

日本医療研究開発機構 創薬基盤推進研究事業 事後報告書

I 基本情報

研究開発課題名：PPI 界面三次元構造に基づく PPI 化合物ライブラリー

PPI compound library based on the recognition of the 3D interface structure of PPI

研究開発実施期間：平成 27 年 11 月 2 日～令和 2 年 3 月 31 日

研究開発代表者 氏名：古谷 利夫

Toshio FURUYA

研究開発代表者 所属機関・部署・役職：

ペプチドリーム株式会社 研究開発部 技術顧問

PeptiDream Inc., R&D Department, Technical Advisor

II 研究開発の概要

2003 年のヒトゲノム解読の後、網羅的なオミックス研究が展開された。その一環でタンパク質間相互作用 (PPI) を網羅的に研究するインタラクトームも数十万のオーダーで明らかにされた。PPI はそれまでも有望な創薬標的と考えられてきたが、GPCR や酵素などの従来から研究されてきた創薬標的と比べて創薬の難易度が高く、優先順位は低かったが、近年、従来型の創薬標的が枯渇してきたこともあり、有力な創薬標的となってきた。しかしながら、製薬企業各社が所有する従来の化合物ライブラリーは Lipinsky の Rule of 5 (Ro5 と略されることが多い) に沿った化合物が多く、分子量が 500 以下を基本とした低分子化合物が揃えられている。これに対して、PPI 阻害を目的とした場合は、分子量が 500 以上の、いわゆる中分子領域の化合物が必要とされることが多いようだ。このため、製薬企業各社が所有する化合物ライブラリーでは対応できない場合が多いことから、新たな PPI 阻害を指向した新規な化合物ライブラリーが求められている。各社が同じような化合物ライブラリーを構築するのでは無駄が多く、かつまた時間を要するため、国レベルでのライブラリーの構築が必要とされた。そこで、我々は独自に PPI 阻害ライブラリーを構築することを計画し、薬物分子が標的となるタンパク質と相互作用するためにはタンパク質の三次元構造を認識することが必要であることから、「PPI 界面三次元構造に基づく PPI 化合物ライブラリー」を提案した。

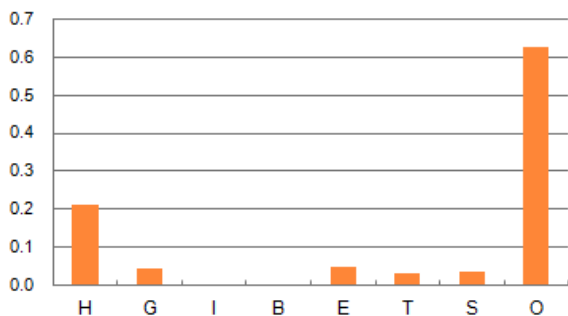


図 1. 既知の標的 PPI の界面における二次構造の出現頻度。二次構造の定義は DSSP に基づく。

H, alpha helix; G, 3-helix (3/10 helix); I, 5 helix (pi helix); B, residue in isolated beta-bridge; E, extended strand, participates in beta ladder; T, hydrogen bonded turn; S, bend; O, others

