

This is a repository copy of *When fragments link: a bibliometric perspective on the development of fragment-based drug discovery*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/132630/>

Version: Accepted Version

---

**Article:**

Romasanta, Angelo K.S., van der Sijde, Peter, Hellsten, Iina et al. (4 more authors) (2018) *When fragments link: a bibliometric perspective on the development of fragment-based drug discovery*. *Drug discovery today*. ISSN 1359-6446

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

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# When Fragments Link: A Bibliometric Perspective on the Development of Fragment-Based Drug Discovery

---

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

<sup>1</sup> Department of Science, Business & Innovation, Faculty of Science, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

<sup>2</sup> Amsterdam Institute of Molecules, Medicines and Systems (AIMMS), Division of Medicinal Chemistry, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

<sup>3</sup> Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

<sup>4</sup> Vernalis Research, Granta Park, Abingdon, Cambridge CB21 6GB, UK.

<sup>5</sup> York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK.

<sup>6</sup> Research Centre for Natural Sciences, Hungarian Academy of Sciences, 1117 Budapest, Magyar Tudósok körútja 2., 1525 Budapest, PO Box 17, Hungary.

## Corresponding Author

De Esch, IJP (i.de.esch@vu.nl)

## Keywords

Fragment-based drug discovery, FBDD, bibliometrics, drug discovery, academia-industry collaboration

## Teaser

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

## **Abstract**

Fragment-based drug discovery (FBDD) is a highly interdisciplinary field, rich in integrating ideas from pharmaceutical sciences, chemistry, biology and physics, amongst others. Enriching our understanding of the field's development, we employed bibliometric techniques to analyze 3,642 publications in FBDD, complementing accounts by key practitioners. Mapping its core papers, we find the transfer of knowledge from academia to industry. Co-authorship analysis shows that university-industry collaboration has grown over time. Moreover, we show how ideas from other scientific disciplines were integrated into FBDD paradigm. Keyword analysis show the field's organization 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**

Fragment-based drug discovery (FBDD) is now a widely adopted approach to lead discovery [1,2]. The field's origin can be traced back to its first demonstration 20 years ago at Abbott Laboratories by Shuker, Hajduk, Meadows and Fesik [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 have also been described in key reviews [7–9]. 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 FBDD's development 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 employ bibliometrics, an approach in information sciences to analyze the relationship among written publications.

In the past, technological breakthroughs have resulted from scientists working together at the interface of diverse disciplines, recombining knowledge from various fields [10]. FBDD's emergence can be seen as various scientific fields coming together: computational methods, molecular biology and medicinal chemistry. With pharmaceutical sciences being more multidisciplinary and the pharmaceutical industry seeking more collaborations especially in pre-clinical development [11–13], it is appealing to investigate the drivers that made FBDD so successful. With the increasing interest of how organizational factors can enable drug discovery [14], 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 how each keyword trend over time, we can get a sense of how the focus of FBDD changed over time and what its current direction is.

## Data and Methods

The papers analyzed in this study were downloaded from Thomson-Reuter's Web of Science (WOS). In order to collect an initial set of papers in the field of fragment based drug discovery, keywords (see **Fout! Verwijzingsbron niet gevonden.**) were used.

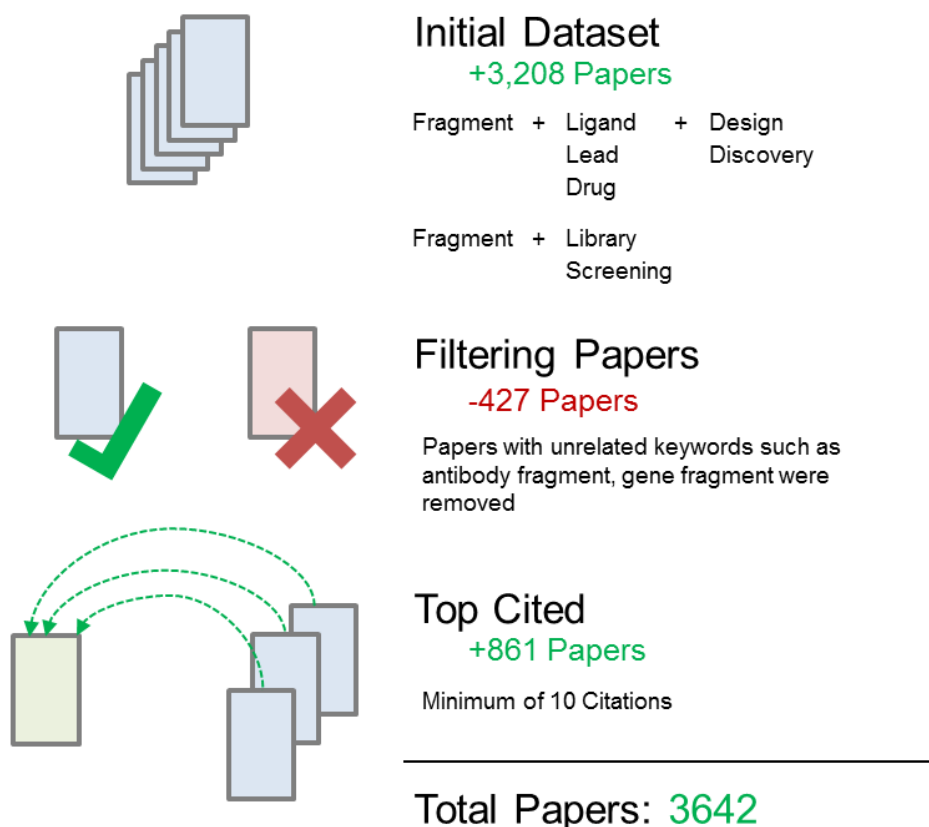


Figure 1. Data Collection for FBDD Publications

The keyword search generated 3,208 papers. To ensure that the keyword “fragment” was used to refer to the field, we looked at the abstract, title and keywords 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 to a dataset of 2,781 papers. To verify whether these papers were representative of FBDD, we inspected the dataset 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 [15] and Hann’s paper on molecular complexity [16], as these have not mentioned any of the used keywords (see **Fout! Verwijzingsbron niet gevonden.**). 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 10 times. It is important to note that this list contains some

