JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: *J. Chem. Inf. Model.* 2018, 58, 1182−1193

pubs.acs.org/jcim

Cheminformatic Insight into the Differences between Terrestrial and Marine Originated Natural Products

Jun Shang,[†,‡,§] Ben Hu,[‡] Junmei Wang,[∥] Feng Zhu,[†] Yu Kang,[†] Dan Li,[†] Huiyong Sun,[†] De-Xin Kong,[*,‡] and Tingjun Hou[*,†,§]

[†]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

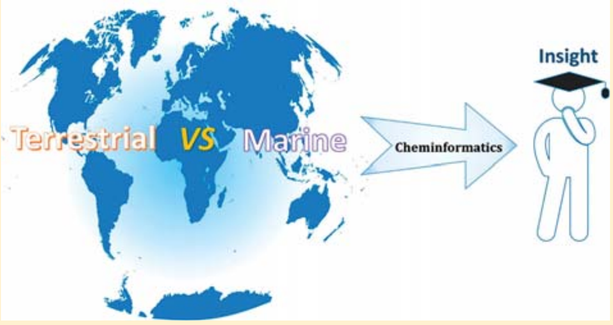
[‡]State Key Laboratory of Agricultural Microbiology and Agricultural Bioinformatics, Key Laboratory of Hubei Province, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[§]State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang 310058, China

[∥]Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States

Supporting Information

ABSTRACT: This is a new golden age for drug discovery based on natural products derived from both marine and terrestrial sources. Herein, a straightforward but important question is "what are the major structural differences between marine natural products (MNPs) and terrestrial natural products (TNPs)?" To answer this question, we analyzed the important physicochemical properties, structural features, and drug-likeness of the two types of natural products and discussed their differences from the perspective of evolution. In general, MNPs have lower solubility and are often larger than TNPs. On average, particularly from the perspective of unique fragments and scaffolds, MNPs usually possess more long chains and large rings, especially 8- to 10-membered rings. MNPs also have more nitrogen atoms and halogens, notably bromines, and fewer oxygen atoms, suggesting that MNPs may be synthesized by more diverse biosynthetic pathways than TNPs. Analysis of the frequently occurring Murcko frameworks in MNPs and TNPS also reveals a striking difference between MNPs and TNPs. The scaffolds of the former tend to be longer and often contain ester bonds connected to 10-membered rings, while the scaffolds of the latter tend to be shorter and often bear more stable ring systems and bond types. Besides, the prediction from the naïve Bayesian drug-likeness classification model suggests that most compounds in MNPs and TNPs are drug-like, although MNPs are slightly more drug-like than TNPs. We believe that MNPs and TNPs with novel drug-like scaffolds have great potential to be drug leads or drug candidates in drug discovery campaigns.



INTRODUCTION

Natural products have been regarded as a rich source of novel drug leads,[1] but the advances of combinatorial chemistry and high-throughput screening (HTS) techniques have shifted the focus of the pharmaceutical industry from natural products to purely synthetic compounds in the past two decades.[2] It was highly expected that combinatorial chemistry techniques could provide most drug-like structures needed for successful lead discovery campaigns. However, the large-scale application of combinatorial chemistry and HTS has not yet boosted the approval rate of new molecular entities (NMEs) significantly.[3] It has been recognized that the chemicals synthesized by combinatorial chemistry techniques usually have limited structural diversity, which may explain why many HTS experiments yielded disappointing outcomes even for large screening collections.[4] The studies reported by Tian and co-workers show that the natural products from traditional Chinese medicines (TCM) exhibit much higher structural complexity than the synthetic compounds in commercial screening collections, and the enrichment of drug-like molecules in TCM is much higher than that in synthetic screening databases.[5−7] Our findings suggest that natural products contain drug-like scaffolds not found elsewhere and TCM is therefore an excellent source for identifying drug leads. Actually, half of the NMEs launched over the last 30 years were derived from natural products.[8] Recently, by discussing a number of successful drug discovery projects from the early development stages to the clinical trials, Crane and Gademann uncovered the fact that the key biological parameters, such as potency and selectivity, of natural products can be retained or even improved by chemical modification of the parent natural products.[9] Therefore, the analysis of structural features of natural products to identify the biologically active molecular

Received: March 5, 2018
Published: May 24, 2018

fragments will play an important role in natural product-based drug design and drug development.

Historically, in many Asian and African countries, traditional medicines from terrestrial herbs have been used for primary health care for more than thousands of years. Moreover, almost all of the available natural product-derived drugs originated from terrestrial sources. However, considering that seas and oceans occupy almost 70% of the earth's surface, natural products derived from marine sources should not be neglected. As a matter of fact, the distribution of marine natural products (MNPs) is quite extensive, even though the majority of MNPs has not been explored due to the limited accessibility of the marine environment. With the advances in sampling and structure determination technologies,[10,11] a large number of new MNPs have been discovered and applied to the pharmaceutical and cosmeceutical industries in the past two decades.[12−14] It is believed that MNPs may become a rich source of drug-like molecules for exploring potential therapeutics.[15]

Chemical compounds in marine and terrestrial organisms have continuously evolved to interact efficiently with the specific biological targets in marine and terrestrial organisms, and therefore marine and terrestrial compounds may occupy different biologically relevant chemical spaces. Moreover, considering that the biochemical reactions of marine and terrestrial organisms triggered by distinct growth environments should be different, it is a reasonable assumption that the marine and terrestrial secondary metabolites are considerably different.[16−20] Here, a straightforward but important question is raised: *what are the structural differences between MNPs and terrestrial natural products (TNPs)?* In 2013, Muigg et al. explored the chemical space of 3802 marine and 29,620 terrestrial molecules defined by physicochemical properties, and they found that despite considerable overlap, the specific regions in the chemical space occupied by these two data sets can be roughly distinguished.[21] Kong and co-workers analyzed the Murcko frameworks and ClogP for MNPs and TNPs, and they found that most scaffolds (71.02%) found in MNPs are unique; but, the drug development potential of MNPs may be hindered by their relatively higher hydrophobicity compared with that of TNPs.[22,23] To the best of our knowledge, a systematic exploration of the structural features of MNPs and comparison of the structural features and drug-likeness of MNPs and TNPs have not been reported.

In this study, with an aim to elucidate the structural differences between TNPs and MNPs, the distributions of important physicochemical properties, structural features, and drug-likeness of TNPs and MNPs were analyzed and compared. First, the distributions of up to 60 important physicochemical molecular properties for TNPs and MNPs were analyzed. Then, the structural features of MNPs and TNPs represented by four types of fragments, including chain assemblies, ring assemblies, Murcko frameworks,[24] and RECAP (Retrosynthetic Combinatorial Analysis Procedure) fragments,[25] were examined. Finally, the drug-likeness of TNPs and MNPs was evaluated by using a drug-likeness classifier established by the naïve Bayesian classification (NBC) technique.[5,26] The findings of our analyses will inspire a mind with cheminformatics and evolutionary biology intuition to gain a deeper understanding why MNPs could bring a bump harvest in this golden age of natural product drug discovery.[27]

METHODS

Preparation of Databases. MNPs were obtained from DMNP (Dictionary of Marine Natural Products, version 2015),[27] and TNPs were obtained by removing the molecules in DMNP from DNP (Dictionary of Natural Products, version 2015)[28] as we did in the previous study.[22] Then, the structures in MNPs and TNPs were standardized by keeping the largest fragments, removing inorganic and tiny (molecular weight <80) compounds, adding hydrogen atoms, and removing duplicated molecules by using Pipeline Pilot 8.5 (PP 8.5).[29,30] Two standardized data sets, TNPs_origin (151 609 molecules) and MNPs_origin (35 883 molecules), were referred to as the original data sets.

Databases of CMC (Comprehensive Medicinal Chemistry, version 2005), MDDR (MACCS-II Drug Data Report, version 2004), ACD (Available Chemical Database), CNPD (Chinese Natural Products Database, version 2005),[31] TCMD (Traditional Chinese Medicine Database, NeoSuite, version 2009),[32] and TCMCD (Traditional Chinese Medicine compound database, version 2015) developed by our group[33,34] were also prepared by the same standardization pipeline mentioned above.

