

# KampoDB をより使いやすくするための大幅更新

奥 牧人<sup>1,a)</sup>

**概要:** 本稿では、富山大学和漢医薬学総合研究所が管理している KampoDB というデータベースをより使いやすくするために実施した、直近の大幅更新の内容について紹介する。このデータベースは、漢方処方、生薬、天然化合物、蛋白質などの関係をまとめたものである。主な変更点は、ページ毎に固有の URL を割り当てたこと、低分子化合物と蛋白質の組のページを除きどのページも一つの検索ボックスで検索出来るようにしたこと、要素数の多い一覧を分割表示にしたこと、分子描画に用いるライブラリを JSmol から 3Dmol.js に変えたこと、元々 3 つに分かれていたデータベースの内容を 1 つにまとめたこと、低分子化合物と蛋白質の結合強度を 6 段階表示にしたことである。1 日当たりの平均ページ閲覧数は更新前の 9.3 回から 17.3 回に増え、改善の効果が示唆されている。

**キーワード:** KampoDB, データベース, 漢方処方, 生薬, 天然化合物, 蛋白質

## A major update of KampoDB for better usability

MAKITO OKU<sup>1,a)</sup>

**Abstract:** In this paper, I introduce the contents of the latest major update of KampoDB, a database maintained by Institute of Natural Medicine, University of Toyama, which has been done for better usability. The database summarizes the relationships between kampo formulas, crude drugs, natural compounds, proteins, and so on. The major changes are: assigned a unique URL to each page, enabled to search any page in one search box except for ligand-protein pages, introduced pagination, changed the molecular visualization library from JSmol to 3Dmol.js, integrated the database contents into one that was originally divided into three, and displayed the binding strength between a small molecule and a protein in six levels. The average number of daily page views has increased from 9.3 before the update to 17.3, indicating the effect of the improvement.

**Keywords:** KampoDB, database, kampo formula, crude drug, natural compound, protein

### 1. はじめに

KampoDB [1–4] は、漢方処方、生薬、天然化合物、蛋白質などの関係をまとめたデータベースである。漢方処方とは、漢方医学において用いられる処方薬のことであり、そのほとんどは 2 つ以上の生薬を組み合わせたものである。生薬とは薬用植物の根などを指し、その中には複数の天然化合物が含まれている。生薬に含まれる天然化合物または

その代謝物がヒトの体内で蛋白質と結合することで、様々な薬効が生じると考えられている。なお、蛋白質も化合物であるが、本稿では化合物という単語は低分子化合物のみを指すこととする。

漢方処方や生薬に関する情報を収録している有用なデータベースは複数存在する。例えば、KEGG (Kyoto Encyclopedia of Genes and Genomes) [5], Metabolomics.jp [6, 7], KNApSACk [8], 伝統医薬データベース [9, 10] などである。KampoDB 独自の情報は大きく 2 種類ある。一つ目は、様々な天然化合物がそれぞれどの蛋白質に結合するかを結合シミュレーション (構造ベース) と機械学習 (リガンドベース) により予測した結果である。もう一つは、個々

<sup>1</sup> 富山大学 和漢医薬学総合研究所  
Institute of Natural Medicine, University of Toyama,  
Toyama 930-0194, Japan

<sup>a)</sup> oku@inm.u-toyama.ac.jp



図 1 新しいページ構成と以前の内容との対応関係

Fig. 1 Correspondence between new page structure and previous contents

の漢方処方や生薬が作用するパスウェイや生物学的プロセスをエンリッチメント解析により予測した結果である。過去の調査報告 [11] で後者は前者に基づくと書いたが、その後の調査で基づいていないことが分かったので、この場を借りて訂正する。

筆者は KampoDB の構築に一切関わっていないが、同データベースが 2019 年 5 月に富山大学和漢医薬学総合研究所 (以下、和漢研) の管理下となったため、それ以降管理を担当している。過去の調査報告 [11] で様々な問題点を指摘したが、2020 年 1 月に和漢研内で行ったアンケート調査でも KampoDB が使いづらいとのご意見を頂いた。