publications that may 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) that marks the pivotal role of protein structural information in FBDD [17]. Merging these publication lists resulted in a total number of publications of 3,642 that span the years from 1953 to 2016.

In order 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 dataset that were published from 1996-2016 in intervals of 5-years. 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 dataset of FBDD articles were identified to find the core papers in FBDD. For further analysis, we used the software CitNetExplorer [18] 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 evolved in the last 20 years, we generated co-authorship network maps using the software VosViewer [19]. In an effort 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 are 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 five years would mostly operate under the general umbrella of drug discovery instead of distinguishing itself as a particular discipline. However, traces of the keywords related to fragment-based drug discovery has already been present as early as the nineties, for example in the computational work of Joseph Moon and Jeffrey Howe [20] at Upjohn, Sergio Rotstein and Mark Murcko [21] at Vertex, and Hans-Joachim Böhm [22] at BASF. Synonyms such as *needles* and *needle screening*, used to describe early applications by Böhm and co-workers, now at F. Hoffmann-La Roche [23], were not adopted by the scientific community as these keywords were used in less than 5 publications in any year. As seen in **Fout! Verwijzingsbron niet gevonden.**, It would take a few more years before research in the field would come together in a term such as '*fragment-based drug discovery*', first appearing in the abstract of the 2002 paper of Murray and Verdonk [24]. Even in those days, the field would alter between the keywords 'lead' and 'drug'. The term *lead discovery* would dominate in the early years, stimulated by influential reviews coming from Astex [25–27] in the mid-2000s. Differentiating between the two, the term *lead* emphasizes the early stage wherein fragments are used (for example, before pharmacokinetic properties are being considered). On the other hand, the term *drug* can be helpful in that it contextualizes the ultimate goal that fragments aim to achieve, and that is to develop drugs.

By 2009, the term fragment-based *drug* discovery would finally overtake as the top keyword that researchers used to identify the field while *lead discovery* has lost favor back from its peak in 2005 as shown in **Fout! Verwijzingsbron niet gevonden.** 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, while drug design puts more emphasis on the rational approaches to build the new drug (candidate). As it is, the abbreviation FBDD is now being *drug discovery* has been favored to *drug design*, being used as much as 3 times more in 2016 according to the Web of Science, although the different words seem to be used as synonyms.

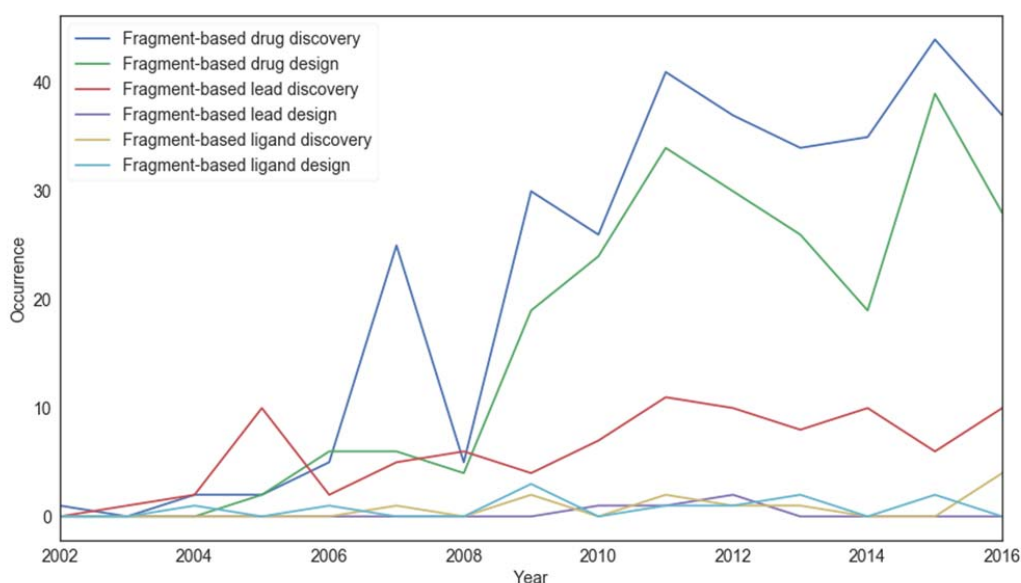


Figure 2. Occurrence of FBDD umbrella keywords in the literature. These keywords were chosen as these were the terms used to refer to the field in various important reviews.

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 its first five years, this has now increased six-fold to 1709 publications from 2011-2016. 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.

