Computationally-Assisted Discovery and computational NMR Assignment of a Highly-binding PXR Natural Diterpenoids

Zhiwei Bian [#[a]], *Hongyi Li* [#[b]], *Jiabao Liu* [#[c]], *Xiaoying Liu*[a], *Shian Hu*[a], *Daneel Ferreira*[d], *Mark T. Hamann*[e], *Shengpeng Wang*[b], *Yeun-Mun Choo*[f], *and Xiaojuan Wang**[a]

[a]School of Pharmacy, Lanzhou University, Lanzhou 730000, Gansu, People′s Republic of China.

[b]State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau. Macau, People′s Republic of China.

[c]Department of Chemistry, The Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3H6, Canada.

[d] National Center for Natural Products Research, University of Mississippi, Oxford, MS 38677, USA.

[e] Drug Discovery and Biomedical Sciences, Medical University of South Carolina, Charleston, SC, US.

[f] Department of Chemistry, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia.

*Corresponding Author: *E-mail: Xiaojuan170@outlook.com

ABSTRACT：We report here the orchestration of molecular ion networking (MoIN) and a set of computationally assisted structural elucidation approaches in the discovery of a new class of 4,5-*seco*-abietane diterpenoids that possess anti-inflammatory bioactivity targeting the pregnane X receptor (PXR). Rubescens A represents the first in a new class of 4,5-*seco*-abietane diterpenoids, discovered from *Isodon. rubescens (I. rubescens)*. The molecule was discovered with the guidance of molecular ion networking (MoIN) analysis and untargeted pull down experiment. The structure was elucidated and determined using extensive spectroscopic analysis in conjunction with computationally assisted quantifiable structure elucidation tools. In the docking study, Rubescens A binds well to PXR, COX-2 and ATP binding domain of some protein kinases involved in inflammatory pathways.

INTRODUCTION

The PXR, a master regulator of drug metabolism, has essential roles in intestinal homeostasis and abrogating inflammation. Existing PXR ligands have substantial off target toxicity[1]. There is still considerable effort required to identify improved PXR legends.

Natural products have provided a diversity of valuable drugs because of complex structure and a wide range of promising bioactivities. It has been estimated that up to 50% of all present day therapeutics are derived from natural products[2]. *I. rubescens* in Chinese folk medicine is mainly rich diversity diterpenoids and has long history to treat jaundice hepatitis, acute cholecystitis, enteritis, and other inflammtory intestinal

inflammation[3]. Diterpenoid natural products cover a vast chemical diversity and include many medicinally and industrially relevant compounds[4]. Continuing our research towards the identification of new bioactive 4,5-*seco*-abietane diterpenoids, *I. rubescens* was phytochemically investigated.

Identification of new molecules with clear target early in the workflow was achieved by MoIN and untargeted LC/MS pull down experiment. LC/MS pull down experiment enables discover the compounds in the complex crude extract binding to target. MoIN technology can visualize observed molecules with molecular weight, it also shows the similarities as well as differences between two or more samples in which similar entities within the network are clustered together while disparate or unique entities are grouped separately[5]. With the application of these technologies, an unrepoted PXR binding diterpenoid was discovered.

X-ray diffraction is golden standard for unequivocal structural elucidation. When well-ordered crystals are not available for x-ray analysis, NMR spectroscopy is the most common structure-elucidation method. However, sometimes it is hard to distinguish isomers with similar spectra, alternative methods must be used[6]. Computational approaches, in conjunction with spectroscopic methods, provide a powerful and emerging method for the assignment of atom connectivity, relative configuration, and absolute configuration of complex molecules. Electronic circular dichroism (ECD) spectroscopy was applied here to determine the absolute configurations of the natural products. The protocol presented here involves orchestration of a number of computational methods in combination with

spectroscopic analyses, leading to the discovery and establishment of the well-defined structure for structurally novel natural products with clear binding target.

Computationally Assisted Discovery and Structural Elucidation.

