https://doi.org/10.51408/issi2025_143

Application of Molecular Docking Technology in Drug Discovery Based on Bibliometric and Patent Analysis

Zhou Haichen[1], Jorge Gulín-González[2], Chen Yunwei[3]

[1]*zhouhc@clas.ac.cn*, [3]*chenyw@clas.ac.cn*
Scientometrics & Evaluation Research Center (SERC) of National Science Library (Chengdu), Chinese Academy of Sciences (People's Republic of China)

[2]*gulinj@uci.cu*
Centro de Estudios de Matemática Computacional (CEMC). Universidad de las Ciencias Informáticas (UCI), La Habana (Cuba)

Abstract

In the field of molecular modeling, molecular docking (MD) is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using scoring functions. MD is frequently used to predict the binding orientation of small molecule drugs candidates to their protein targets in order to in turn predict the affinity and activity of the small molecules. MD plays an important role in the rational design of drugs. On the other hand, patents are a significant output-based indicator of innovation, analyse countries and organizations' technological capabilities, analyse relationship between polices and technological innovation. In this context, it is interesting to investigate the dynamics of the behaviour of granted patents and scientific articles related to them in the topic of MD, with the development of new drugs. In this paper we present a comprehensive assessment about the application of MD technology in drug discovery through bibliometric and patent analysis, revealing research trends, technological hotspots, key participants and their collaboration networks, as well as the connection between academic research and practical applications. The study covers the period 1979-2024. We use a keyword co-occurrence network to analyze high-frequency keywords in publications and identify research hotspots and we apply BertTopic to extract research topics and their evolution over time. Also, it is our objective reveal development patterns and critical milestones in molecular docking technology for drug discovery. Our research provides research hotspot references for academia and strategic insights for industry stakeholders and promotes collaborative innovation between academic research and industrial practice. In future works we will present the analysis of a study case related to the development of drugs for the treatment of COVID-19.

Introduction

The identification of drug candidates is one of the most arduous stages in the design of new drugs (Schnecke & Boström, 2006). Molecular docking (MD) is a technique which examines the conformation and orientation of molecules, mainly ligands, into the binding site of a protein target. Searching algorithms generate likely poses, which are ranked by scoring functions (Liu et al., 2018). To generate a receptor (protein)-ligand structure *in silico* two steps are followed: (i) Docking per se entails conformational and orientational sampling of the ligand within constraints of the receptor binding site and, (ii) Scoring function selects the best pose for a given molecule and rank orders ligands. At present, several software are available to carry out MD, among them: AutoDock (Morris et al., 1998), AutoDock Vina (Trott & Olson, 2010), DockThor (DeMagalhães, et al., 2014), FlexX (Rarey et al, 1996) and

GOLD (Verdonk, M.L,2003). The number of scientific publications related to MD has been increasing significantly in the last 25 years.

On the other hand, patents are as a significant output-based indicator to measure innovation, analyse countries and organizations' technological capabilities, analyse relationship between polices and technological innovation (Banerjee et al., 2000). Interesting studies on the relationship between patents and the economic development of a country, particularly in the field of biotechnology, have been published recently (Qiang et al. 2019). The analysis of the number of patents can give us an idea of the impact of theoretical and computational techniques such as DM on the development of new drugs. These results can be complemented with an analysis of the number of publications per year, their trend and relation with the number of patents, as well as by the analysis of other scientometric parameters.

In this paper we present a comprehensive assessment about the application of MD technology in drug discovery through bibliometric and patent analysis, revealing research trends, technological hotspots, key participants and their collaboration networks, as well as the connection between academic research and practical applications. This research provides research hotspot references for academia and strategic insights for industry stakeholders and promotes collaborative innovation between academic research and industrial practice.

Methods

The selected scientific papers encompass science citation index expanded (SCIE) articles and conference proceedings citation index–science (CPCI-S) proceedings papers from 1979 to 2024, sourced from the Web of Science Core Collection with the search strategy as: TS=("molecular docking" OR "docking") AND TS=("drug* discover*" OR "ligand*" OR "drug* design*") AND PY=1979-2024. We collected 38240 results in the studied period. For this study, only research papers (articles and proceedings papers) were selected, since they are directly related to the study and development of new drugs. The selection of SCIE as the database for this first part of the study was based on the fact that it is the most internationally recognized database and since molecular docking is a novel and high-impact topic, it is to be expected that a significant percentage of the main research papers will be published in journals from this collection. Patent data information related to molecular docking from the PatSnap Analytics database (Alkhazaleh R & Mykoniatis, 2024), with the search strategy as: Keywords=(("molecular docking" OR "docking")" AND "("drug design" OR "drug discovery" OR "ligand")") AND Publication Date="1979/01/01-2024/12/31". The retrieval date is January 2025. The PatSnap's advantage lies in its coverage of 172 patent offices worldwide, containing over 1.96 billion patent records. With daily updates, it ensures real-time access to comprehensive and up-to-date global patent data, supporting precise and efficient intellectual property research.