**Table 1.** Summary of the FBDD dataset from 1996-2016

	1996-2000	2001-2005	2006-2010	2011-2016
<b>Publications</b>	277	496	939	1709
<b>Journals</b>	95	143	220	363
<b>Researchers</b> (min. 5 publications) <sup>a</sup>	102	190	343	389
<b>Organizations</b> (min. 10 publications) <sup>a</sup>				
<b>Academia</b>	1	4	15	53
<b>SMEs</b>	0	3	3	6
<b>Big Pharma</b>	1	7	7	7

<sup>a</sup> Threshold needed to be set as some firms and researchers may co-author publications but not necessarily practice FBDD

### From Ideas to Application: The Role of Industry

Clearly, the industry has played a pivotal role in developing FBDD. Although the approach was first demonstrated at the big pharmaceutical firm 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 application in drug discovery. In the first few years of the field, majority of articles were coming from the industry. This is noteworthy because an inherent bias towards universities is expected when focusing on scientific publications, due to the incentive of academia to publish. Considering that the industry has the opposite incentive of withholding information for competitive advantage [28,29], 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 in the late nineties. Astex (founded in 1999 by University of Cambridge professors Tom Blundell and Chris Abell and former head of structural biology and bioinformatics of GlaxoWellcome dr. Harren Jhoti) led in the next decade. Only in the recent five years would there be a surge in publications from the academia in the top 10 list. Table 2 also shows that biotech companies such as Astex, Vertex and Sunesis have played an important role in establishing the field. However, it is also important to note that some prominent biotechs and pharmaceutical companies in FBDD do not show up in this particular analysis as they might have put less emphasis on authoring scientific publications.

**Table 2.** Top institutional publishers and their total citations in the field of FBDD over time <sup>a</sup>

Rank	1996-2001		2001-2006		2006-2011		2011-2016	
1	Abbott Labs	154	Astex	368	Astex	595	Univ Oxford	368
2	Vertex	77	Abbott Labs	221	Abbott	320	Univ Cambridge	348
3	Univ California San Francisco	52	Sunesis	187	Univ California San Francisco	261	Glaxosmithkline	304
4	Roche	49	Novartis	163	Astrazeneca	249	Astex	232
5	Novartis	43	Pfizer	139	Univ Cambridge	216	Univ California San Francisco	156
6	Univ Sheffield	35	Scripps Inst	112	Novartis	188	Astrazeneca	139
7	BASF	34	Astrazeneca	93	Scripps Inst	187	Heptares	120
8	Univ of California San Diego	29	Glaxosmithkline	90	Glaxosmithkline	184	Pfizer	110
9	Univ Marburg	28	Sanford Burnham	87	Pfizer	155	Cancer Research UK	105
10	CCDC	26	Univ California San Francisco	85	Vernalis	135	Univ Dundee	104

<sup>a</sup> Industry groups in red

The important role that the private sector has played in FBDD innovation is also apparent when looking at the top 10 cited papers from our collection of FBDD papers (Table 3). A staggering number of nine of the ten top publications were written by industry researchers. The only paper in the top 10 coming from academia is Berman's publication on the Protein Data Bank [17], which does not strictly belong to FBDD but is a fundamental resource for drug discovery research in general and for FBDD in particular as many of the hit fragment optimization programs have been guided by protein structural data. Next to some influential reviews, including work from Hajduk, Congreve, Rees (all from 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 made enormous impact (rank 2, Table 3). LE assesses the contribution of every atom on the affinity of the ligand and is being used to select the most promising fragment hits and to guide the growing of the 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. Amongst others, this work raised the realization that fragments should be simple and small molecules that can interrogate the binding sites with higher resolution. Amongst others, this has resulted in the guidelines captured in *the Rule of Three* that 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 soluble and permeable drug-like molecules, the ultimate goal of FBDD efforts.

**Table 3.** The ten most cited papers in the dataset of FBDD articles<sup>a</sup>

	Authors	Title	Journal	Affiliation	Year	Cit
1	Shuker, SB; Hajduk, PJ; Meadows, RP; Fesik, SW	Discovering high-affinity ligands for proteins: SAR by NMR	Science	Abbott Labs	1996	454
2	Hopkins, AL; Groom, CR; Alex, A	Ligand efficiency: a useful metric for lead selection	Drug Discovery Today	Pfizer	2004	403
3	Hajduk, PJ; 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, AJ	Recent developments in fragment-based drug discovery	Journ of Medicinal Chemistry	Astex	2008	290
6	Hann, MM; Leach, AR; Harper, G	Molecular complexity and its impact on the probability of finding leads for drug discovery	Journ Chem Info and Comp Sci	Glaxosmithkline	2001	287
7	Lipinski, CA; Lombardo, F; Dominy, BW; Feeney, PJ	Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings	Advanced Drug Delivery Reviews	Pfizer	1997	286
8	Rees, DC; Congreve, M; Murray, CW; Carr, R	Fragment-based lead discovery	Nature Reviews Drug Discovery	Astex	2004	275
9	Berman, HM; Westbrook, J; Feng, Z; Gilliland, G; Bhat, TN; Weissig, H; Shindyalov, IN; Bourne, PE	The Protein Data Bank	Nucleic Acids Research	Rutgers Univ NIST Burnham Inst UCSD	2000	257
10	Erlanson, DA; McDowell, RS; O'Brien, T	Fragment-based drug discovery	Journ of Medicinal Chemistry	Sunesis	2004	219

<sup>a</sup> Industry groups in red

However, if we look at the top cited journals in the recent years (Table 4), seven out of the 10 most cited publications were authored by academia from 2009. This adoption by academia is validated as well by the

increase in the share of publishing universities and research institutions in FBDD in the last five years. One of the factors for the adoption by academia is the rise of academic medicinal chemistry and drug discovery groups [11,30]. We can also speculate on the mobility of researchers, including experts from industry that move towards university, setting up academic drug discovery research groups. As there has been an increase of interest in how researcher mobility affects innovation [31], the impact of this mobility and transfer of knowledge in FBDD's development will be the topic of future research.