Isolation of the novel chemical class was facilitated under the guidance of HPLC-HRMS/MS (positive mode)-based MoIN. The organized landscape of the MoIN was generated using Cytoscape (for details see Supporting Information) and showed several 4,5-*seco*-abietane diterpenoids compounds (**Figure 1**). Combined with untargeted pull down experiment which was applied to discover compounds binding to the PXR in the crude extract, an unreported and PXR binding diterpenoid signal (*m/z* 329.176) was discovered.

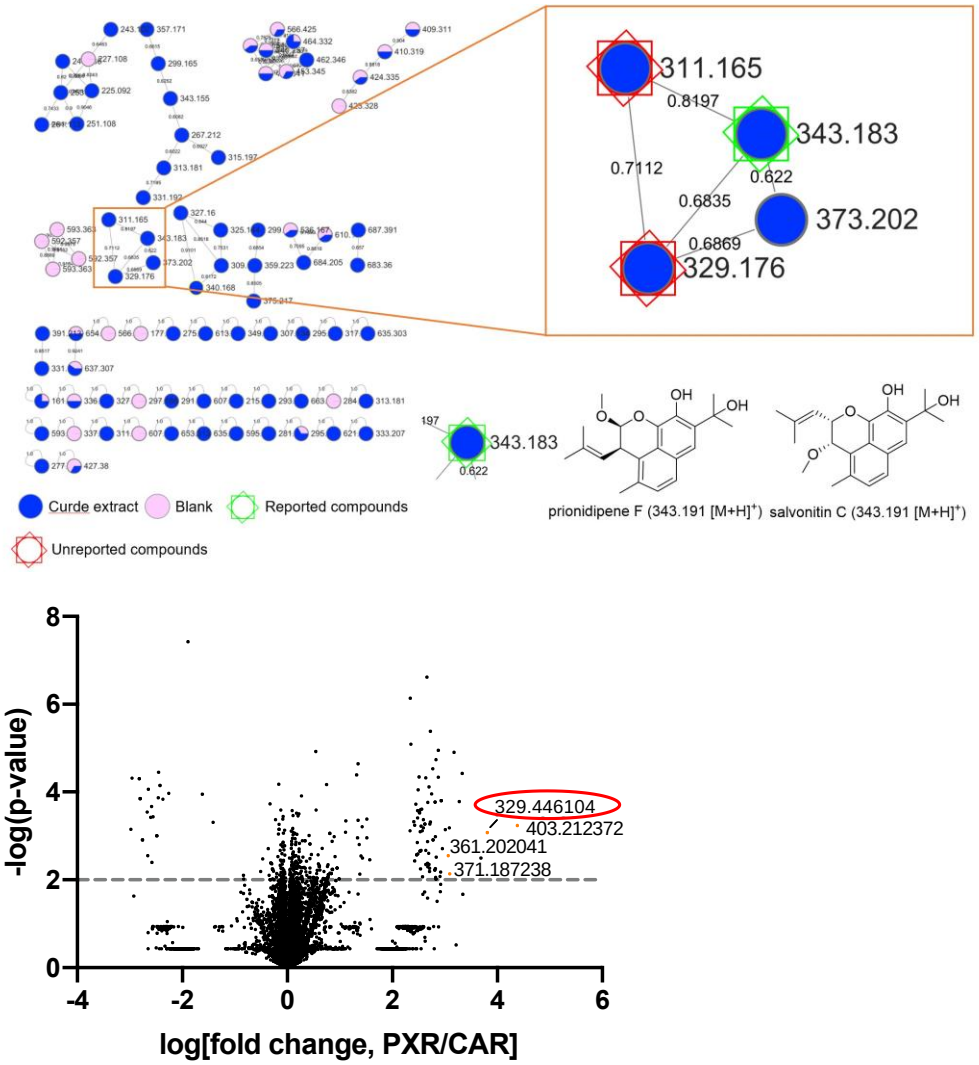

Figure 1. Untargeted pull down experiment and whole MoIN map for ethanol extract of *I. rubescens*. Zoom out part for the diterpenoids of the whole MoIN map.

