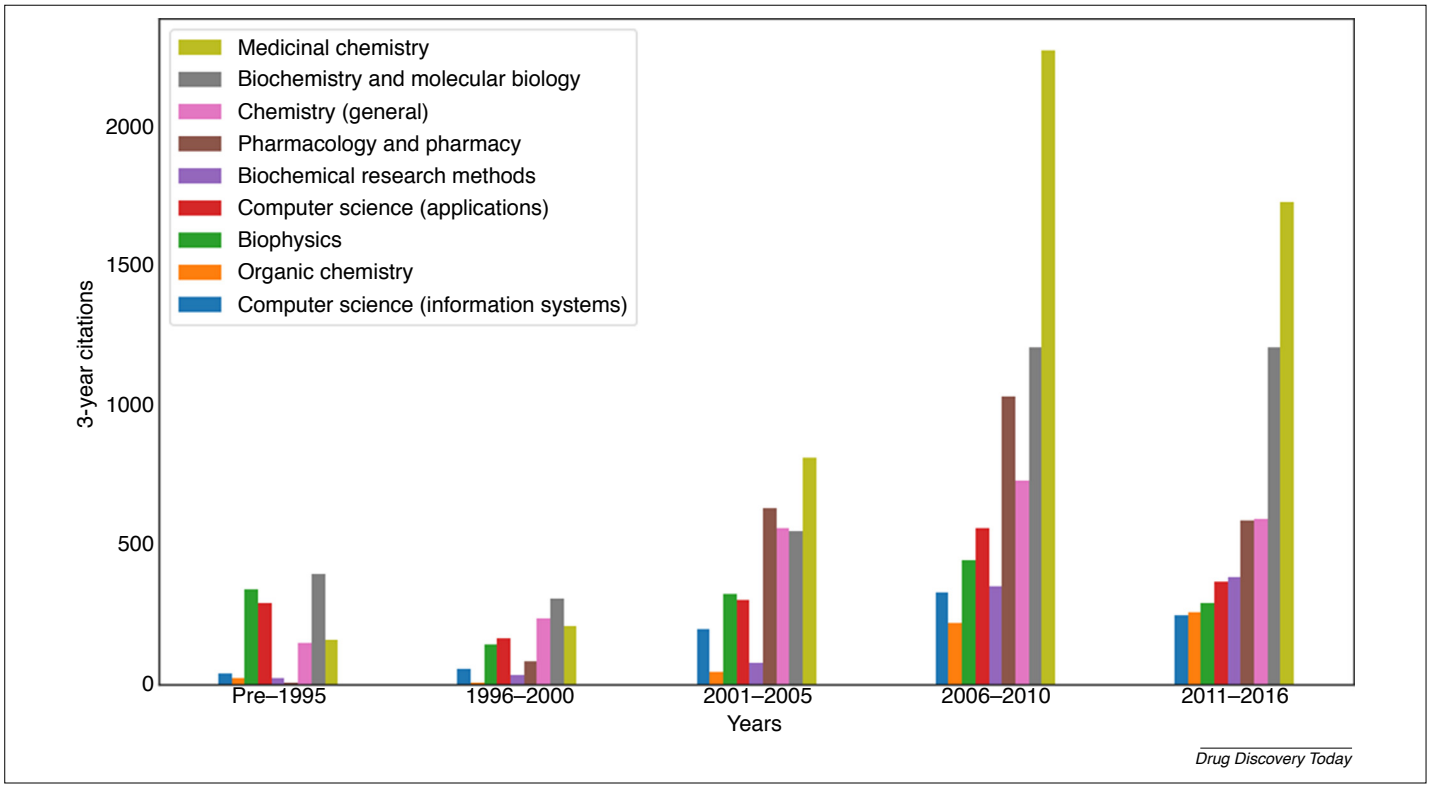

FIGURE 6

Categories of journals over time.

basis of FBDD. By formulating principles from their outsider perspective, these researchers are able to integrate previously unexploited knowledge and technologies into the growing body of FBDD literature. The important role of key opinion leaders can be seen in the central part of the plot, where approximately 2005 scientists, such as Rees, Jhoti, and Abell (Astex and University of Cambridge), Hubbard (University of York and Vernalis), Fesik and Hajduk (Abbott Laboratories), and Erlanson (Sunesis), explicitly integrate the various aspects of FBDD in their publications.

The citation map also shows an orange cluster, which was integrated into FBDD relatively more recently. These are papers in the field of crystallography, such as the CCP4 suite [46] in 1994, Minor's processing of X-ray diffraction data [47], and Dodson's refinement of macromolecular structures [48], both in 1997.

