

## 海外特別研究員最終報告書

独立行政法人 日本学術振興会 理事長 殿

採用年度 平成 29 年度

受付番号 12

氏名 宮尾知幸

(氏名は必ず自署すること)

海外特別研究員としての派遣期間を終了しましたので、下記のとおり報告いたします。

なお、下記及び別紙記載の内容については相違ありません。

記

1. 用務地 (派遣先国名) 用務地: ボーン (国名: ドイツ)

2. 研究課題名 (和文)

創薬を対象とした実用的な分子設計に関する研究3. 派遣期間: 平成 29 年 4 月 24 日 ~ 平成 30 年 3 月 29 日

4. 受入機関名及び部局名

Department of Life Science Informatics, The University of Bonn

5. 所期の目的の遂行状況及び成果

## [初期の研究目標]

創薬における、新規リード化合物としての低分子化合物を提案するシステムを構築することを研究目標とした。特に化学構造のファーマコフォア情報を加味することが必要と考えた。ファーマコフォアとは、標的マクロ分子との結合に影響を与えると考えられる置換基の三次元における配置 (位置関係) を意味する。また提案リード化合物が毒性をもたず、適切に生体内にて吸収されるなどの経口薬として望ましい特性 (ADMET) をもつことが求められる。そのような化学構造を提案するシステムを開発することを目的とした。

## [研究遂行状況]

- 3次元 Matched molecular pairs を利用した SAR 情報の抽出ならびに、その情報を利用した virtual screening

## [構造活性相関 (SAR) 情報の抽出]

Bioactive なコンフォメーションをもつリガンドとタンパクとの結合情報が、structure-based アプローチにおける唯一信頼できる情報源である。そのため、タンパクとリガンドの X 線結晶構造を利用することとした。[文献 1] から抽出したデータは、3D activity cliff という赴任先である Bajorath 研で長年研究されてきた情報である。2つの類似した化学構造 (3次元構造での類似度が 0.8 以上) が 100 倍以上の異なる活性値をとる構造のペアを activity cliff という。3D activity cliff の例を図 1 に示す。図 1 は serine/threonine-protein kinase Chk1 (Chek1) に活性を持つ 2つのリガンドであるが、2HY0 (magenta) の IC50 が 9.19、2HXQ (cyan) の IC50 が 6.85 と大きな乖離がある。この乖離は 2HY0 のリガンド構造におけるピラゾール基が Check1 の Lys38 に結合しているためと推測される。

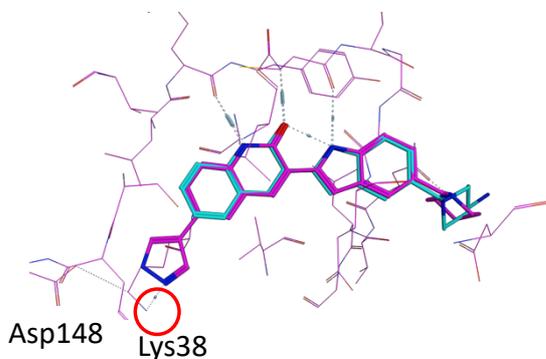


図 1. 3D activity-cliff. Magenta: 2HY0, cyan: 2HXQ (PDBID) in serine/threonine-protein kinase Chk 1.

この情報に基づいて、大量の化学構造をスクリーニングした後の構造を最適化することを考えると、合理的な設計として Lys38 と結合を有するピラゾール基、もしくはその bioisostere を提案することとなる。しかし bioisostere をピラゾール基のみの情報に基づくことは、信頼性に欠ける。そこで、3D activity cliff を形成するそれぞれのリガンドに対する analog を検索し、その analog 同士または、x 線構造を有するリガンドとの新たな 3D activity cliff を検索し、テーブルの形で表現した。重要なことは、analog は ChEMBL19 データベースから抽出しており、大半の活性化化合物の bioactive な conformation は未知であることである。ここでの前提条件 (仮説) は、analog は結晶構造の取得されているリガンドと類似の conformation を取るということである。結果の一例を図 2 に示す。