まず、PPI 複合体の結晶構造を調査した。複合体の相互作用界面を調べてみると、図 1 に示すように、へ

リックスを介して相互作用している場合が約 25%存在していた。また、10%程度ではあるが β シートや γ ターンを介した相互作用も見られるが、60%超は特定の 2 次構造を介さないで相互作用をしていた。そこで、相互作用界面の配列にモチーフ配列 (ELM モチーフ配列の DB に存在する配列) があるかどうかを判定した。最終的に 15,000 化合物を合成する計画であることから、PPI 界面の三次元構造に基づいて、Docking 計算により約 600 万化合物からなるバーチャル化合物ライブラリーから約 12,000 化合物を合成する計画とし、最終的に 11,985 個の化合物の合成を完了した。さらに、これらの化合物の多様性解析を行って、主慣性モーメントを用いて、球状構造を有すると予想される化合物 2,281 個を加えた。また、本プロジェクトでは新規骨格の合成も試みた。その新規構造を含む多様性解析の結果、948 化合物を合成した。合わせて 15,214 個を合成したが、合成化合物の QC チェックを行って最終的に 15,000 化合物を DISC に納品した。全体の 15,000 個の化合物について主慣性モーメント (NRP1&NRP2) で表現した分子の形状を示す分布図を図 2 に示す。逆三角形の右上の部分の球状構造の割合も多くなっていることが分かる。概ね所期の目的を達成したライブラリーが構築されていると判断している。なお、新規母核の創製については、約 20 個の新規骨格を目標に構造を探索し、実際に合成を試みた。具体的には、仮想化合物ライブラリー Generated DataBase (GDB) より scaffold を抽出し、既知化合物ライブラリーである ChEMBL、ZINC、PubChem から抽出した scaffold と比較することにより、新規な scaffold を抽出する。さらに、SciFinder の Substructure search でヒットのない scaffold を新規骨格候補として最終選択し、合成を試みた結果、新規骨格 21 構造を合成した。

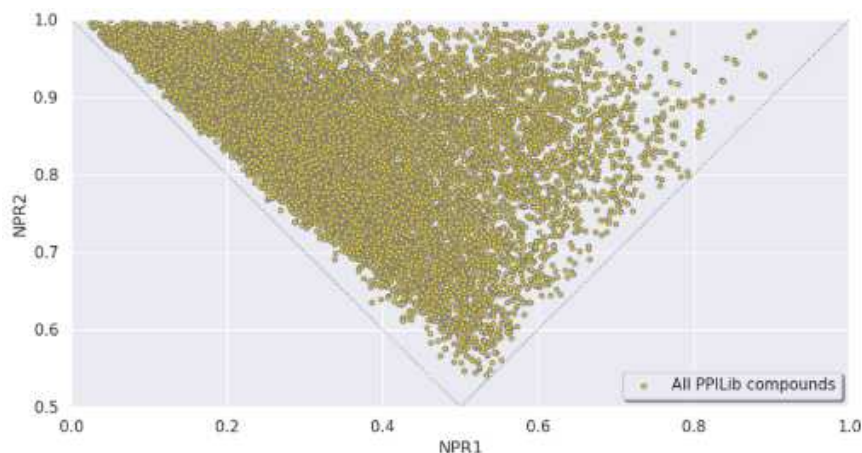


図 2. PPI 化合物ライブラリー (15,000 化合物) の形状解析散布図

以上の結果、本公募要項の「目標」と「求められる成果」に記載の内容は達成した。PPI 標的を対象にした創薬を効率的に推進するためのツールとして、3 年目の後半から本研究で合成された化合物および公開されている PPI 阻害活性を有する化合物の構造情報を格納するデータベース DLiP を開発した。PPI 阻害化合物に関する DB としては TIMBAL、2P2I-DB、iPPI-DB などが知られているが、DLiP は我々が合成した 15,214 化合物に加えて、最新の PPI 阻害化合物に関する文献情報を格納し、現時点で公開されている PPI 化合物関連のデータベースとしては世界初の大規模中分子化合物データベースとなっている。DLiP は 2020 年 4 月から医薬基盤・健康・栄養研究所 (NIBIOHN) の HP に公開した。https://dlip.nibiohn.go.jp/amed_dlip.html

さらに、DLiP の利活用を目的として、AI エンジンを搭載した化合物の活性予測の環境を用意した。AI エンジンと化合物ライブラリーの有用性を検証するために、最終年度に一つの PPI 標的である Keap1/Nrf2 に対して、AI 予測モデルで選抜した化合物にて活性を評価する内容を追加した。600 化合物について、第 1 回目のアッセイを実施し、100 μ M 阻害率で 16 化合物に阻害傾向が見られた。ヒット率は 2.7%である。また、16 個のうち 14 個が AI 予測モデルでスコア上位 (10%以内) に予測された化合物であった。低分子化合物の HTS のヒット率が 100 μ M オーダーで、せいぜい 0.021%とされているので、600 化合物で 16 個 (2.7%) というのは非常に高効率であるといえる。さらに、これらの結果を AI エンジンに学習させて、第 2 回のアッセイを