Besides, according to the studies reported by Koch et al. and Grabowski et al., about 14.8% and 17% of NPs have sugar units.[35,36] The compounds in TNPs and MNPs were deglycosylated using the filters of substructural patterns defined with an in-house pipeline pilot protocol. First, the sugar-like (O- and N-glycosides, including both furanoses and pyranoses) fragments were extracted from the compounds by a substructural search, and the compounds containing glycosides, including all kinds of acyl-O, acyl-N, acetyl, alduronic acid, and other derivatives, were identified. Then, each terminal cyclic glucoside was replaced by a hydroxyl or amino group for the O- or N-glycosides, respectively. The process was executed recursively until all the terminal glycosides were removed. A total of 28 955 compounds (14.03%) in TNPs and 4021 compounds (8.33%) in MNPs were deglycosylated by removing the 1−12 sugar moieties from the parent compounds. The carbohydrates were also removed from the databases. Moreover, after accomplishing the deglycosylation process, another round of duplication check was performed. At last, we obtained 32 937 unique natural products without any sugar unit from DMNP (MNPs_nosugar) and 132071 from DNP (TNPs_nosugar). In the following analyses, for the sake of convenience, we referred to "MNPs_nosugar" as MNPs and "TNPs_nosugar" as TNPs.

Analysis of 60 Important Physicochemical Properties. A total of 60 molecular descriptors listed in Table S1 of the Supporting Information were calculated in PP 8.5. These descriptors can be used to characterize the important physicochemical properties of a molecule, such as bulkiness, solubility, and hydrophobicity, hydrogen bonding capability, composition of elements, bonds, and rings, etc. Then, the data were processed using the *Basic Statistics by Category* component in PP 8.5.

The statistical significance of the difference between the two population means for each molecular property at the significance level of α = 0.05 and critical value $Z_{0.025}$ = 1.96 was evaluated by the u-test or z-test when the variances are known and the sample size is large. The formula to calculate the u-value is as follows:

$$u = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

where $\bar{x}_1$ and $\bar{x}_2$ represent the means of a molecular property for MNPs and TNPs, respectively; S_1 and S_2 represent the standard deviations of a property for MNPs and TNPs, respectively; n_1 and n_2 represent the total numbers of molecules for MNPs and TNPs, respectively. As a result, when the calculated $|u| > 1.96$, this property is considered to have significant difference between TNPs and MNPs. A positive *u*-value means that the property values of MNPs are higher than those of TNPs, and on the contrary, a negative *u*-value means that the properties of MNPs are lower than those of TNPs.

Analysis of Four Types of Fragment Representations. Four types of fragment representations, including chain assemblies, ring assemblies, Murcko frameworks, and RECAP fragments, were used to analyze the scaffolds of MNPs and TNPs. The ring assembly is the continuous rings without any linker in a molecule or fragment, including fused rings and bridged ring systems. The chain assembly is a set of contiguous chain atoms in a molecule or fragment, which include any ring atom that terminates a chain. The Murcko framework is the union of the linkers and ring systems in a molecule. The RECAP fragments are the fragments cleaved from a molecule at the bonds based on 11 predefined bond cleavage rules stemmed from usual chemical reactions. The first three types of fragments were generated by using the *Generate Fragments* component in PP 8.5, and the last one was generated by using the *sdfrag* command in MOE (Molecular Operating Environment).[37] Then the generated fragments were processed by using the *Merge Molecules*, the *Property Value Threshold Filter*, and the *Basic Statistics by Category* components in PP 8.5.

Drug-Likeness Analyses Based on Physicochemical Properties, Structures, and Drug-Likeness Prediction model. The Lipinski's Rule-of-5 (Ro5) implemented in the *Custom Manipulator (PilotScript)* component in PP 8.5 was used to evaluate the property-based drug-likeness of MNPs, TNPs, CMC, MDDR, CNPD, TCMD, and TCMCD.[38] It should be noted that Ro5 can only be used as a qualitative estimator of the absorption and permeability capability of a molecule, but it does not have good capability to distinguish drug-like from nondrug-like molecules.[39−42]

Then, in order to understand the drug-likeness of MNPs and TNPs on the basis of structural features, the scaffold architectures of TNPs and MNPs, characterized by Murcko frameworks, were compared with those of CMC by the *Molecular Similarity* component in PP 8.5. The structural diversity of Murcko frameworks was analyzed by the tree maps generated by the TreeMap software,[43] which can highlight both the structural diversity and the distribution of fragments. Different from the traditional methods using the tree structures by a graph with the root node and children nodes from the top to the bottom, tree maps proposed by Shneiderman uses circles or rectangles in a 2D space-filling way to delegate one property of clustered data sets with clearly intuitive visualization.[44] First, the unique Murcko frameworks were clustered by using the *cluster molecules* component in PP 8.5 based on the ECFP_4 fingerprints.[45−47] Then, the SDF file of the clustered scaffolds for each standardized data set was used as the input of the TreeMap software. In the tree maps, the area of each square is proportional to the scaffold frequency.

Moreover, in order to gain a deeper insight into the drug-likeness of TNPs and MNPs, the drug-likeness classification model established by the NBC technique based on 21 molecular physicochemical properties (as shown in Table 1) and the LCFP_6 fingerprints reported by Tian et al. was used to identify the drug-like molecules in TNPs and MNPs.[5]

■ RESULTS AND DISCUSSION

Differences of Physicochemical Properties between TNPs and MNPs. Among the 60 studied physicochemical properties, 42 including molecular weight, solubility and PSA show significances within the two data sets based on means, medians, and modes (Table 1). The *u*-test was employed to evaluate the statistical significance of the difference between two population means for each molecular property at the significance level of $\alpha = 0.05$ and the critical value $Z_{0.025} = 1.96$. When the absolute value of the calculated *u* is higher than 1.96, the analyzed property is considered to have significant difference between TNPs and MNPs. The important properties are interpreted from an evolutionary perspective as follows. The property-based drug-likeness of TNPs and MNPs is also compared in this part.

Molecular Sizes. As shown in Table 1, the average volume and surface area of MNPs are prominently higher than those of TNPs. That is to say, molecules collected from seas and oceans are usually larger than those collected from the land. The analysis of molecular weight also supports this phenomenon. Certainly, relatively larger molecules in MNPs may bring more challenge to cell permeation and intestinal absorption.[28] More reasons and views can be found in the following analysis of their N_R (number of rings) properties.

Molecular Solubility. In the studied properties, besides logS (molecular solubility), AlogP, and logD are also related to solubility. As shown in Table 1, the *u* values of AlogP and logD are 45.821 and 44.136, respectively, indicating that their means for MNPs are significantly higher than those for TNPs. In our previous study, we also observed that, on average, MNPs are more hydrophobic and less soluble than TNPs.[22] A well accepted explanation is that, to live in the marine environment freely and independently, halobios need to keep their metabolites more hydrophobic to avoid the loss of nutrients. Owing to being more hypoxic in the ocean than on the land, the average number of oxygen atoms of MNPs is much lower than that of TNPs, resulting in relatively lower solubility but higher hydrophobicity of MNPs.[22,48,49] As we know, hydrophobicity is closely related to the ADME (absorption, distribution, metabolism, and excretion) properties of molecules.[40,50−54] Higher hydrophobicity of MNPs may have profound effects on the success rate of turning initial hits into leads.

Element Compositions. As shown in Table 1, except for oxygen, the average atom numbers of the other elements in MNPs are higher than those in TNPs. Meanwhile, except for fluorine, the atom numbers of the other elements shown in Table 1 have pronounced differences between MNPs and TNPs. Obviously, more nonfluorine halogens are detected in MNPs because the oceans provide the largest source of biogenic organohalogens.[55,56] It is known that natural products have less nitrogen, halogen and sulfur atoms while more oxygen atoms than synthetic compounds and drugs.[55,57] Feher and coworkers observed that nitrogen, sulfur and halogen atoms are often introduced in synthetic reactions to make combinatorial synthesis more efficient.[48] As shown in Table 1, however, the

Table 1. Statistics of the 42 Molecular Descriptors for the 32 937 MNPs and 132 071 TNPs