そこで、KampoDB をより使いやすくするために 2021 年 7 月 15 日に大幅にリニューアルした。本稿ではその内容について紹介する。

## 2. 表示と構成に関する変更点

表示と構成に関する変更点について述べる。まず、情報毎に異なる URL を割り当てた。これまではウェブサイト内でボタンを押したり内部リンクをクリックしたりした際、URL が変わらずに画面表示のみが変わる方式が多用されていた。そのため、Google などの検索サイトから個々のページを見つけられず、また、和漢研が管理している伝統医薬データベース [9] と KampoDB をページ毎に相互リンクさせることも出来なかった。この問題を改善した。

次に、化合物と蛋白質の組のページを除き、どのページも一つの検索ボックスで検索出来るようにした。これまで KampoDB は第 1 部から第 3 部までの 3 つに分かれていたが、第 1 部と第 2 部はそれぞれ検索ボックスが 4 つずつあり、第 3 部には検索ボックスが無かった。この問題を改善した。また、元の検索ボックスは大文字と小文字を

区別する方式 (case-sensitive) だったが、区別しない方式 (case-insensitive) に変えた。

また、要素数の多い一覧を 10 件ずつ分割表示にした。英語では pagination と呼ぶ。ウェブスクレイピングをする人にとっては不便になるが、普通の人は分割した方が見やすいだろう。念のため、全件表示ボタンも付けておいた。

続いて、これまで 3 部に分かれていたデータベースの内容を 1 つにまとめた。図 1 に概要を示す。第 1 部の内容は、漢方処方、生薬、化合物、蛋白質の対応関係だった。第 2 部の内容は、漢方処方、生薬、化合物に関するエンリッチメント解析の結果だった\*1。第 3 部の内容は、化合物と蛋白質の相互作用予測結果だった。リガンドベース (Ligand-based) と構造ベース (Structure-based) の結果があった。

更新後は、図 1 に示すように、個々の漢方処方、生薬、化合物、蛋白質について縦長のページを作り、そこに第 1 部、第 2 部、第 3 部の情報を配置した。蛋白質ページにはエンリッチメント解析結果ではなく直接関連する GO (Gene Ontology) 注釈などを載せた。第 3 部の結合シミュレーション結果の 3D 描画は、化合物と蛋白質の組ごとのページを作り、そこに置いた。また、各化合物の 3D 描画は廃止した。理由は、PubChem [12] で見る事が出来るためである。

結合シミュレーション結果の 3D 描画ページでは、化合物と蛋白質の結合強度を 6 段階表示にした (図 2 参照)。理由は、AutoDock Vina [13] の出力する結合親和性の値が、このソフトウェアを多用する人以外には分かりにくいからである。AutoDock Vina では、無作為に化合物と蛋白質を