Compound **1** was obtained as yellow amorphous powder, and had a molecular formula $C_{20}H_{24}O_4$ with nine degrees of unsaturation due to analysis of its positive ion mode HR-ESI-MS (*m/z* 329.1753 $[M+H]^+$, calcd 329.1753, 351.1578 $[M+Na]^+$, calcd 351.1572) and the ^{13}C NMR data. The ^1H-NMR spectrum of **1** (**Table 1**) displayed signals for five singlet methyls [δ_H 1.77 (3H, s, H-16), 1.78 (3H, s, H-17), 1.47 (3H, s, H-18), 1.62 (3H, s, H-19) and 2.51 (3H, s, H-20)], three aromatic protons [δ_H 7.15

(1H, d, J = 8.3 Hz, H-6), 7.52 (1H, d, J = 8.3 Hz, H-7) and 7.09 (1H, s, H-14)].The ^{13}C NMR, DEPT and HSQC spectra exhibited 20 carbon signals, including five methyls (δ_C 30.4, 30.4, 30.1, 28.7, 16.6), two methylene (δ_C 37.4, 37.1), three olefinic methines (δ_C 129.7, 125.3, 112.1), and ten nonprotonated carbons (δ_C 141.8, 137.3, 133.7, 132.2, 129.5, 128.8, 123.3, 123.2, 85.5, 76.5). 2D NMR experiments (COSY, HSQC, and HMBC) enabled the full assignments of all proton and carbon atoms. A naphthalenyl moiety was initially proposed by comparing the aromatic region of the ^{13}C NMR spectrum with the known natural product 15-hydroxysalprionin[7]. Observation of the HMBCs from H-14 (δ_H 7.09) to C-12 (δ_C 133.7) and C-11 (δ_C 141.8) permitted that C-12 and C-11 were oxidized. The HMBCs of H-16 (δ_H 1.77), H-17 (δ_H 1.78), and H-14 (δ_H 7.09) with C-15 (δ_C 76.5) and H-16 (δ_H 1.77) to C-13 (δ_C 137.3) confirmed the 2-hydroxyisopropyl group was ascribed to C-13. The HMBCs between the methyl at δ_H 2.51 and C-5 (δ_C 129.5) attached the methyl at C-5 position. Further analysis indicated an additional [5.5]- spiroketal skeleton in **1**. The COSY correlations of H-2 (δ_H 2.56, 2.72) with H-3 (δ_H 2.12, 2.48) suggested C-2 (δ_C 37.4) was connected to C-3 (δ_C 37.4). HMBC correlations of H-18 (δ_H 1.47), H-19 (δ_H 1.62) with C-4 (δ_C 85.5) attached the two methyls at C-4 position. C-3 is vicinal to C-4 since H-2 (δ_H 2.56, 2.72) displayed HMBC correlation with C-4 (δ_C 85.5). The ^{13}C NMR and DEPT date of C-1 (δ_C 123.2) indicated that two oxygen atoms connected to C-1. Key HMBC correlations from the anomeric H-3 (δ_H 2.56, 2.72) to C-1 (δ_C 123.2) and H-2 (δ_H 2.56, 2.72) to C-10 (δ_C 123.3) confirmed C-1 was connected to C-5. According to the molecular weight and unsaturation, as well as the previous analysis,

C-1 was connected to C-11 via an oxygen atom. By comparison of the calculated spectra to the experimental spectra (**Figure 3**), the absolute configuration of **1** was assigned as 1*R*.

Table 1. ^{1}H- and ^{13}C-NMR spectral data for **1** (in $CDCl_3$). [a]

Position	**1**	
	δ_C	δ_H
1	123.2	
2	37.4	2.72 (m, 1H)
		2.56 (m, 1H)
3	37.1	2.12 (m, 1H)
		2.48 (m, 1H)
4	85.5	
5	129.5	
6	129.7	7.15 (d, *J*= 8.3 Hz)
7	125.3	7.52 (d, *J*= 8.3 Hz)
8	132.2	
9	128.8	
10	123.3	
11	141.8	
12	133.7	
13	137.3	
14	112.1	7.09 (s)
15	76.5	
16	30.4	1.77 (s)
17	30.4	1.78 (s)
18	28.7	1.47 (s)
19	30.1	1.62 (s)
20	16.6	2.51 (s)

[a] δ in ppm, *J* in Hz. ^{1}H-NMR: 600 MHz, ^{13}C-NMR: 150 MHz.