**Table 4.** The ten most cited papers published from 2009 in the dataset of FBDD articles.

	Authors	Title	Journal	Affiliation	Year	Cit
1	Murray, CW Rees, DC	The rise of fragment-based drug discovery	Nature Chemistry	Astex	2009	141
2	Chessari, G Woodhead, AJ	From fragment to clinical candidate-a historical perspective	Drug Discovery Today	Astex	2009	82
3	Murray, CW Verdonk, ML Rees, DC	Experiences in fragment-based drug discovery	Trends In Pharmacological Sciences	Astex	2012	76
4	Scott, DE Coyne, AG Hudson, SA Abell, C	Fragment-based approaches in drug discovery and chemical biology	Biochemistry	Univ Cambridge	2012	75
5	Murray, CW Blundell, TL	Structural biology in fragment-based drug design	Current Opinion In Structural Biology	Univ Cambridge Astex	2010	70
6	de Kloe, GE Bailey, D Leurs, R de Esch, IJP	Transforming fragments into candidates: small becomes big in medicinal chemistry	Drug Discovery Today	IOTA Vrije Univ Amsterdam	2009	69
7	Filippakopoulos,P Bradner, JE et al.	Selective inhibition of BET bromodomains	Nature	Dana Farber Cancer Inst Harvard Univ Univ Notre Dame Univ Oxford	2010	68
8	Baker, M	Fragment-based lead discovery grows up	Nature Reviews Drug Discovery		2013	67
9	Emsley, P Lohkamp, B Scott, WG Cowtan, K	Features and development of Coot	Acta Crystallographica Section D-Biological Crystallography	Karolinska Inst Univ York UC Santa Cruz Univ Oxford	2010	67
10	Chen, Y Shoichet, BK	Molecular docking and ligand specificity in fragment-based inhibitor discovery	Nature Chemical Biology	UC San Francisco	2009	62

### Knowledge Transfer: The Role of University-Industry Collaboration

We then explored the list to the top 100 cited articles in FBDD, representing the core papers of FBDD. By creating a citation map of these articles over time, we can visualize 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 **Fout! Verwijzingsbron niet gevonden.** reveals that ideas and tools developed in academia provided groundwork for the field.

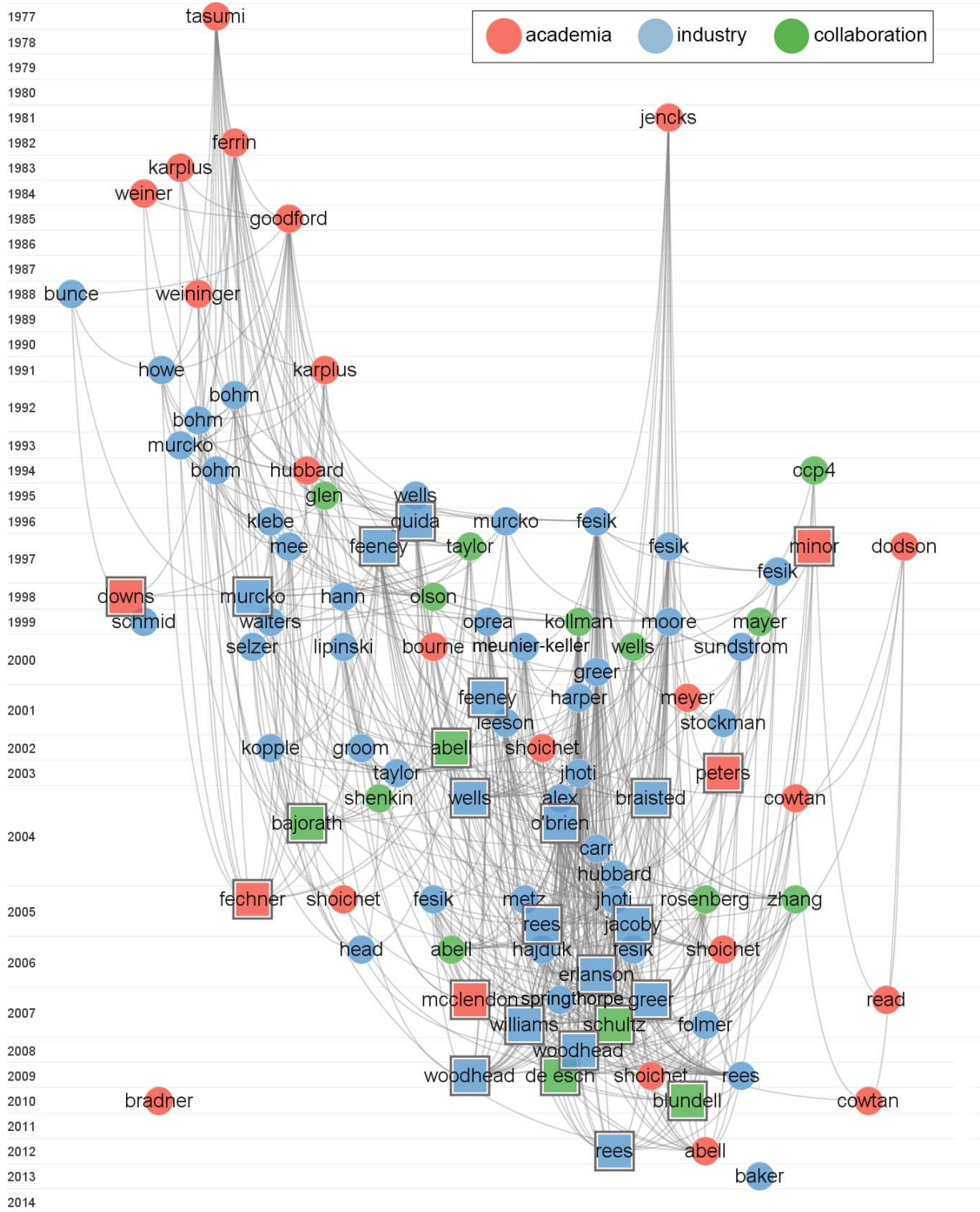


Figure 3. Citation map of 100 core papers in FBDD. Each paper is labelled by its last author. The colors reflect the affiliation of the authors Square highlights mark review articles.