	MNPs						TNPs						
descriptors[a]	mean	std dev	min	max	median	mode	mean	std dev	min	max	median	mode	u
V_M	294.112	155.991	25.380	3535.640	263.760	238.040	273.885	133.797	25.030	2261.050	245.930	180.760	21.632
SA	419.900	222.052	43.130	5519.490	375.310	253.170	391.933	191.892	49.490	3445.900	350.060	340.060	20.986
PSA	92.867	81.872	0.000	2250.360	72.830	20.230	96.392	80.061	0.000	1743.710	79.150	46.530	−7.020
***f*PSA**	0.223	0.126	0.000	1.000	0.203	0.000	0.245	0.117	0.000	1.000	0.230	0.000	−28.462
AlogP	4.071	3.394	−8.812	53.473	3.692	4.366	3.142	2.839	−19.902	41.417	2.856	2.447	45.821
logD	3.682	3.510	−36.321	53.473	3.390	3.814	2.756	2.965	−22.282	41.417	2.575	2.671	44.136
logS	−5.893	3.895	−60.286	5.452	−5.315	−4.832	−4.675	3.146	−45.809	15.740	−4.086	−3.877	−52.621
MW	420.702	221.301	32.042	5,033.753	379.406	304.467	396.057	199.323	32.042	3748.598	357.357	264.317	18.432
$\boldsymbol{N}_{\mathrm{Hdon}}$	2.354	2.939	0.000	73.000	2.000	1.000	2.370	2.827	0.000	57.000	2.000	1.000	−0.891
$\boldsymbol{N}_{\mathrm{N+O}}$	5.847	4.980	0.000	132.000	5.000	4.000	6.229	4.890	0.000	104.000	5.000	4.000	−12.508
N_a	29.063	15.505	2.000	352.000	26.000	24.000	28.431	14.218	2.000	268.000	26.000	24.000	6.729
N_b	30.704	16.521	1.000	351.000	28.000	24.000	30.706	15.632	1.000	299.000	28.000	27.000	−0.013
N_{RB}	14.822	10.890	0.000	171.000	15.000	0.000	17.290	10.976	0.000	204.000	17.000	6.000	−36.749
$\boldsymbol{N}_{\mathrm{rotB}}$	7.598	8.519	0.000	152.000	5.000	1.000	5.608	6.196	0.000	120.000	4.000	2.000	39.852
$\boldsymbol{N}_{\mathrm{aroB}}$	4.352	6.554	0.000	120.000	0.000	0.000	5.841	7.440	0.000	120.000	5.000	0.000	−35.869
$\boldsymbol{N}_{\mathrm{R}}$	2.641	1.938	0.000	32.000	2.000	2.000	3.275	2.090	0.000	32.000	3.000	3.000	−52.232
$\boldsymbol{N}_{\mathrm{aroR}}$	0.768	1.161	0.000	20.000	0.000	0.000	1.007	1.283	0.000	20.000	1.000	0.000	−32.792
N_{Rasb}	1.284	0.964	0.000	20.000	1.000	1.000	1.442	1.033	0.000	26.000	1.000	1.000	−26.301
$\boldsymbol{N}_{\mathrm{R3}}$	0.089	0.318	0.000	3.000	0.000	0.000	0.085	0.308	0.000	6.000	0.000	0.000	2.054
$\boldsymbol{N}_{\mathrm{R4}}$	0.008	0.087	0.000	2.000	0.000	0.000	0.013	0.118	0.000	5.000	0.000	0.000	−9.835
$\boldsymbol{N}_{\mathrm{R5}}$	0.664	0.876	0.000	10.000	0.000	0.000	0.711	0.922	0.000	16.000	0.000	0.000	−8.571
$\boldsymbol{N}_{\mathrm{R6}}$	1.658	1.533	0.000	28.000	1.000	1.000	2.301	1.782	0.000	25.000	2.000	3.000	−65.872
$\boldsymbol{N}_{\mathrm{R7}}$	0.050	0.257	0.000	5.000	0.000	0.000	0.069	0.270	0.000	4.000	0.000	0.000	−12.045
$\boldsymbol{N}_{\mathrm{R8}}$	0.014	0.127	0.000	3.000	0.000	0.000	0.011	0.107	0.000	2.000	0.000	0.000	3.058
$\boldsymbol{N}_{\mathrm{R9+}}$	0.159	0.385	0.000	4.000	0.000	0.000	0.084	0.316	0.000	8.000	0.000	0.000	32.861
$N\mathrm{c}$	40.460	24.389	0.000	518.000	37.000	38.000	36.602	19.517	2.000	320.000	33.000	28.000	26.657
N_{Casb}	11.147	7.469	0.000	108.000	11.000	1.000	12.278	7.224	1.000	124.000	12.000	12.000	−24.733
N_{steA}	0.005	0.137	0.000	7.000	0.000	0.000	0.012	0.228	0.000	15.000	0.000	0.000	−7.002
N_{steB}	1.780	2.037	0.000	59.000	1.000	0.000	1.356	1.646	0.000	37.000	1.000	0.000	35.074
N_{douB}	5.189	3.839	0.000	63.000	4.000	3.000	5.678	4.217	0.000	84.000	5.000	4.000	−20.280
N_{triB}	0.048	0.328	0.000	6.000	0.000	0.000	0.031	0.249	0.000	7.000	0.000	0.000	8.778
Corg	29.063	15.505	2.000	352.000	26.000	24.000	28.431	14.218	2.000	268.000	26.000	24.000	6.729
C_H	32.771	20.817	0.000	376.000	30.000	32.000	28.884	16.107	0.000	288.000	26.000	22.000	31.612
C_C	22.822	11.949	0.000	219.000	21.000	20.000	22.120	10.537	0.000	176.000	20.000	20.000	9.764
C_N	0.988	2.036	0.000	60.000	0.000	0.000	0.599	1.643	0.000	48.000	0.000	0.000	32.184
C_O	4.859	4.170	0.000	117.000	4.000	2.000	5.630	4.365	0.000	104.000	5.000	4.000	−29.761
C_F	0.004	0.117	0.000	12.000	0.000	0.000	0.003	0.122	0.000	16.000	0.000	0.000	1.065
C_P	0.008	0.103	0.000	3.000	0.000	0.000	0.006	0.102	0.000	6.000	0.000	0.000	4.475
C_S	0.112	0.526	0.000	20.000	0.000	0.000	0.047	0.319	0.000	8.000	0.000	0.000	21.523
C_{Cl}	0.088	0.452	0.000	11.000	0.000	0.000	0.024	0.222	0.000	10.000	0.000	0.000	24.965
C_{Br}	0.173	0.657	0.000	8.000	0.000	0.000	0.002	0.071	0.000	8.000	0.000	0.000	47.198
C_I	0.009	0.128	0.000	4.000	0.000	0.000	0.000	0.017	0.000	2.000	0.000	0.000	11.770

Table 1. continued

[a]Molecular descriptors in bold face plus molecular solvent accessible surface area (SASA) and number of hydrogen receptors (NHacc) not shown in this table are 21 molecular physicochemical properties used in the naïve Bayesian classification model.

average atom numbers of nitrogen, sulfur, and nonfluorine halogens of MNPs are obviously higher than those of TNPs, suggesting that MNPs may be synthesized by different and more efficient biosynthetic pathways.

Numbers of Molecular Bonds, Chains, and Rings. According to Table 1, MNPs have remarkably more rotatable bonds, stereo bonds and chains while fewer ring bonds, aromatic bonds, and aromatic rings than TNPs. As for the multiple rings, the average numbers of the four- to seven-membered rings (≥4 and ≤7), particularly the six-membered rings, of MNPs are significantly lower than those of TNPs according to their *u* values. However, the average number of the nine-membered rings of MNPs is statistically higher than that of TNPs. Taken together, MNPs are more flexible than TNPs. It is quite possible that flexible structures can adapt to the marine environment with high pressure easier. In comparison with chains, it is more difficult for rings, especially aromatic rings, to be involved in reactions. Besides, these aromatic compounds have highly adapted by terrestrial species as signaling molecules to attract useful organisms or creatures, defense against natural enemies, prevent from ultraviolet injury, and so on.[58,59] Therefore, structures of TNPs need to be more stable than those of MNPs to get acclimatized to the more volatile and totally different land habitats. Conversely, MNPs are more flexible and active to chemical reactions. The structural difference between MNPs and TNPs is partially contributed by the different evolution requirement on the land and in the marine.[60]

Overall, the molecular size, solubility, element composition and even basic structures of MNPs and TNPs have significant differences, which also highlights some important factors influencing molecular drug-likeness and biosynthetic pathways examined from the evolutionary idea. These let us urge to know more about the big differences they will make.