The impact that crystallography would have on FBDD is continuing. By analyzing the keywords used in the abstract and title of the publications in the field, we can get a sense of the methods that catch the interest of practitioners. As seen in Fig. 7, although nuclear magnetic resonance (NMR) was the dominant technique during the first years of FBDD, it has been replaced by X-ray crystallography over the past 5 years. However, this does not perfectly reflect the usage of such techniques in various laboratories, but rather reflects the identifiers that are used by authors to attract their targeted readership.

Currently, the field is organized into four interrelated practices. To determine these four classifications, the top keywords in FBDD was plotted and clustered according to how often they occur together per paper (Fig. 8). Four clusters were detected, corresponding to the four major fields working together in FBDD: molecular biology, (medicinal) chemistry, biophysics, and computational chemistry. These in turn aid the major processes in FBDD, namely, designing the fragment library, screening them using, for example, biophysical techniques, modeling using computational methods, and optimizing the lead. Although the position of the keywords generally indicates the category and interrelatedness of the keywords, the position must be taken with 'a grain of salt' because keywords are more often than not related to the three other dimensions of FBDD.

To see the trends in FBDD over the years, these keywords were colored according to the average year of publication. As shown in Fig. 8, the colors correspond to the average year of keyword occurrence. Interestingly, there is a trend towards the upper left cluster of molecular biology, with more keywords occurring more recently. This is expected because the field has been moving towards applying FBDD, instead of building basic knowledge that comes from the other clusters.

As FBDD matures, it has been applied to more targets. This can be seen by the curious case of the publication by Bradner [49]. Going back to Fig. 5, this publication does not cite the core FBDD literature, yet is cited by many of the recent papers in FBDD. This publication on the inhibition of BET bromodomains has been an area of interest for FBDD researchers in recent years.

Together with other targets, the focus now for FBDD has been its application. The most cited references in recent years (as seen in Table 4) have been reviews showing how an increasing number of leads originating from FBDD are entering clinical trials. It is not only industry using the technique, but also various academic groups. With the growth of FBDD, small has indeed become big in drug discovery [10].