Results and Discussion

Overall tendency

In Figure 1 we present the annual publication volume related to MD topic over the studied period 1979-2024. A sustained increase in the number of papers can be observed, this increase is even more significant from 2020, which can be related to the health emergency resulting from the COVID-19 pandemic. As an example of that, the most cited docking software AutoDock is used by the FightAIDS@Home and OpenPandemics - COVID-19 projects run at World Community Grid, to explore for antivirals in the treatment of HIV/AIDS and COVID-19. Also, AutoDock has contributed to the design of relevant drugs, including HIV1 integrase inhibitors (Goodsell et al., 2021). Figure 1 also shows the number of patents per year linked to MD between 1979 and 2024. According to the behavior of the numbers of papers and patents, graph of Figure 1 can be divided in three regions: (i) The first goes up to 1999 and both the number of papers and patent applications are very similar and relatively low, which corresponds to the incipient state of research on the subject and the limited computational capabilities in that period. Also, it is expected that at the beginning prevail research of a theoretical nature, (ii) The second period (between 2000 and 2019), in which there is initially an increase in papers and patent application, then, quickly, in the case of patent application, a plateau (2000-2012) is reached. In our opinion, the initial increase corresponds to a higher degree of maturity of the molecular docking technique; the plateau can be explained from the accumulation of knowledge in the previous period and the non-existence of exceptional pandemic episodes at these years. In the interval 2013-2020 both parameters have practically the same behavior, (iii) After 2020, both papers and patent application increase significantly, which is directly related to the appearance of the COVID-19 pandemic.

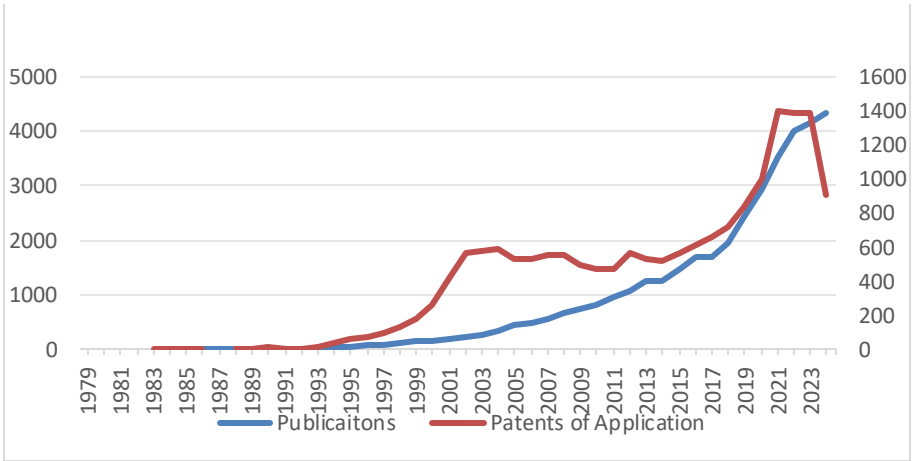

Figure 1. Annual publications and patents related to MD topic in the studied period 1979-2024.

Note: The patent application data for 2023 and 2024 has not yet been fully disclosed.

Other interesting aspect is to study funding availability. According to the dataset of papers (1979 - 2024), 20,502 papers were funded by 973 agencies and institutions from 76 countries/regions. It is a 53.6% of all the published papers. The annual trend of support of five representative funding agencies and institutions are shown in Figure 2. In the last 15 years there has been a sustained growth in funding for this type of research, especially after the COVID-19 pandemic, with funding from the National Natural Science Foundation of China (NNSCF) standing out in this last period.