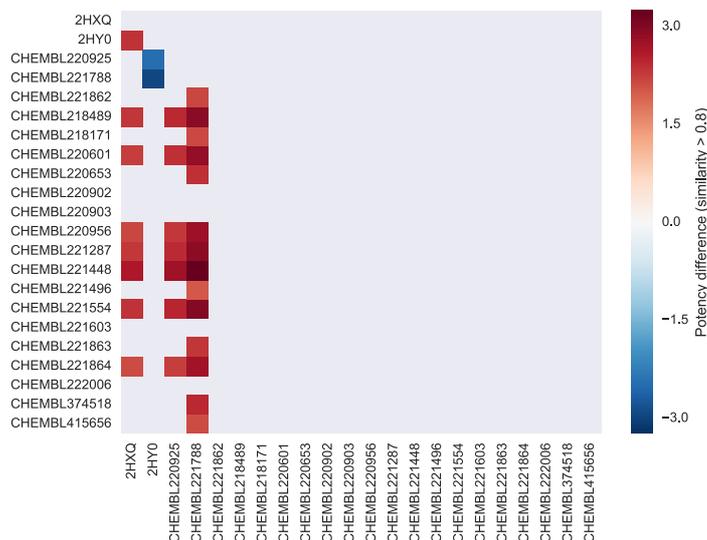


図 2. 3D activity cliff を形成する analog 化合物。活性の差は行 - 列で表現しオーダーが 2 (100 倍) 以上の差があるペアのみをハイライトしている。

ここで、前提条件が真と仮定するならば、ハイライトしているペアは 3D cliff を形成する x-ray 結晶構造と同様に活性に重要と思われる置換基 (部分構造) の情報を保有している。従って、それらの部分構造を抽出することで、screening 後の構造を最適化する指針となることが期待される。以上の解析を文献 1 のデータをもとに、systematic に実施した。対象として、thrombin, phosphodiesterase 4B, serine/threonine protein Chk1, estrogen receptor alpha を採用した。これらのルールは virtual に構築した分子を最適化する際のルールとして使用することが期待される。

#### [3D activity cliff を利用した virtual screening の実施]

3D activity cliff の情報を利用することで、1. virtual screening の精度が上がるか、また、2. 活性の高い構造を優先的に選択しているかどうかを判定する必要がある。1 については、Dekois [文献 2] をスクリーニングデータとした結果、activity cliff を利用することによる有意な差は観測されなかった。ここでは activity cliff から得られたルールを activity cliff による影響  $(x) = \text{類似度}(x, \text{high potent x-ray ligand}) - \text{類似度}(x, \text{less potent x-ray ligand})$  で定義した。類似度  $(x, y)$  は分子  $x$  と  $y$  の類似度を表す。続いて、2 のテストを ChEMBL22 の活性化化合物データセット (最大 1000 化合物/標的) を用いて実施した。スクリーニング対象

データを正例は活性上位 20%、負例は残りの 80%とし、AUC-ROC を評価手法とした。それぞれのスクリーニング化合物のコンフォメーションを OMEGA [文献 3]により発生させ、3次元構造同士の重ね合わせ(Tanimoto Combo Score)をスクリーニング化合物のスコアとした(テンプレート対象として使用したデータセットを table 1 にまとめ、例として estrogen receptor alpha と heat shock protein 90-alpha (HSP90)を対象とした結果を Table 2 に示す。

Table 1. スクリーニング対象データセット ()はサンプリング前の化合物数

|   | # ChEMBL cpds. | max pKi | mean pKi | min pKi | std  | # cliffs [pKi & IC50] |
|---|----------------|---------|----------|---------|------|-----------------------|
| P00734-Thrombin                             | 1000 (1157)    | 12.19   | 6.36     | 1.00    | 1.62 | 166                   |
| P56817-Beta-secretase 1                     | 472            | 9.60    | 6.73     | 1.65    | 1.26 | 57                    |
| P07900-Heat shock protein HSP 90-alpha      | 88             | 8.70    | 7.29     | 5.08    | 0.78 | 43                    |
| P24941-Cyclin-dependent kinase 2            | 89             | 9.00    | 6.82     | 4.70    | 1.14 | 41                    |
| P00742-Coagulation factor X                 | 1000 (1672)    | 11.40   | 7.49     | 3.59    | 1.76 | 28                    |
| P00918-Carbonic anhydrase II                | 1000 (2895)    | 11.10   | 6.92     | 0.60    | 1.50 | 24                    |
| Q16539-MAP kinase p38 alpha                 | 102            | 9.70    | 7.88     | 5.72    | 1.13 | 13                    |
| P00749-Urokinase-type plasminogen activator | 187            | 9.00    | 6.68     | 4.25    | 1.18 | 12                    |
| P03372-Estrogen receptor alpha              | 201            | 10.04   | 7.63     | 4.42    | 1.19 | 10                    |