\*1 エンリッチメント解析とは 2 つの集合の重複度を調べる解析であり、ここでの集合の要素は蛋白質/遺伝子なので、個々の蛋白質/遺伝子に対しては実施できない。

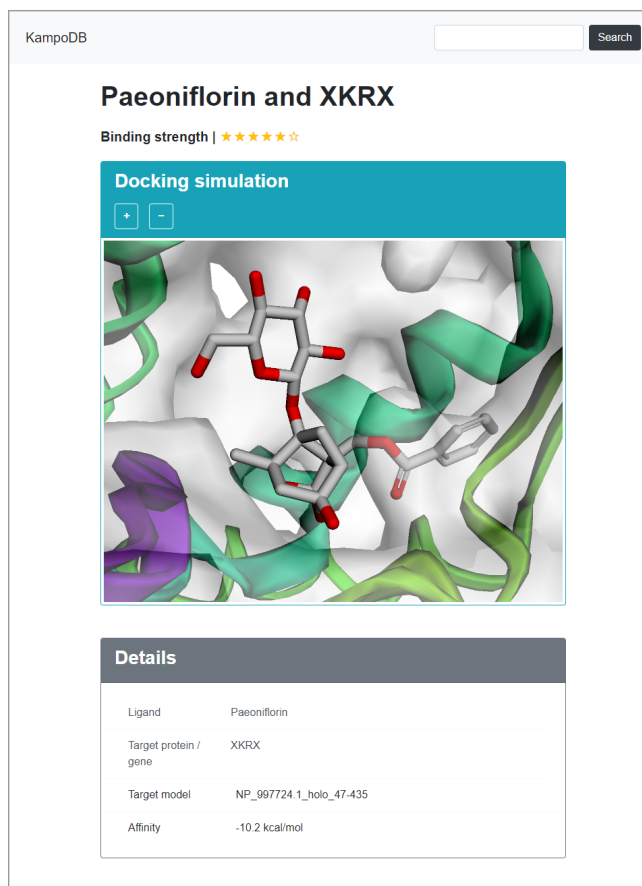


図 2 結合シミュレーション結果の 3D 描画ページの例

Fig. 2 An example 3D visualization page of a docking simulation result

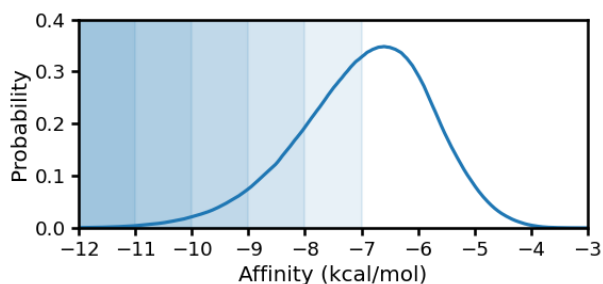


図 3 全ての結合シミュレーション結果における最適結合モードの親和性の分布

Fig. 3 Distribution of the affinity of the best binding modes of all the docking simulation results

一つずつ取ってきても、平均的に  $-7$  kcal/mol 程度の値が出てしまう (図 3 参照)。これを踏まえ、表 1 のように結合強度を 6 段階で表すことにした。さらに、既定で段階 3 以上のみを表示するようにした。

### 3. 実装に関する変更点

実装に関する変更点について述べる。バックエンドの実装は JavaServer Pages から Python の Flask に変更した。これに伴い、Apache との連携方法を Apache Tomcat から

表 1 結合親和性の 6 段階表示

Table 1 Six-level representation of binding affinity

Level	Affinity's range (kcal/mol)	Ratio
1	$-7 < x$	53.3 %
2	$-8 < x \leq -7$	27.2 %
3	$-9 < x \leq -8$	13.5 %
4	$-10 < x \leq -9$	4.6 %
5	$-11 < x \leq -10$	1.2 %
6	$x \leq -11$	0.2 %

mod\_wsgi に変更した。関係データベース管理システムは PostgreSQL から SQLite に変更した。Flask で簡単な Web API も作った。フロントエンドの実装は JavaScript のままだが、jQuery ライブラリから Vue.js フレームワークに変えた\*2。CSS フレームワークは Bootstrap の日本語対応テーマ Honoka を引き続き使用した。

分子描画ライブラリは JSmol [14] から 3Dmol.js [15] に変更した。これにより、分子の回転や拡大・縮小がより滑らかになった。3Dmol.js の他にも NGL Viewer [16] と Molmil [17] を試してみたが、前者は起動がやや遅く、ページ毎に URL を変える方針とは相性が悪く、後者は情報が不足していて描画様式の変更法がよく分からなかったため、断念した。

マシンは物理サーバから仮想マシンに変更した。これまででは和漢研内の物理サーバ上に KampoDB の実体があり、学外からアクセスする際は富山大学総合情報基盤センターが提供している仮想マシンを経由するように設定されていた。その理由の一つとして、元々 KampoDB のデータ容量は約 300GB あり、仮想マシンの容量上限 100GB を越えていたことが挙げられる。この問題は、筆者が 2019 年に不要データを削除し主なデータを zip 圧縮した [11] ことで既に解決していた。そこで今回、KampoDB の実体を、これまで中継に用いていた仮想マシン上に移動した。この仮想マシンが動いている物理サーバは定期的新しいものと交換されるため、故障の危険は大幅に減ったと考えられる。