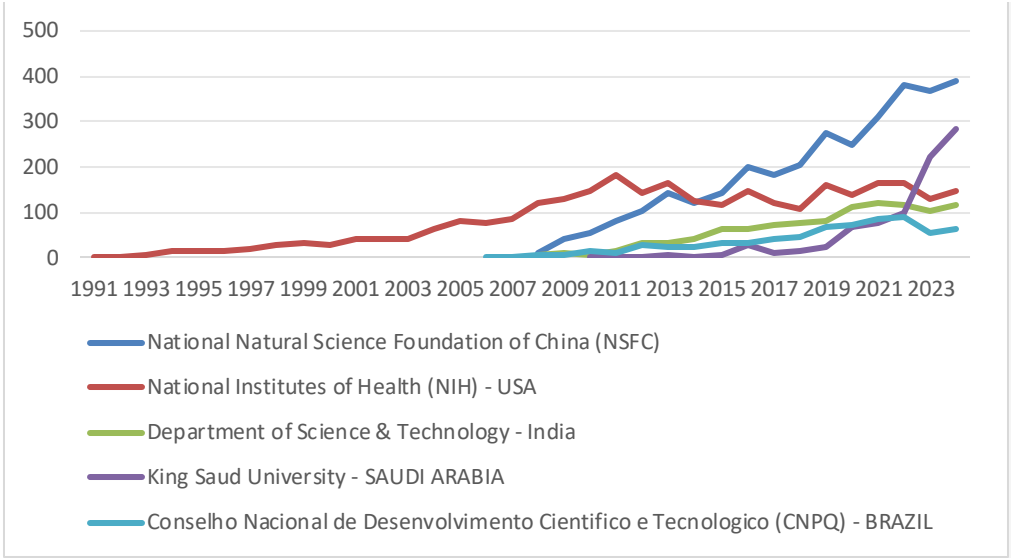

Figure 2. MD publication counts funded for five representative agencies and institutions in the studied period.

Research Hotspot Analysis

In Figure 3 we present a keyword co-occurrence network to reveal the research hotspot in MD. Six hotspots identified are as followed:

1. Drug Design and Discovery (Red Cluster)

This cluster highlights the application of molecular docking in identifying and optimizing potential drug candidates. Keywords such as "virtual screening", "drug design", "prediction", and "binding-affinity" underscore the reliance on computational methods for screening and improving ligand-target interactions. Terms like "flexibility", "force-field", and "genetic algorithm" point to methodological advancements to enhance docking precision and efficiency.

2. Structural and Biological Validation of Molecular Docking (Blue Cluster)

This cluster integrates structural biology, experimental validation, and biological activity evaluation to support molecular docking predictions. Core keywords include "crystal-structure", "x-ray", "fluorescence spectroscopy", and "protein-binding", reflecting the reliance on high-resolution structural data to model and verify molecular interactions. Simultaneously, terms such as "antibacterial", "antioxidant

activity", and "cytotoxicity" emphasize the application of docking techniques in identifying compounds with specific therapeutic properties. The presence of "in-vitro" further indicates the integration of computational predictions with experimental validation to confirm biological relevance.

3. Computational Methods and Algorithms (Yellow Cluster)
This cluster represents the methodological and algorithmic innovations in molecular docking studies. Keywords like "3D-QSAR", "pharmacophore", and "homology model" highlight the use of advanced modeling techniques to predict and optimize molecular interactions. The inclusion of "protein-coupled receptor", "GPCR", and "kinase inhibitors" demonstrates the diverse range of target molecules addressed by these computational methods.

4. Biological Activity and Therapeutic Applications (Green Cluster)
This cluster explores the application of docking in discovering and evaluating compounds for their biological activity and therapeutic potential. Terms such as "cancer", "resistance", "drug discovery", and "biological evaluation" reflect the use of docking for identifying bioactive molecules in disease-related contexts. Keywords like "activation", "target", and "expression" emphasize the subsequent experimental validation of docking results through pathway analysis and functional assays.

5. Neurological and Viral Diseases (Yellow-Green Cluster)
This cluster bridges molecular docking with research on neurological and viral diseases, addressing global health challenges. Key terms such as "Alzheimer's disease", "oxidative stress", "virus", and "SARS-CoV-2" suggest a focus on identifying therapeutic molecules for these diseases. The presence of "network pharmacology" and "pathogenesis" indicates a systems-level approach to understanding disease mechanisms and therapeutic interventions.