Most of the theoretical grounding of FBDD came with ideas from the academia as early as 1970s. This early influence by the academia can be seen explicitly with the paper of Jencks from Brandeis University [32]. In his paper on the additivity of binding energies, he puts out the idea that large molecules can be considered as a combination of fragments.

On the upper left side of the citation map, a number of papers authored in the academia can also be seen. These are foundational publications about influential drug discovery tools such as the Protein Data Bank in 1977 [33], molecular docking approaches by Ferrin and co-workers in 1982 [34], the molecular modelling software CHARMM by Karplus and co-workers in 1983 [35] and Goodford's computational-procedure for determining energetically favorable binding-sites in 1988 [36] and functionality maps of binding sites by Karplus *et al.* in 1991 [37]. Also other computational chemistry efforts (e.g., Karplus, Schneider, Hubbard) to develop *de novo* structure generation and molecular docking software have 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. In those early nineties, the technologies and protocols to determine fragment binding to proteins, using for example sensitive biophysical technologies, were not yet available. The 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 (*vide supra*) 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 in developing the applications is not only apparent from the work of Fesik and co-workers with NMR technology, also the work of Greer and co-workers that focusses on discovering ligands using X-ray crystallographic screening. Later, their crystallographic screening method called CrystaLEAD) was further developed and exploited by influential scientists like Hubbard (University of York, Vernalis), Rees, Jhoti (Astex), Abell, 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 FBDD's foundations and Big Pharma first demonstrating the technique in 1996, the road was now ready for the field's valorization. The next decade of key FBDD publications would almost exclusively come from the industry. Especially in the early 2000s, smaller structure-based drug discovery companies like Astex, Vertex and Sunesis come to play an important role. These biotechs specialized in specific FBDD technologies and approaches (e.g., crystal soaking, biochemical assays, tethering) and perfected them for application in hit finding and lead generation. Fragments provided a way for these companies 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]. It is noted that not all known technologies and FBDD companies do appear in this bibliometric analysis, possibly because their restricted efforts to publish in scientific literature. It is interesting to see that those companies that do publish make an significant impact when considering collaborations that publish FBDD work (**Fout! Verwijzingsbron niet gevonden.**).



Figure 4. Collaboration network map of top 500 publishing institutions in FBDD grouped in five year periods from 1996-2016. Each node corresponds to an institution. The size reflects the number of publications. Red nodes are from academia while blue nodes are from industry. Dark blue nodes are from big pharma while light blue are other industrial firms including small biotechs and firms from adjacent industries. For years 2006-2016, the biggest cluster is shown

### 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 seen in **Fout!**



Verwijzingsbron niet gevonden..