通信プロトコルは 2020 年 2 月に http から https に切り替わっており、今も https を使っている。

### 4. データに関する変更点

データに関する変更点について述べる。全体的に、外部の情報源から取得しているデータの多くを最新版にした。一方、独自データは少し整理しただけで、追加はしていない。識別子も一部変更した。SQL (Structured Query Language) で扱うデータの総量は約 80 MB から約 173 MB に増えた。その他の PDB 形式のデータは約 35 GB から約 14 GB に減少した。

漢方処方の識別子は、独自の通し番号から英語略号 [18]

\*2 正確には 3Dmol.js の中で jQuery が今も使われている。筆者の自作部分では jQuery を使っていない。

に変えた。これは医療用の 148 処方と一般用の 294 処方の和集合である 298 処方と一対一に対応する<sup>\*3</sup>。それらを全て登録したため、処方数は 42 から 298 に増えた。漢方処方のローマ字表記は、漢方処方名ローマ字表記法 [19] に従って修正した。

漢方処方と生薬の対応関係のデータは、主に一般用漢方製剤製造販売承認基準から取得した。医療用のみの 4 処方に関するデータは、KEGG [5] から取得した。

生薬の識別子は、独自の通し番号のままとした。生薬名などの基本情報と、生薬と基原植物の関係データは、最新の第十八改正日本薬局方と日本薬局方外生薬規格 2018 から取得した。どちらにも載っていないが上述の漢方処方に含まれる 20 種類の生薬データは、伝統医薬データベース [9] から取得した。生薬の登録数は 54 から 180 に増えた。

植物と化合物の対応関係のデータは、KNApSACk [8] から最新版を取得した。これまで PCIDB [20] の情報を使っていたようだが、PCIDB の該当部分のデータは KNApSACk から取得したもののようであり、更新もされていない様子なので、情報源を本家に切り替えた。

化合物の識別子は、KNApSACk の C.ID から PubChem [12] の CID に変えた。化合物名などの基本情報は、PubChem から最新版を取得した。

化合物と蛋白質の対応関係のデータは、CTD (Comparative Toxicogenomics Database) [21] から最新版を取得した。これまで ChEMBL [22] などの複数のデータベースを使っていたが、管理しづらいので単一の情報源に切り替えた。CTD の化合物の識別子<sup>\*4</sup>は、一旦 NCBI で PubChem SID に変換してから Pubchem で CID に変換した。

蛋白質/遺伝子の識別子は、Entrez gene ID のままとした。これは遺伝子の識別子であって蛋白質の識別子ではないが、他の複数のデータベースでも識別子として使われていて統合が簡単なので、変更はしなかった。遺伝子記号などの基本情報は、NCBI から最新版を所得した。

GO のデータは、最新版を Gene Ontology Consortium [23] から取得した。GO には BP (Biological Process), MF (Molecular Function), CC (Cellular Component) の 3 種類がある。一般的には BP のみまたは 3 つ全てを使うことが多いと思うが、KampoDB では BP と MF のみを使っている。今回はそのままとした。GO 注釈は UniProt [24] の識別子で書かれていたので、UniProt から最新の対応表を取得して Entrez gene ID に変換した。

KEGG [5] のパスウェイと疾患のデータは、最新版を KEGG から取得した。KEGG Brite の第 3 層を用いたエンリッチメント解析の結果は削除した。理由は、Brite 第 3 層の各要素が KEGG 側で一意の URL を持っておらず、KampoDB からリンクを張れなかったことと、エンリッチ