Table 2. スクリーニングパフォーマンス (estrogen receptor alpha (上)、heat shock protein 90-alpha (下))

| Cliff pair | AUC                      |                                    |                                  | #actives in top30        |                                    |                                  |
|------------|--------------------------|------------------------------------|----------------------------------|--------------------------|------------------------------------|----------------------------------|
|            | Pharmacophore Difference | Similarity to Highly Potent ligand | Similarity to less potent ligand | Pharmacophore Difference | Similarity to Highly Potent ligand | Similarity to less potent ligand |
| 2R6Y-1XQC  | 0.62                     | 0.75                               | 0.74                             | 8                        | 22                                 | 19                               |
| 2AYR-1XQC  | 0.75                     | 0.75                               | 0.74                             | 18                       | 20                                 | 19                               |
| 1XP6-1XQC  | 0.54                     | 0.72                               | 0.74                             | 5                        | 20                                 | 19                               |
| 1XP1-1XQC  | 0.60                     | 0.71                               | 0.74                             | 7                        | 21                                 | 19                               |
| 2R6W-1XQC  | 0.64                     | 0.71                               | 0.74                             | 11                       | 22                                 | 19                               |
| 3ERD-1GWQ  | 0.61                     | 0.51                               | 0.52                             | 2                        | 0                                  | 6                                |
| 3ERD-1X7R  | 0.57                     | 0.51                               | 0.51                             | 1                        | 0                                  | 4                                |
| 3ERD-1X7E  | 0.59                     | 0.51                               | 0.55                             | 1                        | 0                                  | 5                                |
| 3ERD-2FAI  | 0.59                     | 0.51                               | 0.43                             | 3                        | 0                                  | 6                                |
| 2POG-1X7R  | 0.44                     | 0.46                               | 0.51                             | 6                        | 5                                  | 4                                |

| Cliff pair | AUC                      |                                    |                                  | #actives in top30        |                                    |                                  |
|------------|--------------------------|------------------------------------|----------------------------------|--------------------------|------------------------------------|----------------------------------|
|            | Pharmacophore Difference | Similarity to Highly Potent ligand | Similarity to less potent ligand | Pharmacophore Difference | Similarity to Highly Potent ligand | Similarity to less potent ligand |
| 4EGK-4EGH  | 0.76                     | 0.58                               | 0.21                             | 12                       | 9                                  | 2                                |
| 3BMY-2YJW  | 0.64                     | 0.58                               | 0.24                             | 9                        | 8                                  | 2                                |
| 3BMY-2CCT  | 0.70                     | 0.58                               | 0.30                             | 11                       | 8                                  | 4                                |
| 2VCI-2CCT  | 0.44                     | 0.57                               | 0.30                             | 5                        | 8                                  | 4                                |
| 2VCI-2CCT  | 0.43                     | 0.54                               | 0.30                             | 2                        | 6                                  | 4                                |
| 2FWZ-1UY6  | 0.68                     | 0.32                               | 0.27                             | 9                        | 2                                  | 4                                |
| 2FWZ-1UYH  | 0.58                     | 0.32                               | 0.14                             | 6                        | 2                                  | 0                                |
| 2FWZ-1UYE  | 0.72                     | 0.32                               | 0.22                             | 10                       | 2                                  | 2                                |
| 2FWZ-1UYF  | 0.72                     | 0.32                               | 0.21                             | 10                       | 2                                  | 2                                |
| 2FWZ-1UYG  | 0.64                     | 0.32                               | 0.34                             | 7                        | 2                                  | 5                                |
| 3BM9-4EEH  | 0.63                     | 0.27                               | 0.24                             | 10                       | 3                                  | 3                                |

Estrogen receptor alpha を用いた結果、3D activity cliff の情報を取り込むことによる効果を観測することはできなかったが、HSP90 を標的とした場合、明らかな精度の向上が観測された(今回の 9 つの標的では HSP90 のみ)。Activity cliff を利用することでスクリーニング能力が上昇するメカニズムを解明することを今後の課題とするが、その明確な理由がわからない以上は、activity cliff の情報をスクリーニングに使用することはできない。