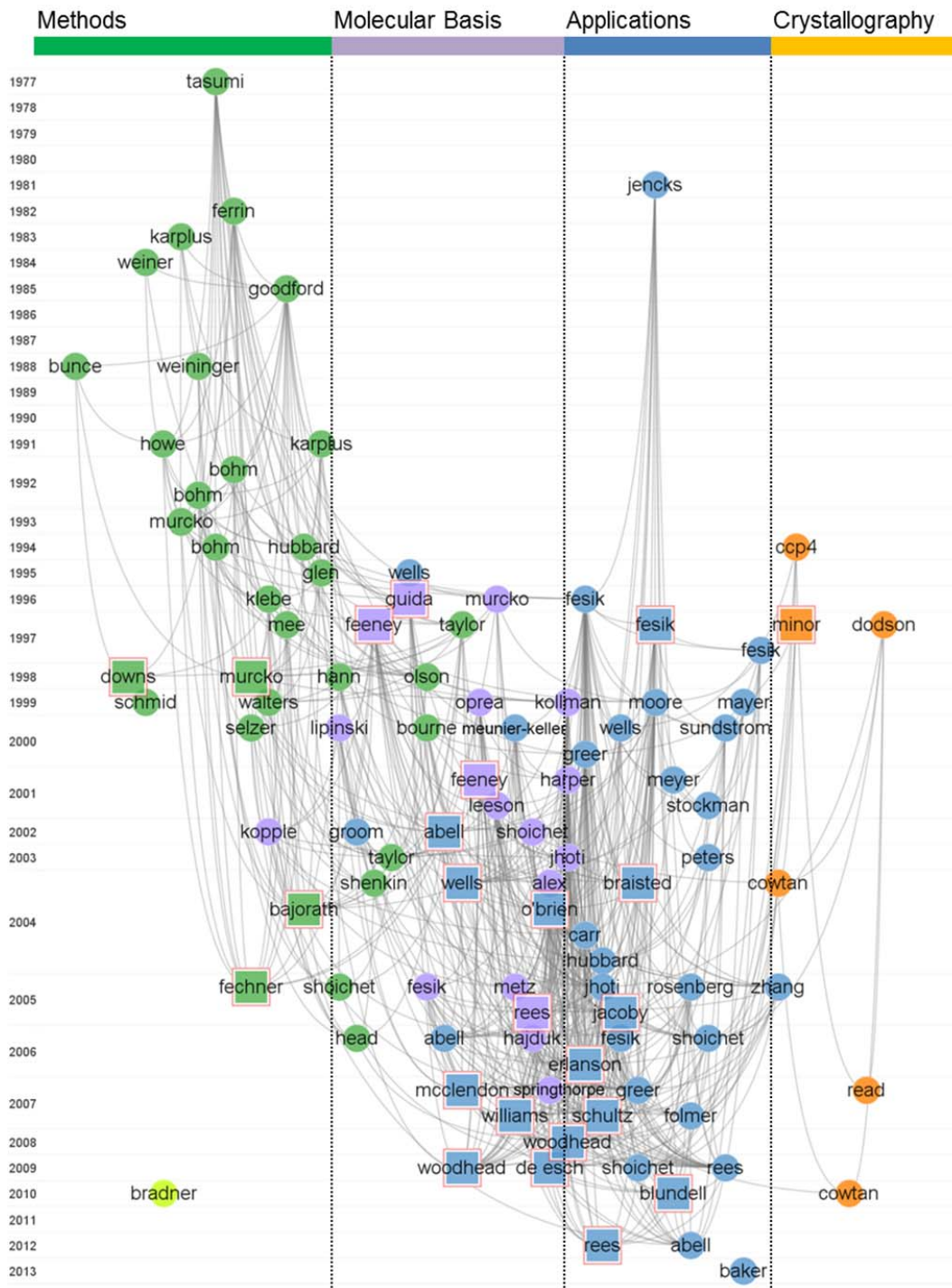


Figure 5. Citation map of 100 core papers in FBDD. The colors show clustering of papers by similarity.

Before 1996, the scientific groundwork that would eventually be integrated in FBDD would come from two separate fronts. As seen on the upper right side of **Fout! Verwijzingsbron niet gevonden.**, on one

end, we have the work of Jencks which provided the theoretical rationalization for fragments. On the other end (green cluster of **Fout! Verwijzingsbron niet gevonden.**), 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 **Fout! Verwijzingsbron niet gevonden.**, there is a clear separation between these two branches with no paper citing the two, prior to 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 **Fout! Verwijzingsbron niet gevonden.**, the paper serves as a hub where a dense amount of publications branch from. The publication of Fesik brought the two separate branches together, explicitly referring to the paper of Jencks while also referring to Bohm's LUDI [22], Hubbard's HOOK [40] and Murcko's GroupBuild [21] at the same time. By this, 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 respectively measure and visualize low affinity fragment binding.

We looked at the categories of the journal sources of FBDD papers. Doing so allows us to see the disciplines that FBDD was building from. In **Fout! Verwijzingsbron niet gevonden.**, before 1995, FBDD literature cited articles coming from the fields of biophysics, biochemistry and molecular biology and computer science. This signals that advancements in knowledge in these various fronts was necessary for FBDD to expand. It also gives a clear indication that FBDD is going mainstream with many publications now appearing in the more applied Medicinal Chemistry field, whereas in the early years most papers were in the field of biochemistry & molecular biology, biophysics and computational chemistry.

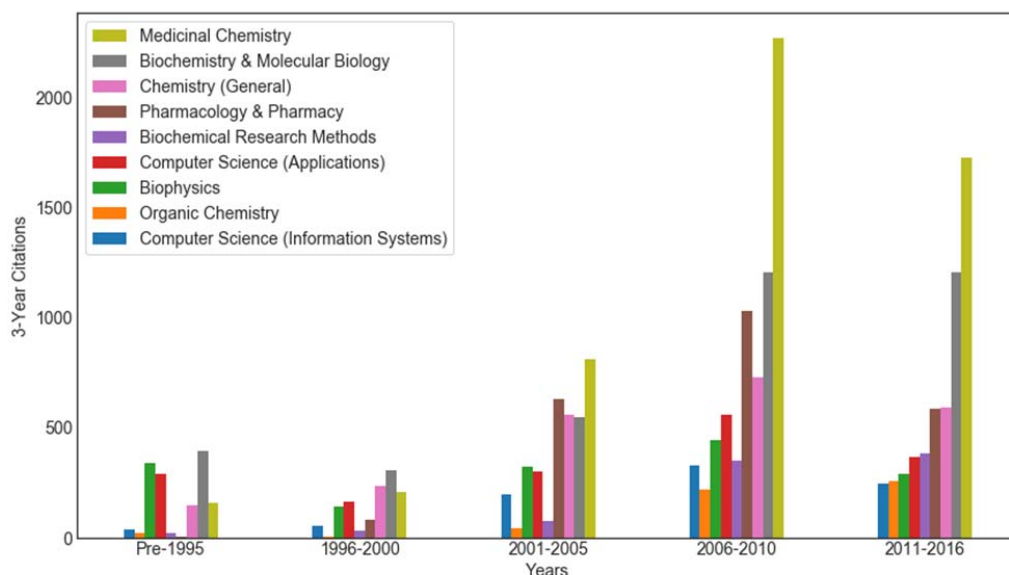


Figure 6. Categories of journals over time



Although this cluster includes the pre-90's computational techniques described before, the influence of this cluster extends into the early 2000s, including *de novo* structure generation and docking algorithms such as Glide[41] in 2004 and development of frequently used databases such as ZINC [42] in 2005.