メント解析で KEGG Brite を使うことはあまり一般的ではないと考えたためである。

エンリッチメント解析は、これまでは事前に計算した結果を表示していたが、その場で計算する方式に変えた。重複数が 5 以上の場合を対象にフィッシャーの正確検定を行い、偽発見率 (False Discovery Rate, FDR) が 5 %未満の注釈を表示するようにした。従って、現在サーバ内にエンリッチメント解析の結果データは保存されていない。

蛋白質構造データは、SAHG (Structure Atlas of Human Genome) [25] のドメインモデリングデータのままとした。SAHG は既に閉鎖しているので<sup>\*5</sup>、PDB (Protein Data Bank) [26] や AlphaFold DB (AlphaFold Protein Structure Database) [27] など別の情報源に切り替えた方が良いが、今回は実施出来なかった。これまで KampoDB 内に 42330 個のファイルがあったが、NP から始まる RefSeq ID から Entrez gene ID を NCBI で逆引き出来なかったものを削除したため、40685 個に減った。

結合シミュレーション結果の独自データは、少し整理した。具体的には、化合物と蛋白質構造データの組ごとに最大 9 個の結合モードの予測結果があったが、親和性が最も強い一つを残し、他を削除した。

## 5. 未解決の課題

未解決の課題について述べる。表示と構成に関して、以前の調査 [11] で見つかった課題は全て解決した。追加課題として、日本語版の作成が挙げられる。

実装に関して、以前の調査 [11] で見つかった課題のうち、docker 化以外は全て解決した。追加課題は今のところ無いが、もしも今後アクセス数が増加し動作が重くなった場合は、学内の仮想マシンからクラウドへ移行する必要があるだろう。

独自データに関して、以前の調査 [11] で見つかった課題は全然解決していない。以下に引用する。

データの正確性については、構造既知の蛋白質データの使用、可動部の設定、元から結合していたペプチドの除去、探索範囲の限定、AutoDock Vina より新しいソフトウェアの使用などが挙げられる。データの網羅性については、化合物の追加、結果の無い箇所の再計算などが挙げられる。

これらのうち、蛋白質構造データについては AlphaFold DB [27] の登場によって実験データと予測データの差が大幅に縮まったため、もし再計算する場合は全て最新の予測データを使えば良いかもしれない。また、AutoDock Vina [13] は現在でも十分実用的と考えられるため変えなくて良いかもしれない。

<sup>\*3</sup> 小青竜湯加石膏の SSTKS は SSTS の誤植と考えられる。

<sup>\*4</sup> 見た目は KNApSACk の C.ID と似ているが全くの別物である。

<sup>\*5</sup> 生命科学系データベースアーカイブから取得可能である。

表 2 大幅更新前後の KampoDB の利用統計 (平均 ± 標準誤差)

Table 2 KampoDB usage statistics of KampoDB before and after the major update (mean ± standard error)

name	before	after
page views per day	9.30 ± 0.52	17.30 ± 1.65
number of users per day	2.31 ± 0.06	4.83 ± 0.25
number of sessions per day	2.56 ± 0.07	5.23 ± 0.28
average session duration (s)	108.2 ± 9.0	90.0 ± 15.0
bounce rate (%)	55.5 ± 1.4	69.4 ± 2.1

## 6. 大幅更新の効果

大幅更新の効果について簡単に述べる。より詳細な分析結果は [28] で報告する予定である。Google Analytics により得られた利用統計を KampoDB の大幅更新前 (正確な記録がある 2019 年 5 月 9 日から 2021 年 7 月 14 日まで) と大幅更新後 (2021 年 7 月 16 日から 2021 年 11 月 30 日まで) で比較したところ, 1 日当たりのページ閲覧数, ユーザ数, セッション数が約 2 倍に増えた (表 2 参照)。それぞれの時系列からも大幅更新以降で増えたことが見てとれる (図 4 参照)。これらの結果より, 大幅更新によって確かに利便性が向上したことが示唆される。