Drug-Likeness Analysis Based on Simple Physicochemical Properties. In practical drug design, simple drug-likeness filters, such as Lipinski's Ro5,[38] have usually been used to filter out nondrug-like molecules. According to Ro5, a molecule would be more likely to be orally absorbed if its properties satisfy the following rules: molecular weight ≤ 500; calculated octanol−water partition coefficient (ClogP) ≤ 5;[61] number of hydrogen bond donors (N_{HD}) ≤ 5; number of hydrogen bond acceptors (N_{HA}) ≤ 10. Here, Ro5 was used to estimate the drug-likeness of MNPs and TNPs, and the results are summarized in Figure 1 and Table S2. It should be noted that Ro5 can only be used as a qualitative estimator of the absorption and permeability capability of a molecule, but it does not have good capability to distinguish drug-like from nondrug-like molecules.[39−42]

As shown in Figure 1, the percentages of the compounds in the CMC, MNPs, MNPs_origin, TNPs, TNPs_origin, MDDR, CNPD, TCMCD, and TCMD databases satisfying all the rules of Ro5 are 76.04%, 55.14%, 50.96%, 65.88%, 57.13%, 62.22%, 61.92%, 53.04%, and 53.01%, respectively. The percentage of the drugs in CMC satisfying all Ro5 rules is the highest (76.04%). In contrast, the percentage of the drugs and drug candidates in MDDR satisfying all Ro5 rules is only 62.22%, which is even lower than that for TNPs (65.88%). However, the percentage of the compounds in MNPs satisfying all Ro5 rules is only 55.14%, suggesting that MNPs are possibly less drug-like than TNPs according to Ro5.

Nevertheless, according to the study reported by Zhu and co-workers, marine-originated natural products have been explored

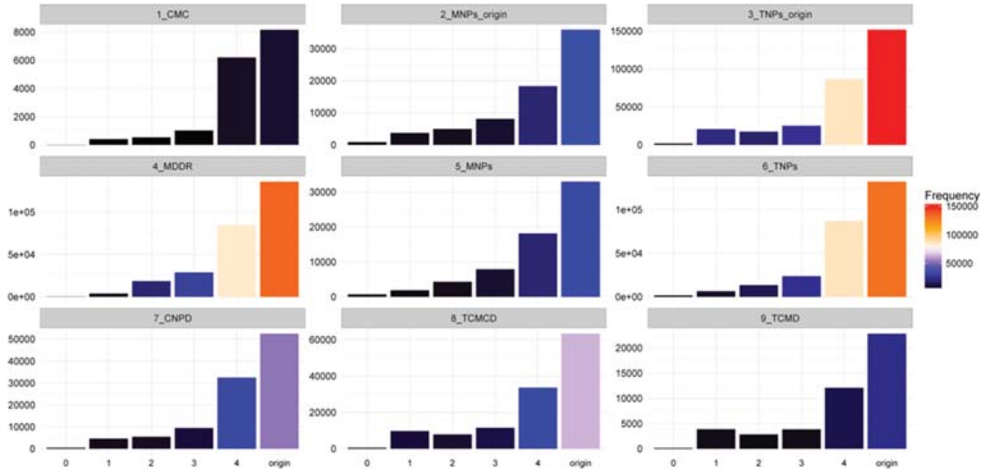


Figure 1. Drug-likeness analyses based on Ro5 ("Rule of 5") for the nine data sets. On the *X* axis, 0−4 means that the molecules only satisfy 0−4 rules of Ro5; the origin is the total number of the molecules in the data set, which has been scaled to 100% in each histogram; frequency means number of molecules.

Table 2. Overview of the Four Types of Fragments in MNPs and TNPs

		statistics of fragments[a]					
fragments	sources	total	nonduplicated	nonduplicated_P	unique	common	unique_P
chain assemblies	MNPs	367156	10188	2.77%	7919	2269	77.73%
	TNPs	1621518	20844	1.29%	18575		89.11%
ring assemblies	MNPs	42279	4868	11.51%	3456	1412	70.99%
	TNPs	190454	17109	8.98%	15697		91.75%
RECAP fragments	MNPs	1777882	142123	7.99%	131872	10251	92.79%
	TNPs	6915803	718662	10.39%	708411		98.57%
Murcko frameworks	MNPs	28833	8066	27.97%	6197	1869	76.83%
	TNPs	121975	29985	24.58%	28116		93.77%

[a] Nonduplicated_P means (the number of the nonduplicated fragments)/(total number of the same kind of fragments); unique_P means (the number of the unique fragments)/(the nonduplicated of the same kind), and the number of the unique fragments represents (the number of the nonduplicated fragments) − (the number of the same kind of common fragments in MNPs and TNPs).

actively and have shown good drug-discovery potential.[15] We then turned our focus to the molecules that only satisfy three rules of Ro5 (referred to as Ro5_3). As shown in Figure 1, the percentages of the Ro5_3 molecules in CMC, MNPs, MNPs_origin, TNPs, TNPs_origin, MDDR, CNPD, TCMCD, and TCMD are 12.42%, 24.00%, 22.63%, 17.93%, 16.62%, 21.19%, 18.07%, 18.27%, and 16.93%, respectively. The statistics of the molecules that obey each Ro5 rule is summarized in Table S2 of the Supporting Information. Up to 31.04% molecules in MNPs (only 18.95% in TNPs) do not obey the rule of logP, which is consistent with our previous analysis. For the Ro5_3 compounds, most of them violate the logP rule, from 57.02% for MDDR to 75.83% for MNPs, suggesting that these Ro5_3 compounds may become potential drug candidates if the logP issue is solved.

Differences of Fragments and Scaffolds between TNPs and MNPs. Four types of fragments or scaffolds, including chain assemblies, ring assemblies, RECAP fragments, and Murcko frameworks, were generated and compared for TNPs and MNPs. The overview of the fragments is shown in Table 2, in which the percentages of the nonduplicated fragments from MNPs are often higher than those from TNPs except the RECAP fragments. Besides, the percentages of the unique fragments in MNPs vary from at least 70.99% for the ring assemblies to 92.79% for the RECAP fragments. Therefore, there are a huge number of novel fragments and scaffolds in MNPs, especially the RECAP fragments (fragments obtained according to the retrosynthetic method) with lower nonduplicated numbers, highlighting the diverse biosynthetic pathways in MNPs.

Analyses of Chain Fragments, Ring Fragments, RECAP Fragments, and Murcko Frameworks. To step further, the enumeration method was used in the following analysis. The top ten frequently occurred unique fragments, common fragments with similar percentages of MNPs and TNPs (P_{MNPs} and P_{TNPs}), and high and low $P_{\mathrm{MNPs}}/P_{\mathrm{TNPs}}$ ratios in TNPs and MNPs are listed in Figures S1−S4 of the Supporting Information.

As shown in Figure S1 of the Supporting Information, the predominant chain assemblies in MNPs contain haloalkanes, multihydroxyls, alkenyls, esters, ethers, and sulfates. As to elements, oxygen and bromine are relatively enriched in MNPs, consistent with the fact that oxygen and halogen are abundant in the seas and oceans. The predominant chain assemblies in TNPs contain ester, carboxyl, hydroxyl, ketone, amine, alkynyl, and alkenyl functional groups. It seems that the ester groups in TNPs are more abundant than those in MNPs. The observation is not surprising because in the alkalescency water of seas or oceans, ester is often hydrolyzed into hydroxyl. As to the common chain assemblies, simple and basic chains are found in the "equal" column. In the MNPs/TNPs "descending" column, the groups with amines and oximes are abundant for