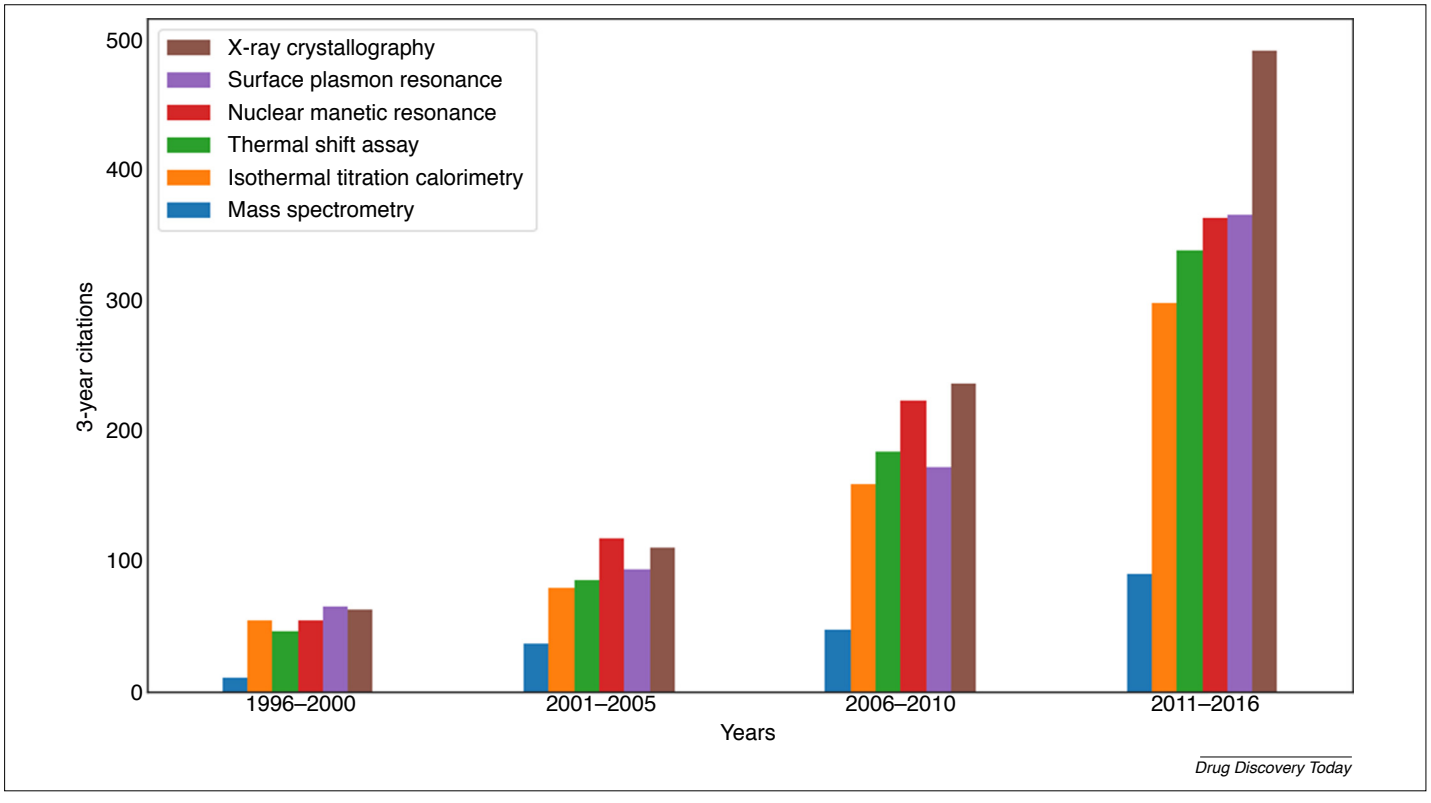

FIGURE 7

Occurrence of various techniques in fragment-based drug discovery (FBDD) papers over time.