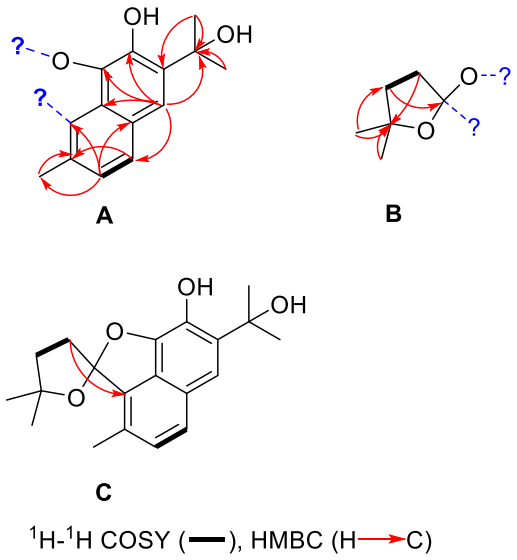

Figure 2. 2D structure elucidation of **1**.

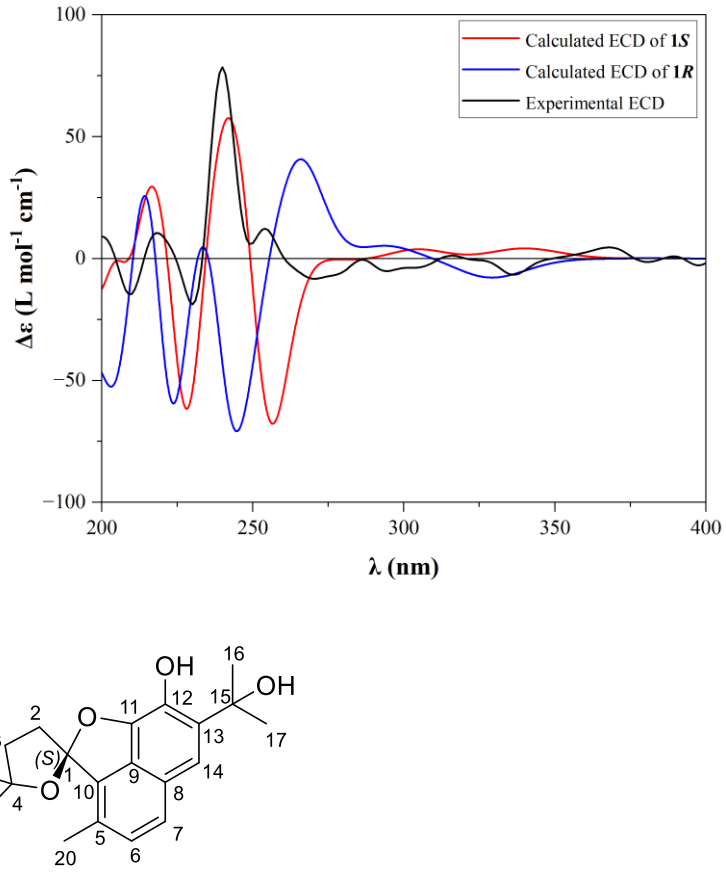

Figure 3. ECD and 3D structure elucidation of **1**.

Biosynthetic Pathway of 1

Structurally, compound **1** represents the first example of 4,5-*seco*-abietane diterpenoids with an unprecedented [5.5]-spiroketal skeleton. **1** could be derived from saproorthoquinone, a known *seco*-abietane diterpenoids co-occurred in *Salvia. prionitis*[8]. The intermediate A derived from the oxidation of saprorthoquinone was followed by reduction to get B. Subsequently, B underwent an intra-molecular cyclization between the hydroxyl at C-11 and the ketone group at C-1 lead to a hemiketal C by ketalation. Finally, C constructed spiroketal by further cyclization between C-4 and C-1 to form **1**.