一方, 平均セッション時間は低下し直帰率は増加した (表 2 参照)。図 4 の直帰率の時系列を見ると, 2020 年 2 月に急激に増加した後はほぼ横ばいである\*6。従って, 直帰率が大幅更新により増加したとは言えないと考えられる。平均セッション時間は変動が大きいため, 現時点で傾向と原因を考察することは難しい。

## 7. まとめ

本稿では KampoDB の大幅更新の内容を紹介した。表示と構成, 実装, データに関する変更についてそれぞれ説明した。一つ一つは些細なことかもしれないが, それらの積み重ねで, ユーザが以前より使いやすくなったと感じてくれることを期待している。

### 参考文献

- [1] Sawada, R., Iwata, M., Umezaki, M., Usui, Y., Kobayashi, T., Kubono, T., Hayashi, S., Kadowaki, M. and Yamanishi, Y.: KampoDB, database of predicted targets and functional annotations of natural medicines, *Sci. Rep.*, Vol. 8, p. 11216 (online), DOI: <https://doi.org/10.1038/s41598-018-29516-1> (2018).
- [2] 梅寄雅人, 門脇 真, 立川仁典, 山西芳裕: 和漢薬のターゲットタンパク質のインシリコ探索, 和漢薬の科学基盤形成拠点 平成 26 年度共同研究報告書, pp. 11-15 (オンライン), 入手先 ([https://www.inm.u-toyama.ac.jp/collabo/doc/2014fy\\_report\\_03.pdf](https://www.inm.u-toyama.ac.jp/collabo/doc/2014fy_report_03.pdf)) (2015).
- [3] 梅寄雅人, 門脇 真, 立川仁典, 山西芳裕, 五斗 進: 和漢薬のターゲットタンパク質のインシリコ探索, 和漢