Referring back to **Fout! Verwijzingsbron niet gevonden.**, the blue cluster on the right side is composed of what is considered as integral FBDD publications. These include principles and demonstrations of how various biophysical techniques can be used in the paradigm of FBDD. Included as well are applications of FBDD to various therapeutic targets, i.e., the actual use in drug discovery [43,44]. 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 1996-2002, which describes concepts on the molecular basis of the approach. One way of interpreting this is that there are researchers (like Fesik) who bridge the gap between a new field and established methods. In this case, they provided the molecular basis of FBDD. By formulating principles from their outsider perspective, they are able to integrate previously unexploited knowledge and technologies to the growing body of FBDD literature. The important role of key opinion leaders can be seen in the central part of the plot, where around 2005 scientists like Rees, Jhoti, Abell (Astex and University of Cambridge), Hubbard (University of York, Vernalis), Fesik, Hajduk (Abbott Laboratories) and Erlanson (Sunesis) explicitly integrate the various

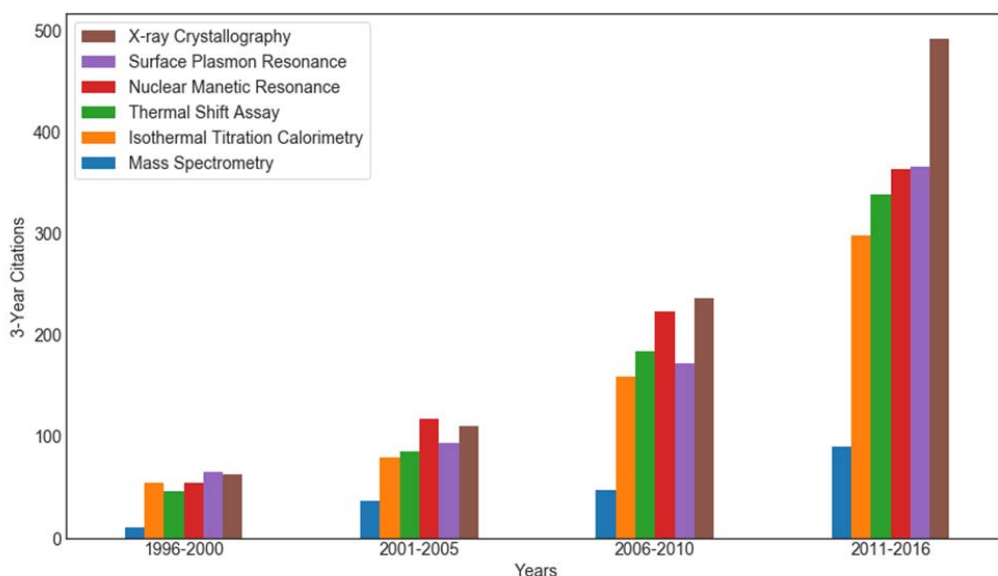


Figure 7. Occurrence of various techniques in FBDD papers over time aspects of FBDD in their publications.

The citation map also shows an orange cluster that was integrated into FBDD relatively more recently. These are papers in the field of crystallography such as the CCP4 suite[45] in 1994, Minor's processing of X-ray diffraction data[46] and Dodson's refinement of macromolecular structures[47] both in 1997.

The impact that crystallography would bring to FBDD carries on today. 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 **Fout! Verwijzingsbron niet gevonden.**, although nuclear magnetic resonance has been the dominant technique in the first years of FBDD, it has been replaced by x-ray crystallography in the last five years. It is important to note that 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.

Today, the field is organized into four interrelated practices. To come up with these four classifications, the top keywords in FBDD was plotted and clustered according to how often they occur together per paper (**Fout! Verwijzingsbron niet gevonden.**). Four clusters were detected, corresponding to the four major interrelated activities in FBDD – 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 indicate the category and interrelatedness of the keywords with each other, the position must be taken with a grain of salt as keywords are more often than not related to the three other dimensions of FBDD.

In order to see what the trends have been in FBDD in the years, these keywords were colored according to the average year of publication. As seen in **Fout! Verwijzingsbron niet gevonden.**, 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 as 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 publication of Bradner [48]. Going back to **Fout! Verwijzingsbron niet gevonden.**, this publication does not cite the core FBDD literature yet is cited by a lot of the recent papers in FBDD. This publication on the inhibition of BET bromodomains has been an area of interest for FBDD researchers in the recent years.

Together with other targets, the focus now for FBDD has been its application. The most cited references in the recent years (as seen in Table 4) have been reviews showing how more and more leads originating from FBDD are entering the clinical trials. Not only industry is using the technique, but also various academic groups. As Baker [2] puts it on a news article in 2013, FBDD has indeed grown up.



## Conclusions

In our paper, we have shown the history of FBDD through bibliometric methods. In the early days of the field, research in FBDD was highly fragmented, operating under the general umbrella of drug discovery. Today, scientific progress in FBLD and FBDD are organized with the leading keywords “fragment-based lead discovery”, “fragment-based drug discovery” and “fragment-based drug design.” Although all these terms 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 on how recombining knowledge from various worlds can advance science. This was seen in 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 can be later implemented industrially. With the basic science laid out, industry was able to valorize the knowledge and integrate them in 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 in the same years that the pharmaceutical research landscape was undergoing major changes, with the big pharmaceutical companies outsourcing more and more pre-clinical research work [49]. 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 the knowledge on how collaborations are formed. Also, with collaborations in FBDD being increasingly created, 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 FBDD’s development show 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 can address the limitations of our current approach. In this bibliometric analysis, we only focused on scientific publications in FBDD. This analysis has identified the key opinion leaders of the field and clearly, publications that are accessible for the world-wide research community make an obvious impact. However, certain key contributions to the FBDD field are excluded from the analysis. As 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 companies disclosures, conference attendance and new chemical entities in the market could provide a comprehensive picture on 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 towards 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.