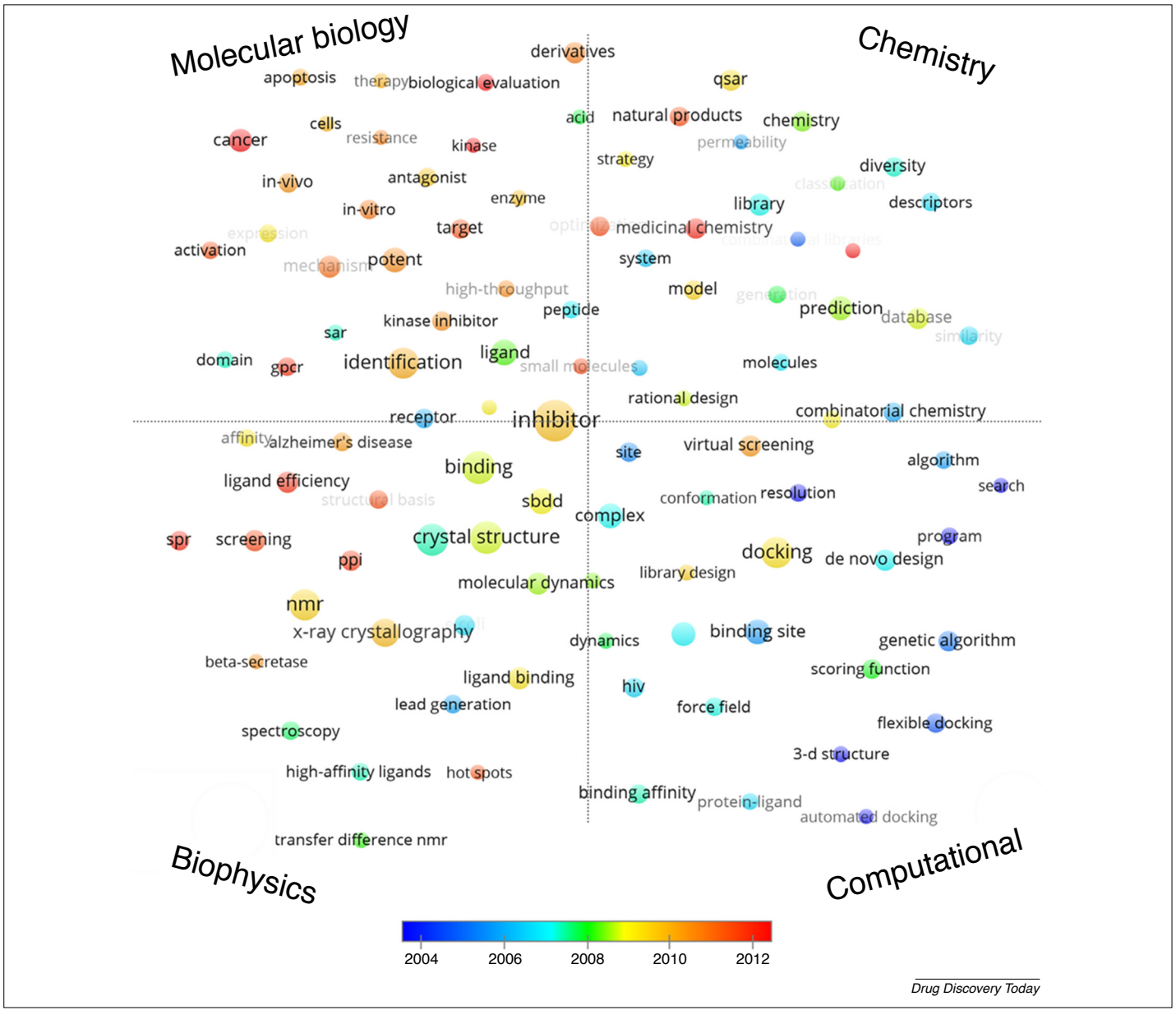

FIGURE 8

Occurrence network of top 100 keywords in fragment-based drug discovery (FBDD). Color correspons to the average year of occurrence of each keyword.

Concluding remarks

Here, we have shown the history of FBDD by using bibliometric methods. During the early days of the field, research in FBDD was highly fragmented, operating under the general umbrella of drug discovery. Today, scientific progress in FBDD are organized with the leading keywords 'fragment-based lead discovery' (FBLD), 'fragment-based drug discovery', and 'fragment-based drug design.' Although these terms all refer to the same approaches, they put emphasis on different aspects of work and the ultimate aim of the endeavors.

The history of FBDD provides a solid case for how recombining knowledge from various worlds can advance science. This was seen at two levels. First, on the organizational level, industry and academia played their respective roles reliably well. Academia laid down the theoretical foundations and also generated research on methods that could be later implemented industrially. With the basic science laid out, industry was able to valorize the knowledge and integrate it into actual drug discovery efforts. Progress in FBDD was able to occur alongside a growing interconnected network of collaborations among various institutions. The studies clearly identify an increasing interconnectedness between academia and industry. Interestingly, FBDD research field has developed over the same years that the pharmaceutical research landscape has undergone major changes, with big pharmaceutical companies outsourcing an increasing amount of preclinical research work [50]. As such, FBDD forms an interesting topic to further explore business development and innovation management in the pharmaceutical sciences. Using the bibliometric database as a

premise, we would like to deepen understanding of how collaborations are formed. Also, with collaborations in FBDD increasing, it is of value to understand how these collaborations are maintained so that all the complementary abilities of each partners are synergized instead of working separately. Finally, it is of essence to evaluate the success of these initiatives towards open innovation.

The technical aspect of the development of FBDD shows us that integration of outsider technologies with solid theoretical grounding is a useful approach to innovation. Being able to spot opportunities for integrating is becoming a more valuable skill for researchers wanting to stay on top of their fields. It is of interest then to understand how both academia and industry cope with this need. Further surveys should be done on this front.

Future studies should also address the limitations of our current approach. In this bibliometric analysis, we only focused on scientific publications in FBDD. This analysis identified the key opinion leaders of the field and publications that are accessible to the world-wide research community make an obvious impact. However, certain key contributions to the FBDD field are excluded from the analysis. Given that pharmaceutical companies and biotechs are often not incentivized to publish, analyzing the patent landscape might be able to characterize better the current state of collaborations in the field. Collecting additional data sources, such as company disclosures, conference attendance, and new chemical entities in the market, could provide a comprehensive picture of the growth of FBDD. By connecting and analyzing these data together, it would be possible to better understand the factors that allow companies to successfully bring their laboratory results to the market. We believe that building a better understanding of business development and innovation management in such a well-defined and recently developed research area as FBDD offers useful case studies to describe the changing landscape of pharmaceutical sciences.

Acknowledgment

This work was supported by the European Union's Horizon2020 MSCA Program under grant agreement 675899 (FRAGNET).