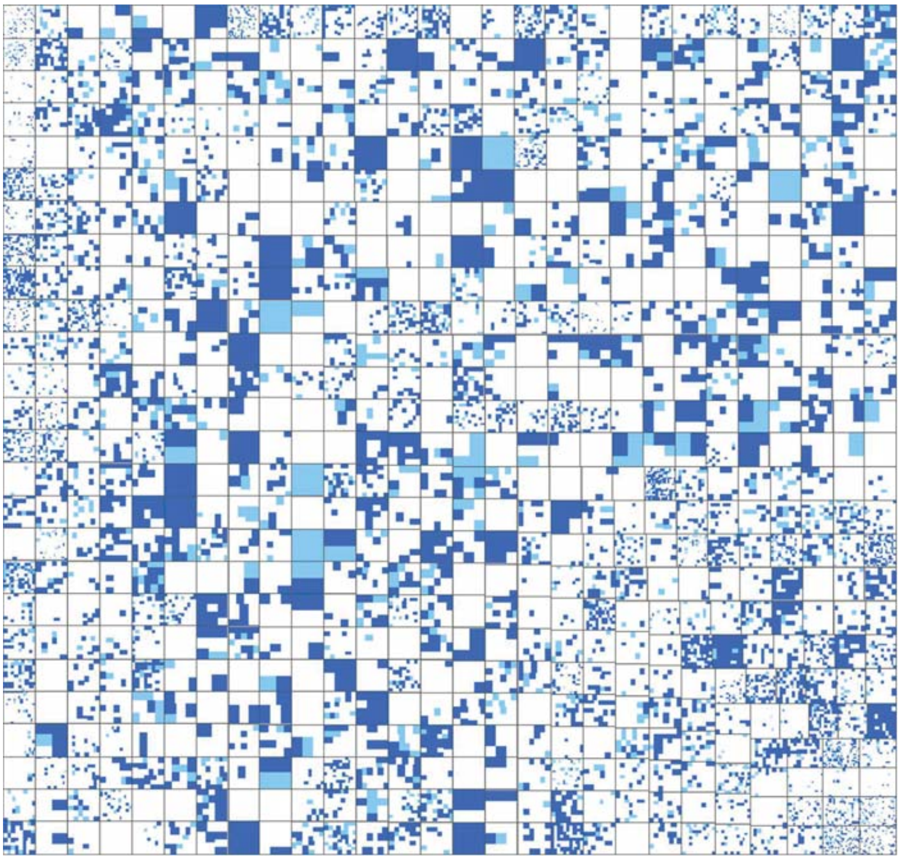


Figure 2. Comparison of the scaffold differences between TNPs (white) and MNPs (blue). The light blue color represents the common scaffolds of TNPs and MNPs; each square surrounded by the gray perimeters stands for a cluster of scaffolds.

MNPs. Both the amine and oxime functional groups have strong reducibility to induce redox reactions, which also explains the reason why oxygen is highly utilized in the anoxic water of seas and oceans. Nevertheless, as shown in the "ascending" column, the typical structural features of TNPs are still characterized by carboxyl, ester and carbonyl. The analysis of the chain assemblies shows that the fragments produced by organisms in the seas and oceans may have relatively high reactivity and thus can efficiently use oxygen in anoxic environment.

As shown in Figure S2 of the Supporting Information, there is a large proportion of the unique ring assemblies in MNPs contain ten-membered rings. Compared with the unique ring assemblies in MNPs, those in TNPs often contain five- or six-membered rings, especially the benzene and condensed rings, which contribute to the good structural stability of TNPs. In the "equal" column of common rings, all the fragments contain five- or six-membered rings, including benzene and cyclic ethers, which are building blocks of growth hormone or other substances important for life activities. Apparently, these common rings are vital for not only terrestrial but also marine species. As for the "descending" and "ascending" columns, the similar observations can be made: more complex ring assemblies are found in MNPs than in TNPs.

Similarly, significant differences between the unique RECAP fragments in MNPs and TNPs can be observed in Figure S3 of the Supporting Information. Most unique RECAP fragments from MNPs contain long chains instead of rings commonly found in TNPs. Moreover, the phosphate components are frequently observed in MNPs.

According to the Murcko frameworks shown in Figure S4 of the Supporting Information, most of the top ten occurred unique scaffolds in MNPs contain long linkers or ten-membered rings, suggesting that these unique frameworks in MNPs are flexible to facilitate the adaption of organisms with MNPs to the water habitats. The unique scaffolds in TNPs, on the other hand, contain more stable structures with lower complexity. The element compositions, chemical groups, chains, and rings to scaffolds between MNPs and TNPs have substantial differences, which guide us to investigate the differences in other aspects induced by these basic structural differences.

Insight into Unique Scaffold Types between TNPs and MNPs. In order to reveal the structural difference of the scaffolds between TNPs and MNPs more clearly in scaffold types, the scaffolds derived from TNPs and MNPs were compared by TreeMaps. In Figure 2, the pure color squares mean that all the scaffolds in the same cluster come from the same data set, TNPs (white) or MNPs (blue). In other words, the scaffolds deposited in these squares are the unique chemotypes for TNPs or MNPs. There are 18 pure blue squares and 88 pure white squares, among which 9 blue squares and 71 white squares contain more than two cluster members. The largest cluster within the pure blue squares is cluster 676 which has 11 members of Murcko scaffolds of MNPs; while the largest cluster within pure white squares is cluster 288 which has 66 members of Murcko scaffolds of TNPs (these unique