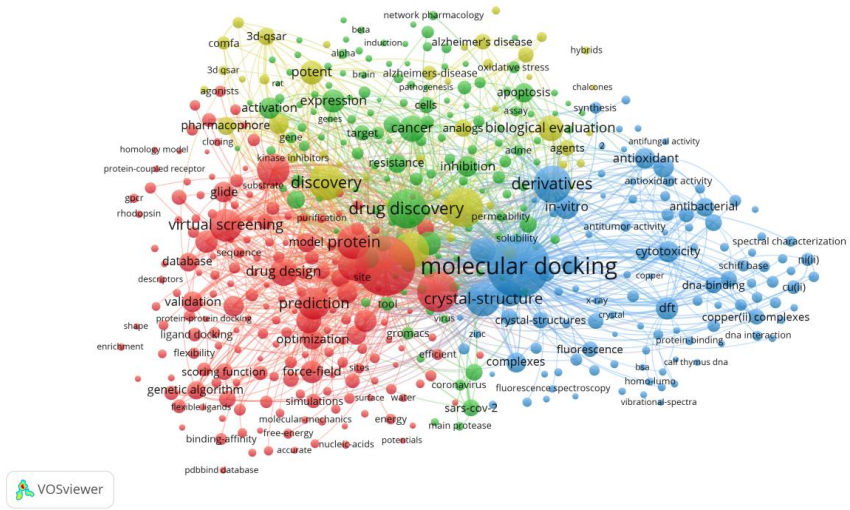

Figure 3. Keyword co-occurrence network in MD (Occurrences ≥ 100).

National/Regional Collaboration Networks

In Figure 4 we present the national collaboration in MD. A short description of these regions regarding to the MD techniques is shown in the following.

1. North America and East Asia as Pioneering Regions in MD Research

North America, led by the United States, and East Asia, dominated by China, emerge as the most influential regions in the global molecular docking research network. The United States maintains its leadership position through pioneering computational methodologies and consistent funding for drug discovery initiatives, while China has rapidly ascended in this domain, leveraging its strategic investment in bioinformatics and increasing collaboration with both developing and developed nations.

2. Europe's Collaborative Network in Multidisciplinary Research

European countries, notably Germany, Italy, and France, form a dense and interlinked cluster, underscoring the region's collaborative research culture. This interconnectedness stems from European Union funding initiatives, such as Horizon Europe, which promote cross-border partnerships and multidisciplinary projects. The European network's focus likely extends to fundamental research and the development of innovative molecular docking algorithms, emphasizing their application in diverse fields such as cancer therapeutics and personalized medicine. The prevalence of collaboration within this cluster highlights the synergistic potential of smaller countries working collectively with leading research nations.

3. South Asia and the Middle East as Emerging Contributors

The emergence of South Asia and the Middle East, represented prominently by India and Saudi Arabia, reflects the growing contributions of these regions to molecular docking research. These countries prioritize computational approaches to tackle region-specific health issues, such as neglected tropical diseases and antimicrobial resistance. India, in particular, has become a regional hub for bioinformatics research, leveraging its strong pharmaceutical industry and a rapidly expanding academic infrastructure. Saudi Arabia's position suggests strategic investments in life sciences, likely aimed at diversifying its research portfolio and fostering international collaborations in computational drug discovery.

4. Cross-Cluster Collaboration Driven by Global Health Challenges

The strong interconnectivity across clusters signifies a global effort to address pressing biomedical challenges through molecular docking. Two major themes emerge from these collaborations: the discovery of antiviral agents, accelerated by the COVID-19 pandemic, and the development of targeted therapies for cancer and other chronic diseases. These thematic focuses necessitate partnerships that bridge technological expertise, such as that found in North America and East Asia, with diverse research perspectives from Europe, South Asia, and the Middle East. The network visualization thus underscores the centrality of molecular docking as a

unifying research domain, fostering innovation through cross-border scientific exchange.

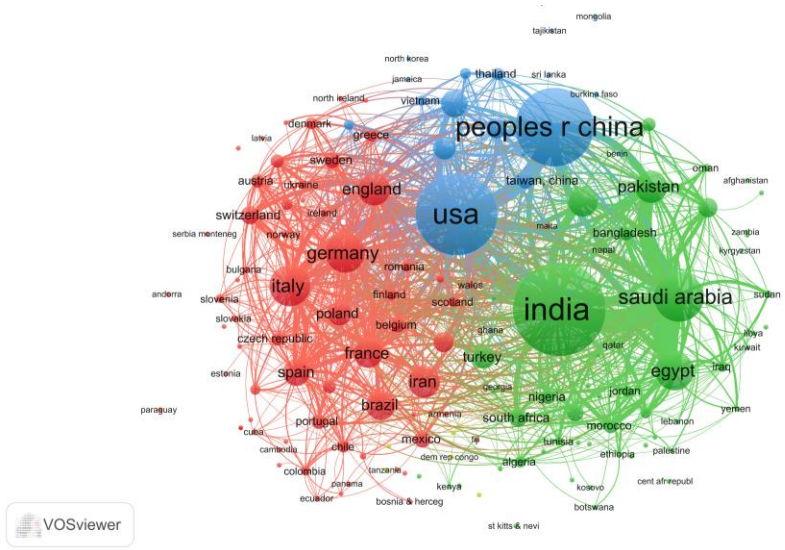

Figure 4. National/ regional collaboration in MD.