References

1 Murray, C. and Rees, D.C. (2009) The rise of fragment-based drug discovery. *Nat. Chem.* 1, 187–192
2 Baker, M. (2013) Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* 12, 5–7
3 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531
4 Nowak, T. and Hubbard, R.E. (2013) Introduction to FBDD. In *Fragments Workshop 2013*, p. 6, Publisher
5 Chessari, G. and Woodhead, A.J. (2009) From fragment to clinical candidate: a historical perspective. *Drug Discov. Today* 14, 668–675
6 Hammonds, T. and Simpson, P.B. (2015) *The Handbook of Medicinal Chemistry: Principles and Practice*. The Royal Society of Chemistry
7 Keserű, G.M. *et al.* (2016) Design principles for fragment libraries: maximizing the value of learnings from pharma fragment-based drug discovery (FBDD) programs for use in academia. *J. Med. Chem.* 59, 8189–8206
8 Erlanson, D.A. *et al.* (2016) Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* 15, 605–619
9 Hall, R.J. *et al.* (2014) Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* 116, 82–91
10 de Kloe, G.E. *et al.* (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discov. Today* 14, 630–646
11 Keijl, S. *et al.* (2016) The two faces of inventions: the relationship between recombination and impact in pharmaceutical biotechnology. *Res. Policy* 45, 1061–1074
12 Tralau-Stewart, C.J. *et al.* (2009) Drug discovery: new models for industry–academic partnerships. *Drug Discov. Today* 14, 95–101
13 Schuhmacher, A. *et al.* (2013) Models for open innovation in the pharmaceutical industry. *Drug Discov. Today* 18, 1133–1137
14 Ekins, S. *et al.* (2013) Four disruptive strategies for removing drug discovery bottlenecks. *Drug Discov. Today* 18, 265–271
15 Sams-Dodd, F. (2005) Optimizing the discovery organization for innovation. *Drug Discov. Today* 10, 1049–1056
16 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431
17 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
18 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
19 van Eck, N.J. and Waltman, L. (2014) CitNetExplorer: a new software tool for analysing and visualizing citation networks. *J. Informetr.* 8, 802–823
20 Van Eck, N.J. and Waltman, L. (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 523–538
21 Moon, J.B. and Howe, W.J. (1991) Computer design of bioactive molecules: a method for receptor-based de novo ligand design. *Proteins Struct. Funct. Bioinf.* 11, 314–328
22 Rotstein, S.H. and Murcko, M.A. (1993) GroupBuild: a fragment-based method for de novo drug design. *J. Med. Chem.* 36, 1700–1710
23 Böhm, H.-J. (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 6, 61–78
24 Boehm, H.-J. *et al.* (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* 43, 2664–2674
25 Murray, C.W. and Verdonk, M.L. (2002) The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput. Aided Mol. Des.* 16, 741–753
26 Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* 48, 403–413
27 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672
28 Carr, R.A.E.R. *et al.* (2005) Fragment-based lead discovery: leads by design. *Drug Discov. Today* 10, 987–992
29 Dasgupta, P. and David, P.A. (1994) Toward a new economics of science. *Res. Policy* 23, 487–521
30 Lexchin, J. *et al.* (2003) Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 326, 1167–1170
31 Frye, S. *et al.* (2011) US academic drug discovery. *Nat. Rev. Drug Discov.* 10, 409–410
32 Azoulay, P. *et al.* (2017) The mobility of elite life scientists: professional and personal determinants. *Res. Policy* 46, 573–590
33 Jencks, W.P. (1981) On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U. S. A.* 78, 4046–4050
34 Bernstein, F.C. *et al.* (1977) The protein data bank. *Eur. J. Biochem.* 80, 319–324
35 Kuntz, I.D. *et al.* (1982) A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 161, 269–288
36 Brooks, B.R. *et al.* (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217
37 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
38 Miranker, A. and Karplus, M. (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins Struct. Funct. Bioinf.* 11, 29–34
39 Kneller, R. (2010) The importance of new companies for drug discovery: origins of a decade of new drugs. *Nat. Rev. Drug Discov.* 9, 867–882
40 Wang, L. *et al.* (2015) Racing to define pharmaceutical R&D external innovation models. *Drug Discov. Today* 20, 361–370
41 Eisen, M.B. *et al.* (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins Struct. Funct. Bioinf.* 19, 199–221
42 Friesner, R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749
43 Irwin, J.J. and Shoichet, B.K. (2005) ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182

44 Hajduk, P.J. *et al.* (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* 119, 5818–5827
45 Oltersdorf, T. *et al.* (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681
46 Collaborative Computational Project 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760
47 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326
48 Murshudov, G.N. *et al.* (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255
49 Filippakopoulos, P. *et al.* (2010) Selective inhibition of BET bromodomains. *Nature* 468, 1067–1073
50 Fishburn, C.S. (2013) Translational research: the changing landscape of drug discovery. *Drug Discov. Today* 18, 487–494