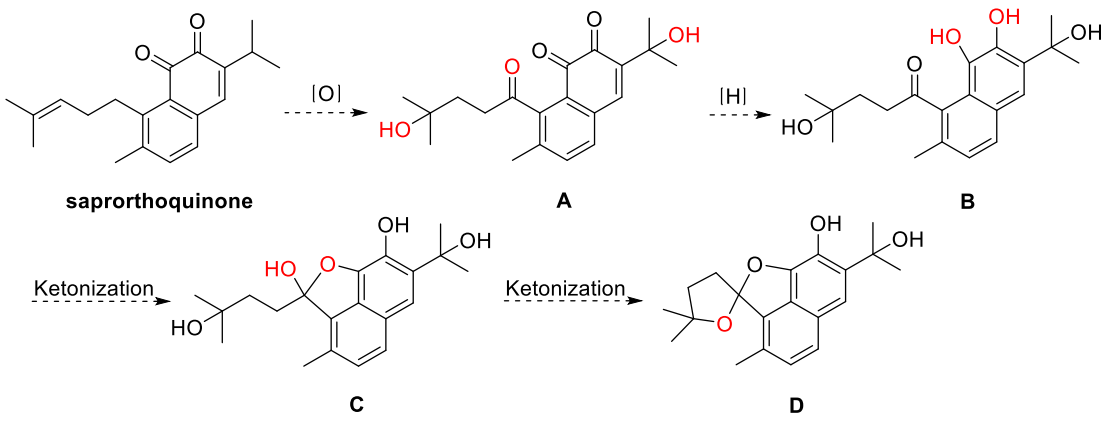

Figure 4. Proposed biosynthesis of **1**.

Molecular Docking Study

The computational molecular docking study was employed to assess the interaction between Rubescens A (**1**) and the PXR protein. Previous research has identified the PXR binding domain, which is situated at the core of the PXR protein. This binding

domain has the capability to bind various agonist compounds, ranging from larger molecules like paclitaxel and rifampicin to smaller molecules such as estradiol and oxadiazon (ref). Furthermore, the binding domain can be subdivided into four distinct subpockets, each with its preferred binding ligand: subpocket 1 - α-zearalanol, subpocket 2 - oxadiazon, subpocket 3 - pretilachlor, and subpocket 4 - estradiol (**Figure 5**).[9] To evaluate its binding preferences and affinity, compound **1** was docked separately into these subpockets.