Institutions with the most patent applications

In Table 1 we present the Top 20 current assignee of patent in MD. Academic and research-focused organizations account for a significant share of patent filings, with The Regents of the University of California leading with 381 patents. Universities such as Sanskriti University, Harvard College, and Stanford University follow suit. These institutions dominate the rankings, reflecting their emphasis on foundational research and innovation. Novartis AG (119 patents), Bristol-Myers Squibb (92 patents), Genentech, Inc. (89 patents), and Allergan, Inc. (75 patents) are prominent pharmaceutical companies. Their focus is on translating molecular docking innovations into clinical applications. Research hospitals and non-profit institutions also feature prominently in this domain. Notable examples include:Dana-Farber Cancer Institute, Inc. (109 patents), City of Hope (84 patents), The General Hospital Corporation (70 patents).

Table 1. Top 20 current assignee of patent in MD in the studied period 1979-2024.

Current Assignee	*Patent Count*
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA	381
SANSKRITI UNIVERSITY	246
PRESIDENT AND FELLOWS OF HARVARD COLLEGE	137
THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY	131
NOVARTIS AG	119
DANA-FARBER CANCER INSTITUTE, INC.	109
MASSACHUSETTS INSTITUTE OF TECHNOLOGY	109
IMMUNOMEDICS, INC.	104
BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM	97
BRISTOL-MYERS SQUIBB COMPANY	92
YALE UNIVERSITY	91
THE SCRIPPS RESEARCH INSTITUTE	90
GENENTECH, INC.	89
CITY OF HOPE	84
THE BROAD INSTITUTE, INC.	79
ALLERGAN, INC.	75
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK	74
THE GENERAL HOSPITAL CORPORATION	70
VANDERBILT UNIVERSITY	70
DUKE UNIVERSITY	67

Conclusions

In this paper we present a comprehensive assessment about the application of MD technology in drug discovery through bibliometric and patent analysis, revealing research trends, technological hotspots, key participants and their collaboration networks. We found an increasing trend in the number of papers and patents in the field of MD in the studied period. Furthermore, a relationship has been found between the number of papers and patents. Also, a preliminary analysis of the funding of agencies and institutions for the support of research in MD was carried out. In our study, main countries and institutions with patents in the drug design have been identified. Our results can help to better understand the dynamic relationship between scientific work expressed in papers and the development of new drugs based on the number of patents granted. Future developments will focus on quantifying the relationships found and studying these relationships in detail for a case study related to the development of drugs for the treatment of COVID-19.

Acknowledgments

This work was supported by the key project of innovation fund from National Science Library
(Chengdu), Chinese Academy of Sciences (E3Z0000902).

References

Alkhazaleh, R., & Mykoniatis, K. (2024). Unveiling predictors influencing patent licensing: Analyzing patent scope in robotics and automation. *World Patent Information*, 77, 102276.

Banerjee, P., Gupta, B., & Garg, K. (2000). Patent statistics as indicators of competition an analysis of patenting in biotechnology. *Scientometrics*, 47(1), 95-116.

De Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C., & Dardenne, L. E. (2014). A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Information Sciences*, 289, 206-224.

Goodsell, D. S., Sanner, M. F., Olson, A. J., & Forli, S. (2021). The AutoDock suite at 30. *Protein Science*, 30(1), 31-43.

Liu, Z., Liu, Y., Zeng, G., Shao, B., Chen, M., Li, Z., ... & Zhong, H. (2018). Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere*, 203, 139-150.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14), 1639-1662.

Qiang L., Chen Y., Li X., Qi L., J. Gulín-González, Zhang Z (2019). A citation iteration method for publications and scientists' evaluation. (2021). *Data Science and Informetrics*, 1(2).

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3), 470-489.

Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., & McCammon, J. A. (2004). Discovery of a novel binding trench in HIV integrase. *Journal of medicinal chemistry*, 47(8), 1879-1881.

Schnecke, V., & Boström, J. (2006). Computational chemistry-driven decision making in lead generation. *Drug discovery today*, 11(1-2), 43-50.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein−ligand docking using GOLD. Proteins: *Structure, Function, and Bioinformatics*, 52(4), 609-623.