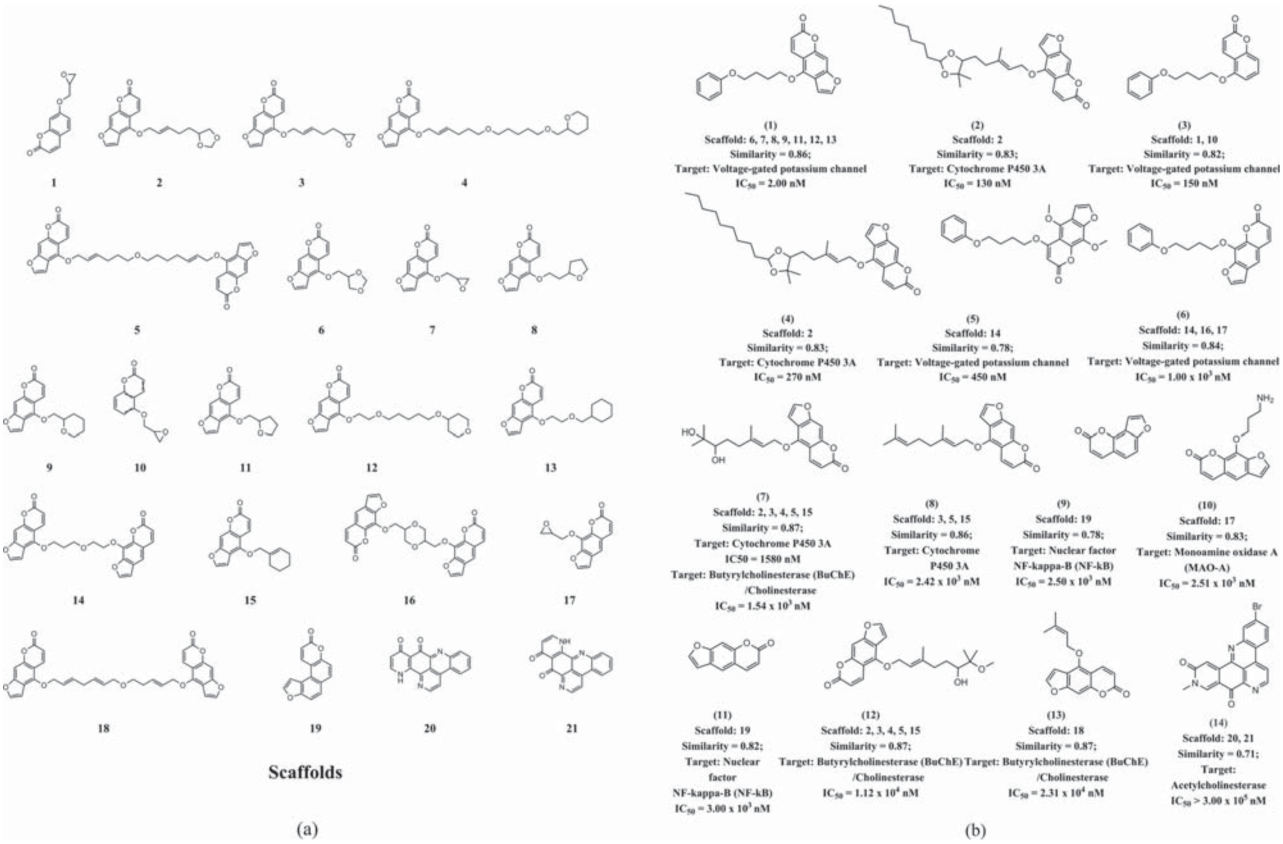

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Scaffolds
(a)
(1)
Scaffold: 6, 7, 8, 9, 11, 12, 13
Similarity = 0.86;
Target: Voltage-gated potassium channel
IC$_{50}$ = 2.00 nM
(2)
Scaffold: 2
Similarity = 0.83;
Target: Cytochrome P450 3A
IC$_{50}$ = 130 nM
(3)
Scaffold: 1, 10
Similarity = 0.82;
Target: Voltage-gated potassium channel
IC$_{50}$ = 150 nM
(4)
Scaffold: 2
Similarity = 0.83;
Target: Cytochrome P450 3A
IC$_{50}$ = 270 nM
(5)
Scaffold: 14
Similarity = 0.78;
Target: Voltage-gated potassium channel
IC$_{50}$ = 450 nM
(6)
Scaffold: 14, 16, 17
Similarity = 0.84;
Target: Voltage-gated potassium channel
IC$_{50}$ = 1.00 x 10^3 nM
(7)
Scaffold: 2, 3, 4, 5, 15
Similarity = 0.87;
Target: Cytochrome P450 3A
IC50 = 1580 nM
Target: Butyrylcholinesterase (BuChE) /Cholinesterase
IC$_{50}$ = 1.54 x 10^3 nM
(8)
Scaffold: 3, 5, 15
Similarity = 0.86;
Target: Cytochrome P450 3A
IC$_{50}$ = 2.42 x 10^3 nM
(9)
Scaffold: 19
Similarity = 0.78;
Target: Nuclear factor NF-kappa-B (NF-kB)
IC$_{50}$ = 2.50 x 10^3 nM
(10)
Scaffold: 17
Similarity = 0.83;
Target: Monoamine oxidase A (MAO-A)
IC$_{50}$ = 2.51 x 10^3 nM
(11)
Scaffold: 19
Similarity = 0.82;
Target: Nuclear factor NF-kappa-B (NF-kB)
IC$_{50}$ = 3.00 x 10^3 nM
(12)
Scaffold: 2, 3, 4, 5, 15
Similarity = 0.87;
Target: Butyrylcholinesterase (BuChE) /Cholinesterase
IC$_{50}$ = 1.12 x 10^4 nM
(13)
Scaffold: 18
Similarity = 0.87;
Target: Butyrylcholinesterase (BuChE) /Cholinesterase
IC$_{50}$ = 2.31 x 10^4 nM
(14)
Scaffold: 20, 21
Similarity = 0.71;
Target: Acetylcholinesterase
IC$_{50}$ > 3.00 x 10^5 nM
(b)

Figure 3. (a) Twenty-one unique scaffolds in cluster 288 (1−19) of TNPs and in cluster 676 (20, 21) of MNPs. (b) Representative 14 molecules by the substructure searching based on the scaffolds in cluster 288 of TNPs (1−13) and those in cluster 676 of MNPs (14).

Table 3. Similarity Comparison of the Murcko Frameworks Based on Different Similarity Cutoffs between MNPs and CMC and between TNPs and CMC

	number				percentage			
similarity	MNPs vs CMC[a]	TNPs vs CMC	CMC vs MNPs	CMC vs TNPs	MNPs vs CMC	TNPs vs CMC	CMC vs MNPs	CMC vs TNPs
=1	361	738	390	754	4.48%	2.46%	9.71%	18.77%
≥0.9	405	879	413	780	5.02%	2.93%	10.28%	19.42%
≥0.8	491	1122	459	849	6.09%	3.74%	11.43%	21.14%
≥0.7	679	1714	577	971	8.42%	5.72%	14.37%	24.18%
≥0.6	1110	3213	853	1334	13.76%	10.71%	21.24%	33.22%
≥0.5	2153	7221	1492	2079	26.69%	24.08%	37.15%	51.77%
≥0.4	3821	14430	2459	2986	47.37%	48.12%	61.23%	74.35%
≥0.3	6428	24391	3524	3825	79.69%	81.33%	87.75%	95.24%
≥0.2	7991	29871	3998	4015	99.07%	99.61%	99.55%	99.98%
≥0.1	8066	29989	4016	4016	100.00%	100.00%	100.00%	100.00%
≥0	8066	29989	4016	4016	100.00%	100.00%	100.00%	100.00%

[a] MNPs/TNPs vs CMC means MNPs/TNPs as the reference, CMC vs MNPs/TNPs means CMC as the reference.

scaffolds are listed in the Supporting Information). One may notice that there are more nitrogen atoms in the scaffolds of MNPs and more oxygen atoms in the scaffolds of TNPs. This phenomenon may be explained by the distinct habitats of marine and land as we have analyzed above. The molecules in the two largest clusters are served as the representative unique scaffolds of the two types of natural products, respectively.

In order to evaluate the potential pharmacological functions of the representative unique scaffolds of MNPs and TNPs (Figure 3a), they were used as queries to identify the similar molecules in the BindingDB database.[62] Only 21 of them have similar molecules at a high similarity cutoff around 0.85 (Figure 3a). Then a total of 35 molecules were obtained and 13 of which have $IC_{50} \leq 30\ \mu M$ (Figure 3b) in BindingDB were found similar to the representative scaffolds of TNPs. However, only one molecule in BindingDB is found using the representative scaffolds of MNPs as queries (Figure 3b, 14) even when the cutoff dropped to 0.7. Apparently, TNPs have provided sufficient and novel scaffolds for drugs or drug candidates, but the scaffolds derived from MNPs have not been widely identified as pharmacophoric fragments. Therefore, it is quite possible that marine-originated natural products have not

Table 4. Predicted Percentages of Drug-Likeness for 16 Datasets

	all			molecular weight ≤ 600		
		drug-likeness			drug-likeness	
data set	total	number	percentage	total	number	percentage
CMC	8162	6115	74.92%	7468	5679	76.04%
MNPs_origin[a]	35883	26137	72.84%	29352	22533	76.77%
MNPs	32937	25709	78.06%	28276	22479	79.50%
TNPs_origin[a]	151609	102806	67.81%	123215	93187	75.63%
TNPs	132071	101176	76.61%	118002	92672	78.53%
CNPD	52442	37267	71.06%	45791	34351	75.02%
TCMCD	63266	37557	59.36%	50589	34556	68.31%
TCMD	22797	13700	60.10%	18006	12750	70.81%
MDDR	135898	126902	93.38%	123913	116049	93.65%

[a] MNPs_origin and TNPs_origin represent original DMNP and DNP.

been well exploited by traditional drug discovery campaigns and they should have great potential to become novel drugs or drug leads in drug discovery. It is also anticipated that the unique scaffolds from MNPs and TNPs with potential pharmacological functions may provide valuable clues for ligand-based drug design.[63,64]

Drug-Likeness Analysis Based on Murcko Frameworks. After analyzing the drug-likeness of MNPs and TNPs based on physicochemical properties, we expect to gain a deeper understanding of their drug-likeness based on structures. The scaffolds in MNPs and TNPs, represented by the Murcko frameworks, were compared to those in CMC, a widely used drug database as a reference to quantitatively determine how much a compound is drug-like, using the ECFP_4 fingerprints to calculate structure similarities.[65] The numbers of the similar molecules obtained under 11 similarity cutoffs are listed in Table 3. When the cutoff was set to 0.5 (similarity ≥ 0.5), the number of the Murcko frameworks in MNPs similar to those in CMC is 2153, and the number of the Murcko frameworks in CMC similar to those in MNPs is 1492. Therefore, about 26.69% (2153/8066) of the Murcko frameworks in MNPs have similar counterparts in CMC, while 37.15% (1492/4016) of the Murcko frameworks in CMC have similar counterparts in MNPs. As to TNPs, about 24.08% of the Murcko frameworks in TNPs have similar structures in CMC, while 51.77% of the Murcko frameworks in CMC have similar structures in TNPs. Thus, it appears that MNPs are a little bit more drug-like than TNPs since the percentage of MNPs scaffolds found in CMC is higher than that of TNPs. This result is contrary to the conclusion we have made according to the property-based drug-likeness analysis. However, our explanation is that the molecular property-based drug-like rules may be biased in favor of TNPs, as 51.77% of CMC Murcko frameworks have similar counterparts in TNPs while only 37.15% in MNPs. The fact that the Murcko frameworks of MNPs are not well represented in CMC also suggests that MNPs have not been well exploited in traditional drug discovery campaigns and there is a great potential of MNPs to become novel drugs or drug leads in this golden age of drug discovery.

Differences of Drug-Likeness between TNPs and MNPs. Drug-likeness analyses based on molecular properties or structures are too arbitrary to provide reliable estimation of the drug-likeness for TNPs and MNPs.[1,5] Therefore, a naïve Bayesian classification model of drug-likeness based on both the molecular properties and structural fingerprints developed in our group was employed.[5] According to the previous study,[5] the best naïve Bayesian classification model established based on 21 simple physicochemical properties and the LCFP_6 fingerprints yields an overall leave-one-out cross-validated (LOOCV) accuracy of 91.4% for the 140 000 molecules in the training set and 90.9% for the 40 000 molecules in the test set. The drug-likeness classifier was then used to evaluate the drug-likeness of the studied data sets. The percentages of the predicted drug-likeness for all the data sets are summarized in Table 4. Among the 18 data sets, MDDR_600 (molecular weight ≤600) has the highest percentage of drug-likeness (93.65%), which is slightly higher than that of MDDR (93.38%). Besides, the drug-likeness percentage of MNPs_origin_600 (76.77%) is even higher than that of CMC_600 (76.04%), although the percentage of MNPs_origin (72.84%) is slightly lower than that of CMC (74.92%). In contrast, the drug-likeness percentages of TNPs_origin_600 and TNPs_origin are both lower than those of CMC. Interestingly, there is a big difference between the drug-likeness percentages of TNPs_origin (67.81%) and CMC (74.87%) because there are too many compounds in DNP having molecular weights higher than 600.