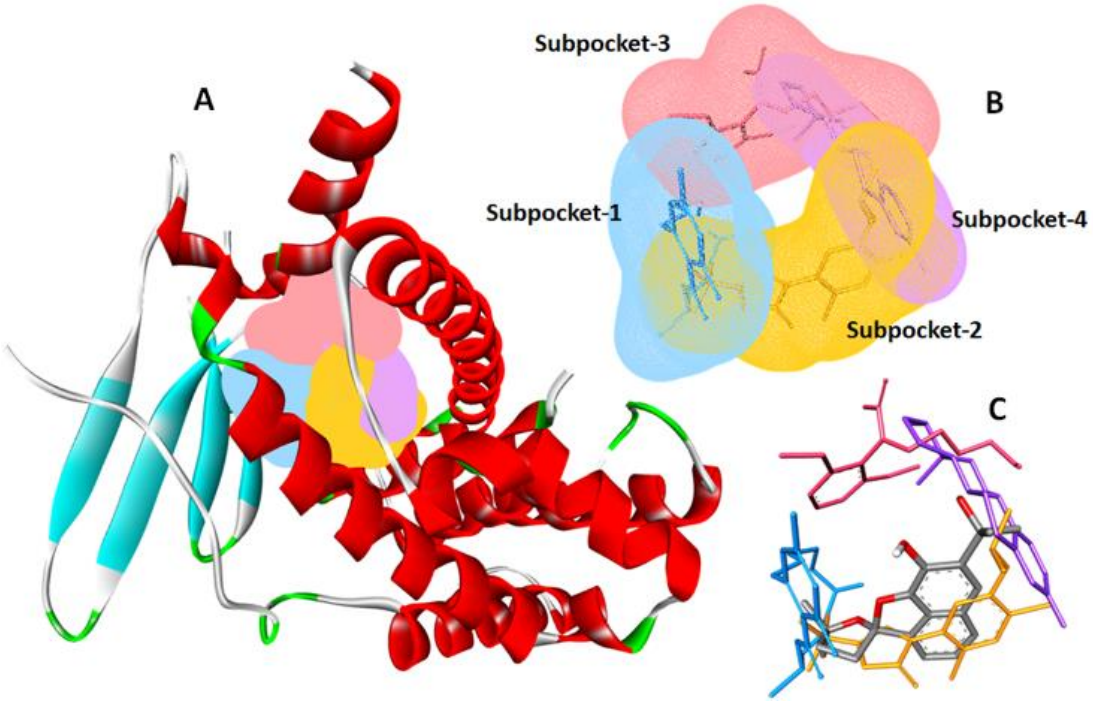

Figure 5. (A) The PXR protein structure and its corresponding binding domain; (B) Subpocket 1-4 subdivisions within the binding domain; (C) Comparative binding positions of Rubescens A (**1**) (-9.4 kcal/mol) and oxadiazon (orange; -9.8 kcal/mol) in relation to α-zearalanol (blue; -10.7 kcal/mol), pretilachlor (red; -7.0 kcal/mol), and estradiol (purple; 9.7 kcal/mol).

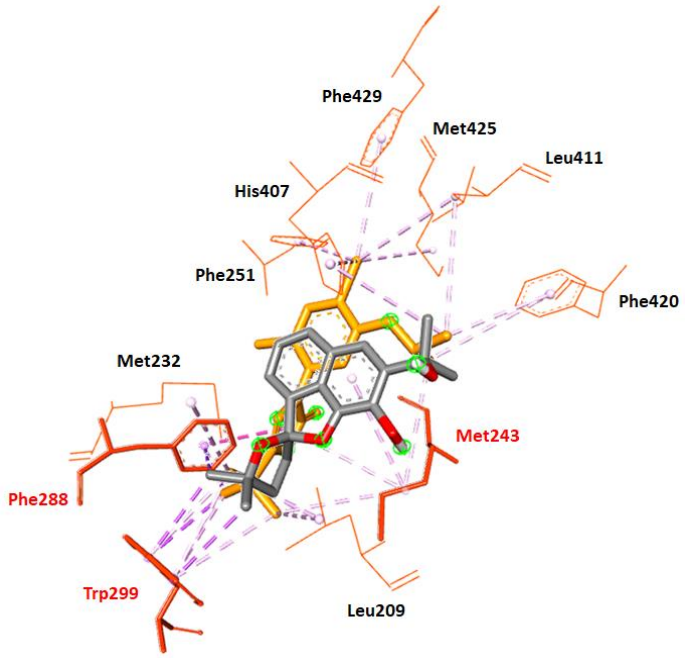

Figure 6. Interaction of **1** and oxadizon with the amino acid residues within the PXR binding domain.

The findings revealed that compound **1** exhibited a specific affinity for subpocket-2, displaying the lowest binding affinity of -9.4 kcal/mol when bound to subpocket-2. Interestingly, during the docking experiment in which **1** was intended to bind to subpocket-4, it instead exhibited a very strong binding affinity to subpocket-2. By closely analyzing the binding configurations of **1** and oxadiazon (the favored ligand for subpocket-2), it was observed that **1** occupied a binding space similar to that of oxadiazon. Both **1** and oxadiazon demonstrated π-σ non-bonding interactions with Phe288 and Trp299, which are located within the aromatic cage in the core of the PXR protein. Additionally, they both exhibited π-alkyl interactions with Met243 (**Figure 6**). These residues, Phe288 and Trp299, are known to establish interactions with the hydrophobic/aromatic components of the ligand and secure it within the

binding domain.[9] Oxadiazon was proven to be a potent PXR agonist with EC50 values of 9.0E-07 ± 2.4E-07 μM. The observed similarities in binding preference and ligand-protein interactions between compound **1** and oxadiazon suggest that **1** possesses a similar potential against PXR.