## GRANT SUPPORT

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

## References

- [1] Murray C, Rees DC. The rise of fragment-based drug discovery. *Nat Chem* 2009;1:187–192. doi:10.1038/nchem.217.
- [2] Baker M. Fragment-based lead discovery grows up. *Nat Rev Drug Discov* 2013;12:5–7.
- [3] Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. *Science (80- )* 1996;274:1531.
- [4] Nowak T, Hubbard RE. Introduction to FBDD. *Fragm. Work.*, 2013, p. 6.
- [5] Chessari G, Woodhead AJ. From fragment to clinical candidate—a historical perspective. *Drug Discov Today* 2009;14:668–75.
- [6] Hubbard RE. *The handbook of medicinal chemistry: principles and practice*. Cambridge: The Royal Society of Chemistry; 2015.
- [7] Keserü GM, Erlanson DA, Ferenczy GG, Hann MM, Murray CW, Pickett SD. 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* 2016.
- [8] Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* 2016;15:605–19.
- [9] Hall RJ, Mortenson PN, Murray CW. Efficient exploration of chemical space by fragment-based screening. *Prog Biophys Mol Biol* 2014;116:82–91..
- [10] Keijl S, Gilsing VA, Knobben J, Duysters G. The two faces of inventions: The relationship between recombination and impact in pharmaceutical biotechnology. *Res Policy* 2016;45:1061–74.
- [11] Tralau-Stewart CJ, Wyatt CA, Kleyn DE, Ayad A. Drug discovery: new models for industry--academic partnerships. *Drug Discov Today* 2009;14:95–101.
- [12] Schuhmacher A, Germann P, Trill H, Gassmann O. Models for open innovation in the pharmaceutical industry. *Drug Discov Today* 2013.
- [13] Ekins S, Waller CL, Bradley MP, Clark AM, Williams AJ. Four disruptive strategies for removing drug discovery bottlenecks. *Drug Discov Today* 2013;18:265–71. doi:10.1016/j.drudis.2012.10.007.
- [14] Sams-Dodd F. Optimizing the discovery organization for innovation. *Drug Discov Today* 2005;10:1049–56.
- [15] Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* 2004;9:430–1.

- [16] Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* 2001;41:856–64.
- [17] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- [18] van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *J Informetr* 2014;8:802–23.
- [19] Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84:523–38.
- [20] Moon JB, Howe WJ. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *Proteins Struct Funct Bioinforma* 1991;11:314–28.
- [21] Rotstein SH, Murcko MA. GroupBuild: a fragment-based method for de novo drug design. *J Med Chem* 1993;36:1700–10.
- [22] Böhm H-J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* 1992;6:61–78.
- [23] Boehm H-J, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, et al. 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* 2000;43:2664–74.
- [24] Murray CW, Verdonk ML. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J Comput Aided Mol Des* 2002;16:741–53.
- [25] Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H. Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 2005;48:403–13.
- [26] Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov* 2004;3:660–72.
- [27] Carr RAER, Congreve M, Murray C, Rees DDC. Fragment-based lead discovery: leads by design. *Drug Discov Today* 2005;10:987–92.
- [28] Dasgupta P, David PA. Toward a new economics of science, *Research Policy*, 23, 487–521. *Res Policy* 1994;23:487–521.
- [29] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *Bmj* 2003;326:1167–70.
- [30] Frye S, Crosby M, Edwards T, Juliano R. US academic drug discovery. *Nat Rev Drug Discov* 2011;10:409–10.
- [31] Azoulay P, Ganguli I, Zivin JG. The mobility of elite life scientists: Professional and personal determinants. *Res Policy* 2017;46:573–90.
- [32] Jencks WP. On the attribution and additivity of binding energies. *Proc Natl Acad Sci* 1981;78:4046–50.
- [33] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The protein

- data bank. *Eur J Biochem* 1977;80:319–24.
- [34] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982;161:269–88.
- [35] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217.
- [36] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985;28:849–57.
- [37] Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins Struct Funct Bioinforma* 1991;11:29–34.
- [38] Kneller R. The importance of new companies for drug discovery: origins of a decade of new drugs. *Nat Rev Drug Discov* 2010;9:867–82.
- [39] Wang L, Plump A, Ringel M. Racing to define pharmaceutical R&D external innovation models. *Drug Discov Today* 2015;20:361–70. d
- [40] Eisen MB, Wiley DC, Karplus M, Hubbard RE. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins Struct Funct Bioinforma* 1994;19:199–221.
- [41] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47:1739–49.
- [42] Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–82.
- [43] Hajduk PJ, Sheppard G, Nettlesheim DG, Olejniczak ET, Shuker SB, Meadows RP, et al. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J Am Chem Soc* 1997;119:5818–27.
- [44] Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, et al. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 2005;435:677–81.
- [45] Collaborative CP, others. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760.
- [46] Otwinowski Z, Minor W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–26.
- [47] Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol Crystallogr* 1997;53:240–55.
- [48] Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, et al. Selective inhibition of BET bromodomains. *Nature* 2010;468:1067–73.
- [49] Fishburn CS. Translational research: the changing landscape of drug discovery. *Drug Discov Today* 2013;18:487–94.