When looking back at our central question of this work, i.e., "what is the difference of drug-likeness between TNPs and MNPs?", we concluded that the drug-likenesses of MNPs and TNPs are comparable, as supported by the fact that the drug-likeness percentage of MNPs (78.06%) is only marginally larger than that of TNPs (76.61%). In contrast, TCMD and TCMCD are not as drug-like as we perceived before, especially when compared with MNPs_origin.

According to the evolutionary and coevolving perspectives, it is not surprising that MNPs are slightly more drug-like than TNPs. In 2009, Ma and Wang demonstrated that, through targeted screening of natural compounds from ancient species (the marine originated ones are usually more ancient than the terrestrial obviously), the rate of anticancer drug discovery can be accelerated greatly based on evolutionary theories.[66] Zhang et al. took antioxidant paradox as an example to illustrate the evolving biological roles of natural polyphenols in plants and microbes by analyzing the corresponding gene expression profiles.[67] Because polyphenols are mainly evolved for assisting plants to adapt to terrestrial life, the primary biological role of them is to defend against microorganisms and herbivores, filter ultraviolet light, and so forth instead of scavenging radicals directly.[59] It highlights the coevolutionary influence between organisms and molecules.[68] This "evolution lens" elucidates the high drug-likeness of MNPs, which are produced by more "ancient" organisms after all.

CONCLUSIONS

The differences between MNPs and TNPs were explored by the drug-likeness analyses based on 42 physicochemical properties, four kinds of fragments, and a naïve Bayesian classification model of drug-likeness. In general, MNPs have lower solubility and are often larger than TNPs. MNPs usually have longer chains and larger rings, especially the 8- to 10-membered rings, facilitating marine organisms to adapt to the water habitat. MNPs have more halogens, especially bromine, and nitrogen than TNPs, suggesting that MNPs may be synthesized by more diverse biosynthetic pathways than TNPs. Moreover, according to the predictions given by the naïve Bayesian drug-likeness classifier, most compounds in MNPs and TNPs are drug-like. Despite the percentages of drug-likeness for TNPs and MNPs are quite similar, MNPs should have greater potential than TNPs in developing new drugs because marine-originated natural products have not been well exploited in traditional drug discovery campaigns and MNPs possess some unique drug-like scaffolds according to our structural analysis. Also, evolutionary perspectives offer us convincing lines of evidence, suggesting the potential of natural products, especially the marine originated compounds, as sources of drug design.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00125.

> Table S1 information of the 60 physicochemical properties; Table S2 statistics of numbers (N) and percentages (P) of the compounds satisfying different numbers of Ro5 rules; Figure S1 unique and common chain assemblies of MNPs and TNPs; Figure S2 unique and common ring assemblies of MNPs and TNPs; Figure S3 (A) unique and (B) common RECAP fragments of MNPs and TNPs; Figure S4 unique and common Murcko frameworks of MNPs and TNPs (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: tingjunhou@zju.edu.cn (T.H.).
*E-mail: dxkong@mail.hzau.edu.cn (D.-X.K.).

ORCID

Junmei Wang: 0000-0002-9607-8229
Feng Zhu: 0000-0001-8069-0053
Huiyong Sun: 0000-0002-7107-7481
Tingjun Hou: 0000-0001-7227-2580

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by National Key R&D Program of China (2016YFB0201700; 2016YFA0501701), National Science Foundation of China (21575128; 81773632), and National Institutes of Health of USA (R01-GM079383; R21-GM097617).

REFERENCES

(1) Newman, D. J.; Cragg, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311−335.

(2) Koehn, F. E.; Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discovery* **2005**, *4*, 206−220.

(3) Kinch, M. S. 2015 in review: FDA approval of new drugs. *Drug Discovery Today* **2016**, *21*, 1046−1050.

(4) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discovery* **2015**, *14*, 111−129.

(5) Tian, S.; Wang, J.; Li, Y.; Xu, X.; Hou, T. Drug-likeness Analysis of Traditional Chinese Medicines: Prediction of Drug-likeness Using Machine Learning Approaches. *Mol. Pharmaceutics* **2012**, *9*, 2875−2886.

(6) Tian, S.; Li, Y.; Wang, J.; Xu, X.; Xu, L.; Wang, X.; Chen, L.; Hou, T. Drug-likeness analysis of traditional Chinese medicines: 2. Characterization of scaffold architectures for drug-like compounds, non-drug-like compounds, and natural compounds from traditional Chinese medicines. *J. Cheminf.* **2013**, *5*, 5.

(7) Shang, J.; Sun, H.-Y.; Liu, H.; Chen, F.; Tian, S.; Pan, P.-C.; Li, D.; Kong, D.-X.; Hou, T.-J. Comparative analyses of structural features and scaffold diversity for purchasable compound libraries. *J. Cheminf.* **2017**, *9*, 16.

(8) Szychowski, J.; Truchon, J.-F.; Bennani, Y. L. Natural Products in Medicine: Transformational Outcome of Synthetic Chemistry. *J. Med. Chem.* **2014**, *57*, 9292−9308.

(9) Crane, E. A.; Gademann, K. Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chem., Int. Ed.* **2016**, *55*, 3882−3902.

(10) Marris, E. Marine natural products - Drugs from the deep. *Nature* **2006**, *443*, 904−905.

(11) Li, J. W. H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **2009**, *325*, 161−165.

(12) Andersen, R. J.; Williams, D. E. *Chemistry in the Marine Environment*; The Royal Society of Chemistry: Cambridge, 2000.

(13) Martins, A.; Vieira, H.; Gaspar, H.; Santos, S. Marketed Marine Natural Products in the Pharmaceutical and Cosmeceutical Industries: Tips for Success. *Mar. Drugs* **2014**, *12*, 1066−1101.

(14) Spanò, V.; Attanzio, A.; Cascioferro, S.; Carbone, A.; Montalbano, A.; Barraja, P.; Tesoriere, L.; Cirrincione, G.; Diana, P.; Parrino, B. Synthesis and Antitumor Activity of New Thiazole Nortopsentin Analogs. *Mar. Drugs* **2016**, *14*, 226.

(15) Zhu, F.; Qin, C.; Tao, L.; Liu, X.; Shi, Z.; Ma, X.; Jia, J.; Tan, Y.; Cui, C.; Lin, J.; Tan, C.; Jiang, Y.; Chen, Y. Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 12943−12948.

(16) Pietra, F. Secondary metabolites from marine microorganisms: bacteria, protozoa, algae and fungi. Achievements and prospects. *Nat. Prod. Rep.* **1997**, *14*, 453−464.

(17) Faulkner, D. J. Marine natural products. *Nat. Prod. Rep.* **2002**, *19*, 1−48.

(18) Suzuki, M.; Kawamoto, T.; Vairappan, C. S.; Ishii, T.; Abe, T.; Masuda, M. Halogenated metabolites from Japanese Laurencia spp. *Phytochemistry* **2005**, *66*, 2787−2793.

(19) de Carvalho, L. R.; Fujii, M. T.; Roque, N. F.; Lago, J. H. G. Aldingenin derivatives from the red alga Laurencia aldingensis. *Phytochemistry* **2006**, *67*, 1331−1335.

(20) Laird, D. W.; van Altena, I. A. Tetraprenyltoluquinols from the brown alga Cystophora fibrosa. *Phytochemistry* **2006**, *67*, 944−955.

(21) Muigg, P.; Rosen, J.; Bohlin, L.; Backlund, A. In silico comparison of marine, terrestrial and synthetic compounds using ChemGPS-NP for navigating chemical space. *Phytochem. Rev.* **2013**, *12*, 449−457.

(22) Kong, D.-X.; Jiang, Y.-Y.; Zhang, H.-Y. Marine natural products as sources of novel scaffolds: achievement and concern. *Drug Discovery Today* **2010**, *15*, 884−886.

(23) Kong, D.-X.; Guo, M.-Y.; Xiao, Z.-H.; Chen, L.-L.; Zhang, H.-Y. Historical Variation of Structural Novelty in a Natural Product Library. *Chem. Biodiversity* **2011**, *8*, 1968−1977.