Table 2. In-silico binding affinity (kcal/mol).

	Rubescens_A	**ATP**
PXR	-9.4	-
COX-2	-9.9	-
TNF-α	-8.1	-
IκB	-9.6	-7.9
P38	-9.2	-8.6
AKT	-9	-9.2
PI3K	8.5	-7.7
JNK-1	-8.9	-8.6
JNK-2	-9.3	-7.6
MKK-4	-9.1	-9.5
MKK-7	-8.2	-7.5
JAK2	8.5	-7.7
STAT3	-6	-

Inflammatory responses can be triggered through various pathways, including the NFkB pathway, AKT/PI3K pathway, JAK-STAT pathway, and MAPK pathway (**Figure S1**).[10-14] The molecular docking study focused on evaluating the key targets of these pathways, which include COX-2, TNF-α, IκB, P38, AKT, PI3K, JNK-1, JNK-2, MKK-4, MKK-7, JAK-2, and STAT-3 (**Table 2**). Activation of these pathways leads to the expression of proinflammatory cytokines such as COX-2, TNF-α, IL-6, and IL-1β. In the molecular docking study, **1** was docked at the ligand binding domains (LBD) of COX-2, TNF-α, and STAT-3. The results showed that **1** had weak binding to STAT-3, moderate binding to TNF-α, but exhibited strong binding to COX-2. Upon examining the structure of **1**, it was observed that it

possesses a relatively small size, characterized by a planar naphtofuran moiety linked to a spirofuran moiety. This structural configuration allows **1** to fit comfortably within the tightly packed binding domain of COX-2 (**Figure S2**). The binding affinity of **1** is comparable to the known COX-2 inhibitor (meloxicam) and better than aspirin.

IκB, P38, AKT, PI3K, JNK-1, JNK-2, MKK-4, MKK-7, and JAK-2 are ATP-dependent protein kinases. **1** was docked at the ATP-binding domains of these kinases. The results demonstrated that **1** bound comparably well at the ATP-binding domain at these kinases (**Table 2**). The ATP-binding domain is known for its high conservation. The results revealed that **1** displays favorable binding within the restricted crevice of the ATP-binding domain (**Figure S3**). Additionally, the results suggest that **1** may function as an anti-inflammatory inhibitor of these protein kinases, thereby inhibiting the expression of proinflammatory cytokines and deactivating their downstream activities. Within the spectrum of protein kinases, **1** demonstrated the highest affinity for IκB (-9.3 kcal/mol), a key regulator preceding NFκB activation. Inhibiting the activity of IκB results in the attenuation of NFκB signaling, consequently leading to a decrease in the expression of downstream cytokines. In a conclusion, the molecular docking study provided valuable insights into the potential mechanisms of compound **1** in regulating inflammation, highlighting its potential as a PXR agonist, COX-2 inhibitor, and protein kinase inhibitor. The unique three-dimensional structure of **1**, characterized by a planar naphtofuran moiety linked to a spirofuran moiety, facilitates its binding within the inner core or narrow groove of the protein targets. These

findings contribute to a deeper understanding of the molecular interactions and therapeutic potential of **1** in the context of inflammatory processes.

ACKNOWLEDGMENT

This work was supported by grants from the National Natural Science Foundation of China (No. 82204227). Thanks to the Mississippi Center for Supercomputing Research.