行った。4種類のAIモデルで予測した312化合物のうち100uM阻害率で45化合物に阻害活性が見られヒット率は14.4%となった。特に、4種類のAIモデルのうちDeep Learningで予測した110化合物で、30化合物がヒットしており、ヒット率は27.3%と驚異的な数字であり、大幅な改善が見られ、良好なAIモデルが構築できていることが示された。また、一つのPPI標的に対して、活性測定を実施したに過ぎないが、概ね良好な化合物ライブラリーができているものと思われる。また、AI予測モデルはアッセイ結果をフィードバックして繰り返し最適化していくとさらに良い結果をもたらすとされており、今回それが示された。RDBの活用とAI予測モデルの活用とそれに基づく予測結果は中分子創薬に極めて有効であることが示唆されたので、今後、多いに本ライブラリーを活用されることを期待したい。

After decoding the human genome in 2003, comprehensive omics research was developed. Many studies of protein-protein interaction (PPI) have been done in part of these. An interactome that comprehensively researches of PPI has been revealed in the order of hundreds of thousands PPIs and are still promising. Although it has been considered as a drug discovery target until then, it has a higher difficulty level and lower priority than drug discovery targets that have been studied in the past such as GPCRs and enzymes, but conventional drug discovery targets are exhausted in recent years. As a result, it has become a powerful target for drug discovery. However, conventional compound libraries owned by pharmaceutical companies are mostly compounds according to Lipinsky's Rule of 5, often abbreviated as Ro5, and low molecular weight compounds based on molecular weight of 500 or less are available. On the other hand, when it comes to PPI inhibition, it seems that compounds with a molecular weight of 500 or more, in the so-called medium sized molecules, are often required. For this reason, there are many cases where the compound libraries owned by the pharmaceutical companies cannot handle the above, and therefore a new compound library directed to new PPI inhibition is required. Since it is wasteful and time consuming for each company to construct a similar compound library, it was necessary to construct the library at the national level. Therefore, we plan to independently construct a PPI inhibitor library, and because it is necessary for the drug molecule to recognize the three-dimensional structure of the protein surface in order to interact with the target protein. Thus a PPI compound library based on the interfacial three-dimensional structure was proposed.

First, crystal structures of the PPI complex were investigated. Examination of the interaction interface of the complex revealed that almost about 25% of the cases interacted via α -helix. In addition, although the interaction through the β -sheet or γ -turn was also observed, although it was about 10% or so. However, in many cases the interaction did not occur through a specific secondary structure. Therefore, it was determined whether the sequence at the interaction interface had a motif sequence, sequences existing in DB of ELM motif sequence. Since we plan to finally synthesize 15,000 compounds, we synthesize about 12,000 compounds from a virtual compound library consisting of about 6 million compounds by Docking calculation based on the three-dimensional structure of the PPI interface. As a result, the synthesis of 11,985 compounds was completed. In addition, a diversity analysis of these compounds was performed based on using the principal moment of inertia, 2,281 compounds expected to have a spherical structure were added. In this project, we also tried to synthesize a new skeleton. As a result of diversity analysis including the novel structure, 948 compounds were synthesized. Although a total of 15,214 compounds were synthesized, QC check of the synthesized compounds was performed and finally 15,000 compounds were installed to DISC. As a result, the contents described in "Aim" and "Required results" in this open call for participants were achieved.

As a tool for efficiently promoting drug discovery targeting PPI targets, we developed a database, DLiP that stores the structural information of the compounds synthesized in this study and also the compounds with published PPI inhibitory activity. DLiP is the world's first large-scale medium-sized molecular compound database to be published as a PPI compound-related database. DLiP has been published on the website of the Institute of Medical Science, Health and Nutrition (NIBIOHN) since April 2020.