\*6 この増加が https 化の影響かどうかは分からない。

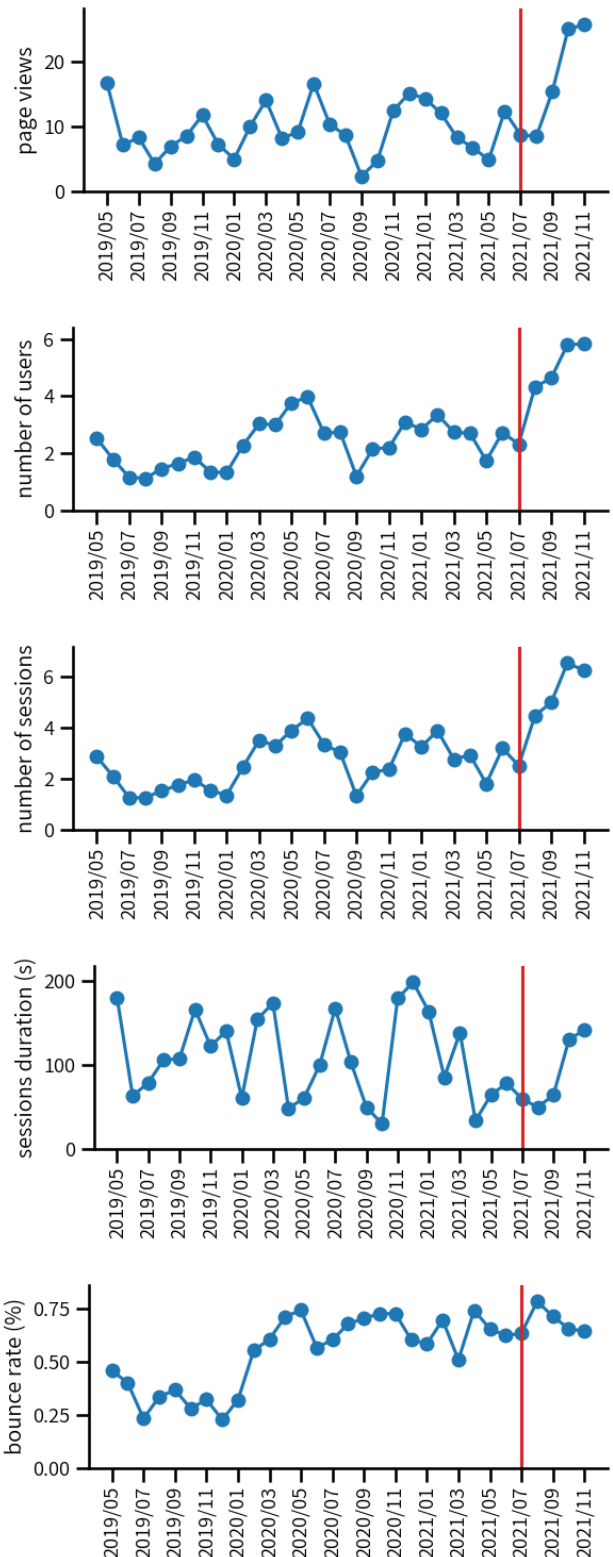


図 4 KampoDB の利用統計の時系列。赤線は大幅更新の時点を表す。

Fig. 4 Time series of KampoDB usage statistics. The red lines show the time point of the major update.

薬の科学基盤形成拠点 平成 27 年度共同研究報告書, pp. 6-10 (オンライン), 入手先 ([https://www.inm.u-toyama.ac.jp/collabo/doc/2015fy\\_report\\_03.pdf](https://www.inm.u-toyama.ac.jp/collabo/doc/2015fy_report_03.pdf)) (2016).