REFERENCES

1. Dvořák, Z.; Kopp, F.; Costello, C. M.; Kemp, J. S.; Li, H.; Vrzalová, A.; Štěpánková, M.; Bartoňková, I.; Jiskrová, E.; Poulíková, K., Targeting the pregnane X receptor using microbial metabolite mimicry. *EMBO molecular medicine* **2020,** *12* (4), e11621.
2. Zou, Y.; Wang, X.; Sims, J.; Wang, B.; Pandey, P.; Welsh, C. L.; Stone, R. P.; Avery, M. A.; Doerksen, R. J.; Ferreira, D., Computationally assisted discovery and assignment of a highly strained and PANC-1 selective alkaloid from Alaska's deep ocean. *Journal of the American Chemical Society* **2019,** *141* (10), 4338-4344.
3. Wan, J.; Liu, M.; Jiang, H.-Y.; Yang, J.; Du, X.; Li, X.-N.; Wang, W.-G.; Li, Y.; Pu, J.-X.; Sun, H.-D., Bioactive ent-kaurane diterpenoids from Isodon serra. *phytochemistry* **2016,** *130*, 244-251.
4. Hopwood, D. A., *Natural product biosynthesis by microorganisms and plants*. Academic Press: 2012.
5. Xu, Y.; Li, J.; Mao, H.; You, W.; Chen, J.; Xu, H.; Wu, J.; Gong, Y.; Guo, L.; Liu, T.; Li, W.; Xu, B.; Xie, J., Structural annotation, semi-quantification and toxicity prediction of pyrrolizidine alkaloids from functional food: In silico and molecular networking strategy. *Food and Chemical Toxicology* **2023,** *176*, 113738.
6. Liu, Y.; Saurí, J.; Mevers, E.; Peczuh, M. W.; Hiemstra, H.; Clardy, J.; Martin, G. E.; Williamson, R. T., Unequivocal determination of complex molecular structures using anisotropic NMR measurements. *Science* **2017,** *356* (6333), eaam5349.
7. Chen, W.-F.; Wong, L.-L.; Zhang, X.; Zhou, F.-X.; Xia, F.; Tang, L.-P.; Li, X., New 4,5-seco-20(10 → 5)-abeo-Abietane Diterpenoids with Anti-Inflammatory Activity from Isodon lophanthoides var. graciliflorus (Benth.) H.Hara. *Chemistry & Biodiversity* **2019,** *16* (6), e1900206.

8. Li, L.; Zhou, M.; Xue, G.; Wang, W.; Zhou, X.; Wang, X.; Kong, L.; Luo, J., Bioactive seco-abietane rearranged diterpenoids from the aerial parts of Salvia prionitis. *Bioorganic Chemistry* **2018,** *81*, 454-460.
9. Delfosse, V.; Huet, T.; Harrus, D.; Granell, M.; Bourguet, M.; Gardia-Parège, C.; Chiavarina, B.; Grimaldi, M.; Le Mével, S.; Blanc, P.; Huang, D.; Gruszczyk, J.; Demeneix, B.; Cianférani, S.; Fini, J. B.; Balaguer, P.; Bourguet, W., Mechanistic insights into the synergistic activation of the RXR-PXR heterodimer by endocrine disruptor mixtures. *Proc Natl Acad Sci U S A* **2021,** *118* (1).
10. Domitrović, R.; Potočnjak, I., A comprehensive overview of hepatoprotective natural compounds: mechanism of action and clinical perspectives. *Arch Toxicol* **2016,** *90* (1), 39-79.
11. Rayego-Mateos, S.; Morgado-Pascual, J. L.; Opazo-Ríos, L.; Guerrero-Hue, M.; García-Caballero, C.; Vázquez-Carballo, C.; Mas, S.; Sanz, A. B.; Herencia, C.; Mezzano, S.; Gómez-Guerrero, C.; Moreno, J. A.; Egido, J., Pathogenic Pathways and Therapeutic Approaches Targeting Inflammation in Diabetic Nephropathy. *International journal of molecular sciences* **2020,** *21* (11).
12. Shultz, R. B.; Zhong, Y., Minocycline targets multiple secondary injury mechanisms in traumatic spinal cord injury. *Neural regeneration research* **2017,** *12* (5), 702-713.
13. Do, H. T. T.; Bui, B. P.; Sim, S.; Jung, J. K.; Lee, H.; Cho, J., Anti-Inflammatory and Anti-Migratory Activities of Isoquinoline-1-Carboxamide Derivatives in LPS-Treated BV2 Microglial Cells via Inhibition of MAPKs/NF-κB Pathway. *International journal of molecular sciences* **2020,** *21* (7).
14. Coggins, M.; Rosenzweig, A., The fire within: cardiac inflammatory signaling in health and disease. *Circulation research* **2012,** *110* (1), 116-25.