以上から、これらの解析において、1. リガンドを最適化するための指針としての activity cliff 情報の収集、2. 仮想的スクリーニングにその情報を使用するメリットは少ないことが明らかになった。

## 2. Inverse QSAR/QSPR を利用した望ましい予測物性・活性を持つ化学構造の提案手法の開発

Inverse QSAR/QSPR では通常、分子の 2次元構造に基づいて、望ましい予測物性や活性を持つ化合物を提案する。以前提案した混合正規分布と線形重回帰を組み合わせた手法[文献 4]では、モデルの適用範囲を考慮することはできるが、予測精度が低いという問題点があった。そこで、differential evolution (DE)により、support vector regression (SVR)と one-class support vector machine の制約付き最適化問題を解くことにより、活性(物性)が望ましいと期待される 1 点を導き、その点に記述子の値を持つ化学構造を生成する手法を提案した。成果として、論文を発表した。

発表論文: Miyao T, Funatsu K, Bajorath J, Exploring differential evolution for inverse QSAR analysis, *F1000Research*, 2017, 6(Chem Inf Sci):1285.

### 3. 仮想的な化合物を生成することによる、化学合成プロジェクト評価手法の開発

製薬企業等におけるリード化合物を合成するプロジェクトにおいては、スクリーニングなどで得られた hit 化合物 (弱い活性を持つ化合物) を、合成展開することでプロジェクトは進行する。問題となるのが、現在進行中のプロジェクトをいつ終了させるべきか、また、複数のプロジェクトをどのように比較するのか、ということである。この問題を解決することで、ほぼ目の出していないプロジェクトを中止するなど、企業におけるコスト削減の効果は大きい。この問題を解決するため、analog series (analog 構造群) に着目した。Analog series のコアの骨格から、仮想的な analog を生成し、それら構造群の分布とプロジェクトにおいて現在合成された化合物群の比較に基づく指標 (local and global scores) を開発した。この研究成果を論文に発表した。

発表論文: Kunimoto R\*, Miyao T\*, Bajorath J, Computational method for estimating progression saturation of analog series, *RSC Advances*, 2018, 8, 5484 (\*: equal contribution)

### 4. 機械学習を利用した物性予測または活性予測モデルの構築

複数の物性や活性 (ドラックライクネス、毒性) を ligand に基づく情報から判断するためには、定量的構造活性 (物性) 相関 (QSAR (QSPR)) モデルを利用することが現実的である。最近では深層学習を利用したモデルによる活性・物性予測の精度向上の報告があるが、実際のスクリーニングデータにおける inactive 化合物への偏り (hit rate は 3% 以下) を考慮した報告は少ない。所属研究室において、PubChem のデータベースに含まれる化合物から、false positive を除外した化合物のみで構成した約 10 万化合物 × 49 assays (= 標的) の全ての値が実験結果として報告された matrix 形態のデータを構築するプロジェクトがあった。その matrix 形態のデータを現実的なスクリーニングデータとして利用できると考え、種々の回帰モデルを適用することで、回帰モデルの性能評価を行なった。結果として、すべての種類の回帰モデル (深層学習やマルチタスク学習を含む) の中で、random forest (RF) が最も予測性が安定して高い性能を示した。この結果から、QSAR/QSPR モデルをフィルタとして発生させた構造に適用する場合、RF で十分であるとの知見が得られた。研究成果として論文を投稿中である。

投稿論文: Pérez R., Miyao T, Jasial S, Vogt M, Bajorath J, Prediction of Compound Profiling Matrices Using Machine Learning, *ACS Omega*, submitted

#### 文献

1. *RSC Adv.*, 2015, 5, 43006-43015.
2. *J. Chem. Inf. Model*, 2015, 55, 2297-2307.
3. *J. Chem. Inf. Model*, 2010, 50, 572-584.
4. *J. Chem. Inf. Model*, 56, 2, 286-299

#### [まとめ]

今回の派遣期間では、実際に de novo 構造を生成するシステムを構築することはできなかったが、そのシステムを構築するための知見を獲得することができ、概ね計画通りの進行となった。また、プロジェクトに付随する解析も複数実施し、成果を発表することができた。