- [4] 門脇 真: 漢方薬のドラッグ・リポジショニングのための生命科学ビッグデータを活用したアルゴリズム/データベース KampoDB の開発, *Prog. Med.*, Vol. 39, No. 11, pp. 1081–1084 (2019).
- [5] Kanehisa, M. and Goto, S.: KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res.*, Vol. 28, No. 1, pp. 27–30 (online), DOI: <https://doi.org/10.1093/nar/28.1.27> (2000).
- [6] Arita, M. and Suwa, K.: Search extension transforms Wiki into a relational system: a case for flavonoid metabolite database, *BioData mining*, Vol. 1, No. 1, p. 7 (online), DOI: <https://doi.org/10.1186/1756-0381-1-7> (2008).
- [7] Arita, M., Yoshimoto, M., Suwa, K., Hirai, A., Kanaya, S., Shibahara, N. and Tanaka, K.: Database for crude drugs and Kampo medicine, *Genome Inform.*, Vol. 25, No. 1, pp. 1–11 (online), DOI: <https://doi.org/10.11234/gi.25.1> (2011).
- [8] Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L. K., Saito, K. and Kanaya, S.: KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research, *Plant Cell Physiol.*, Vol. 53, No. 2, p. e1 (online), DOI: <https://doi.org/10.1093/pcp/pcr165> (2012).
- [9] : 伝統医薬データベース, <https://dentomed.toyama-wakan.net>.
- [10] 奥 牧人: 伝統医薬データベースに登録された生物検定データの統合クラスタ解析, 日本薬学会第 140 年会 (2020).
- [11] 奥 牧人: KampoDB の結合シミュレーション結果の概説, 情処研報, Vol. 2019-BIO-59, No. 1, pp. 1–6 (オンライン), 入手先 (<http://id.nii.ac.jp/1001/00198992/>) (2019).
- [12] Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J. and Bryant, S. H.: PubChem substance and compound databases, *Nucleic Acids Res.*, Vol. 44, pp. D1202–D1213 (online), DOI: <https://doi.org/10.1093/nar/gkv951> (2016).
- [13] Trott, O. and Olson, A. J.: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, Vol. 31, No. 2, pp. 455–461 (online), DOI: <https://doi.org/10.1002/jcc.21334> (2010).
- [14] Hanson, R. M., Prilusky, J., Renjian, Z., Nakane, T. and Sussman, J. L.: JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia, *Isr. J. Chem.*, Vol. 53, No. 3–4, pp. 207–216 (online), DOI: <https://doi.org/10.1002/ijch.201300024> (2013).
- [15] Rego, N. and Koes, D.: 3Dmol.js: molecular visualization with WebGL, *Bioinformatics*, Vol. 31, No. 8, pp. 1322–1324 (online), DOI: <https://doi.org/10.1093/bioinformatics/btu829> (2015).
- [16] Rose, A. S. and Hildebrand, P. W.: NGL Viewer: a web application for molecular visualization, *Nucleic Acids Res.*, Vol. 43, pp. W576–W579 (online), DOI: <https://doi.org/10.1093/nar/gkv402> (2015).
- [17] Bekker, G.-J., Nakamura, H. and Kinjo, A. R.: Molmil: a molecular viewer for the PDB and beyond, *J. Cheminform.*, Vol. 8, No. 1, p. 42 (online), DOI: <https://doi.org/10.1186/s13321-016-0155-1> (2016).
- [18] Hagihara, K., Yakubo, S. and Namiki, T.: Abbreviation of kampo formulations and basic terminology in kampo medicine, *Trad. Kampo Med.*, Vol. 4, No. 2, pp. 65–88 (online), DOI: <https://doi.org/10.1002/tkm2.1078> (2017).
- [19] 津谷喜一郎, 佐竹元吉, 鳥居塚和生, 引網宏彰, 山田和男: 漢方処方名ローマ字表記法, 日本東洋医学雑誌, Vol. 56, No. 4, pp. 611–622 (オンライン), DOI: <https://doi.org/10.3937/kampomed.56.609> (2005).
- [20] 西村陽介: 植物の効能メカニズムを解明するためのデータベース PCIDB の開発, 京都大学化学研究所スーパーコンピュータシステム研究成果報告書, Vol. 2014, pp. 53–54 (オンライン), 入手先 (<http://hdl.handle.net/2433/197634>) (2015).
- [21] Mattingly, C. J., Colby, G. T., Forrest, J. N. and Boyer, J. L.: The Comparative Toxicogenomics Database (CTD), *Environ. Health Perspect.*, Vol. 111, No. 6, pp. 793–795 (online), DOI: <https://doi.org/10.1289/ehp.6028> (2003).
- [22] Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. and Overington, J. P.: ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, Vol. 40, pp. D1100–D1107 (online), DOI: <https://doi.org/10.1093/nar/gkr777> (2012).
- [23] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G.: Gene Ontology: tool for the unification of biology, *Nat. Genet.*, Vol. 25, No. 1, pp. 25–29 (online), DOI: <https://doi.org/10.1038/75556> (2000).
- [24] Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. and Yeh, L.-S. L.: UniProt: the Universal Protein knowledgebase, *Nucleic Acids Res.*, Vol. 32, pp. D115–D119 (online), DOI: <https://doi.org/10.1093/nar/gkh131> (2004).
- [25] Motono, C., Nakata, J., Koike, R., Shimizu, K., Shirota, M., Amemiya, T., Tomii, K., Nagano, N., Sakaya, N., Misoo, K., Sato, M., Kidera, A., Hiroaki, H., Shirai, T., Kinoshita, K., Noguchi, T. and Ota, M.: SAHG, a comprehensive database of predicted structures of all human proteins, *Nucleic Acids Res.*, Vol. 39, pp. D487–D493 (online), DOI: <https://doi.org/10.1093/nar/gkq1057> (2011).
- [26] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E.: The Protein Data Bank, *Nucleic Acids Res.*, Vol. 28, No. 1, pp. 235–242 (online), DOI: <https://doi.org/10.1093/nar/28.1.235> (2000).
- [27] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. and Hassabis, D.: Highly accurate protein structure prediction with AlphaFold, *Nature*, Vol. 596, No. 7873, pp. 583–589 (online), DOI: 10.1038/s41586-021-03819-2 (2021).
- [28] 奥 牧人: KampoDB の大幅更新による影響の利用統計に基づく評価, 日本薬学会第 142 年会 (2022). 発表予定.