(24) Bemis, G. W.; Murcko, M. A. The properties of known drugs 0.1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(25) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP - Retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *Journal Of Chemical Information And Computer Sciences* **1998**, *38*, 511−522.

(26) Tian, S.; Wang, J.; Li, Y.; Li, D.; Xu, L.; Hou, T. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv. Drug Delivery Rev.* **2015**, *86*, 2−10.

(27) Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **2015**, *163*, 1297−1300.

(28) Hewitt, W. M.; Leung, S. S. F.; Pye, C. R.; Ponkey, A. R.; Bednarek, M.; Jacobson, M. P.; Lokey, R. S. Cell-Permeable Cyclic Peptides from Synthetic Libraries Inspired by Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 715−721.

(29) Pipeline Pilot 8.5. http://accelrys.com/ (accessed April 2016); Accelrys: San Diego, CA, USA, 2016.

(30) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.

(31) Zheng, S. X.; Luo, X. M.; Chen, G.; Zhu, W. L.; Shen, J. H.; Chen, K. X.; Jiang, H. L. A new rapid and effective chemistry space filter in recognizing a drug-like database. *J. Chem. Inf. Model.* **2005**, *45*, 856−862.

(32) In NeoTrident, Ed.; 2009.

(33) Hou, T. J.; Qiao, X. B.; Xu, X. J. Research and development of 3D molecular structure database of traditional Chinese drugs. *Acta Chim. Sin.* **2001**, *59*, 1788−1792.

(34) Qiao, X. B.; Hou, T. J.; Zhang, W.; Guo, S. L.; Xu, S. J. A 3D structure database of components from Chinese traditional medicinal herbs. *Journal Of Chemical Information And Computer Sciences* **2002**, *42*, 481−489.

(35) Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17272−17277.

(36) Grabowski, K.; Baringhaus, K.-H.; Schneider, G. Scaffold diversity of natural products: inspiration for combinatorial library design. *Nat. Prod. Rep.* **2008**, *25*, 892−904.

(37) *Molecular Operating Environment (MOE)*, 2014; Chemical Computing Group: Montreal, Quebec, Canada.

(38) 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. *Adv. Drug Delivery Rev.* **2012**, *64*, 4−17.

(39) Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME evaluation in drug discovery. 6. Can oral bioavailability in humans be effectively predicted by simple molecular property-based rules? *J. Chem. Inf. Model.* **2007**, *47*, 460−463.

(40) Hou, T.; Li, Y.; Zhang, W.; Wang, J. Recent Developments of In Silico Predictions of Intestinal Absorption and Oral Bioavailability. *Comb. Chem. High Throughput Screening* **2009**, *12*, 497−506.

(41) Tian, S.; Li, Y.; Wang, J.; Zhang, J.; Hou, T. ADME Evaluation in Drug Discovery. 9. Prediction of Oral Bioavailability in Humans Based on Molecular Properties and Structural Fingerprints. *Mol. Pharmaceutics* **2011**, *8*, 841−851.

(42) Shen, M.; Tian, S.; Li, Y.; Li, Q.; Xu, X.; Wang, J.; Hou, T. Drug-likeness analysis of traditional Chinese medicines: 1. property distributions of drug-like compounds, non-drug-like compounds and natural compounds from traditional Chinese medicines. *J. Cheminf.* **2012**, *4*, 31.

(43) *TreeMap*, v. 3.8.3; Macrofocus GmbH: Swiss.

(44) Shneiderman, B. TREE VISUALIZATION WITH TREE-MAPS - 2-D SPACE-FILLING APPROACH. *Acm Transactions on Graphics* **1992**, *11*, 92−99.

(45) Thanh Le, G.; Abbenante, G.; Becker, B.; Grathwohl, M.; Halliday, J.; Tometzki, G.; Zuegg, J.; Meutermans, W. Molecular diversity through sugar scaffolds. *Drug Discovery Today* **2003**, *8*, 701−709.

(46) Gorse, D.; Lahana, R. Functional diversity of compound libraries. *Curr. Opin. Chem. Biol.* **2000**, *4*, 287−294.

(47) Khanna, V.; Ranganathan, S. Structural diversity of biologically interesting datasets: a scaffold analysis approach. *J. Cheminf.* **2011**, *3*, 30.

(48) Feher, M.; Schmidt, J. M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *Journal Of Chemical Information And Computer Sciences* **2003**, *43*, 218−227.

(49) Raymond, J.; Segre, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **2006**, *311*, 1764−1767.

(50) Tute, M. S. Lipophilicity: A History. In *Lipophilicity in Drug Action and Toxicology*; VCH: Weinheim, 1996; pp 7−26.

(51) Hou, T. J.; Zhang, W.; Xia, K.; Qiao, X. B.; Xu, X. J. ADME evaluation in drug discovery. 5. Correlation of Caco-2 permeation with simple molecular properties. *Journal Of Chemical Information And Computer Sciences* **2004**, *44*, 1585−1600.

(52) Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J. Chem. Inf. Model.* **2007**, *47*, 208−218.

(53) Hou, T.; Wang, J.; Li, Y. ADME evaluation in drug discovery. 8. The prediction of human intestinal absorption by a support vector machine. *J. Chem. Inf. Model.* **2007**, *47*, 2408−2415.

(54) Hou, T.; Wang, J.; Zhang, W.; Wang, W.; Xu, X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr. Med. Chem.* **2006**, *13*, 2653−2667.

(55) Lee, M. L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **2001**, *3*, 284−289.

(56) Gribble, G. W. *The Handbook of Environmental Chemistry*; Springer-Verlag: Berlin, 2003; Vol. *3*.

(57) Henkel, T.; Brunne, R. M.; Muller, H.; Reichel, F. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem., Int. Ed.* **1999**, *38*, 643−647.

(58) Li, D.-Z.; Yu, G.-Q.; Yi, S.-C.; Zhang, Y.; Kong, D.-X.; Wang, M.-Q. Structure-Based Analysis of the Ligand-Binding Mechanism for DhelOBP21, a C-minus Odorant Binding Protein, from Dastarcus helophoroides (Fairmaire; Coleoptera: Bothrideridae). *Int. J. Biol. Sci.* **2015**, *11*, 1281−1295.

(59) Rozema, J.; Bjorn, L. O.; Bornman, J. F.; Gaberscik, A.; Hader, D. P.; Trost, T.; Germ, M.; Klisch, M.; Groniger, A.; Sinha, R. P.; Lebert, M.; He, Y. Y.; Buffoni-Hall, R.; de Bakker, N. V. J.; van de Staaij, J.; Meijkamp, B. B. The role of UV-B radiation in aquatic and terrestrial ecosystems - an experimental and functional analysis of the evolution of UV-absorbing compounds. *J. Photochem. Photobiol., B* **2002**, *66*, 2−12.

(60) Darwin, C. *The Origin of Species*; Random House: New York, 1999.

(61) Ru, J.; Li, P.; Wang, J.; Zhou, W.; Li, B.; Huang, C.; Li, P.; Guo, Z.; Tao, W.; Yang, Y.; Xu, X.; Li, Y.; Wang, Y.; Yang, L. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminf.* **2014**, *6*, 13.

(62) Gilson, M. K.; Liu, T. Q.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045−D1053.

(63) Jasial, S.; Hu, Y.; Bajorath, J. Assessing the Growth of Bioactive Compounds and Scaffolds over Time: Implications for Lead Discovery and Scaffold Hopping. *J. Chem. Inf. Model.* **2016**, *56*, 300−307.

(64) Hu, Y.; Stumpfe, D.; Bajorath, J. Recent Advances in Scaffold Hopping. *J. Med. Chem.* **2017**, *60*, 1238−1246.

(65) Vogt, M.; Bajorath, J. Modeling tanimoto similarity value distributions and predicting search results. *Mol. Inf.* **2017**, *36*, 1600131.

(66) Ma, X.; Wang, Z. Anticancer drug discovery in the future: an evolutionary perspective. *Drug Discovery Today* **2009**, *14*, 1136−1142.

(67) Zhang, H.-Y.; Chen, L.-L.; Li, X.-J.; Zhang, J. Evolutionary inspirations for drug discovery. *Trends Pharmacol. Sci.* **2010**, *31*, 443−448.

(68) Morcos, F. Molecular Coevolution of Fli Proteins Provides a Guide to Accurate Models of Flagellar Protein Complexes and Dynamics. *Biophys. J.* **2017**, *112*, 469a−470a.