

構造活性相関部会・ニュースレター

<14 April, 2015>

SAR News No.28

「目次」

//// Perspective/Retrospective ////

超分子モデリングパイプラインによる創薬支援

土方 敦司、辻 敏之、白井 剛 … 2

//// Cutting Edge ////

天然変性タンパク質の機能予測法について

Munmee Dutta, Ashwini Patil, 中井 謙太 … 8

PoSSuM: ポケット類似性情報に基づく合理的薬剤設計支援に向けて

池田 和由、富井 健太郎 … 14

//// SAR Presentation Award ////

2014 年度選考結果について … 21

受賞コメント … 22

授賞講演要旨 … 23

//// Activities ////

<報告>

第 42 回構造活性相関シンポジウム開催報告 今井 輝子 … 29

第 1 回 QSAR 解析ハンズオンセミナー開催報告 山下 富義 … 30

<会告>

構造活性フォーラム 2015

「様々なリード創製手段において創薬に貢献するインシリコ技術」 … 31

第 43 回構造活性相関シンポジウム

第 10 回薬物の分子設計と開発に関する日中合同シンポジウム 共同開催 … 32

///// Perspective/Retrospective /////

超分子モデリングパイプラインによる創薬支援

長浜バイオ大学 土方敦司、辻 敏之、白井 剛

1. はじめに

タンパク質や核酸など生体分子の大部分は、細胞内では複合体(超分子複合体)を形成して複雑な生体機能を司っている。従って、超分子複合体の構造やタンパク質間の相互作用面(分子間の接触に関与する分子表面。以下インターフェースとよぶ)は、効率的なドラッグターゲットの選択や、ドラッグによる機能制御の予測に必須の情報である。しかしながら、超分子複合体の立体構造やタンパク質間インターフェースを実験的に求めることは、多くのケースについて未だに困難な課題である。

そこで我々の研究グループでは、超分子複合体構造を計算機を用いて組み上げるためのバイオインフォマティクス技術を高度化し、さらにはそれらの統合により超分子モデリングパイプラインを構築して、超分子複合体の研究を行う実験系研究者を支援することを目的として研究を進めている(図1)。超分子モデリングパイプラインは、(1) データベースなどの情報を統合する知識モデリングシステム、(2) 未知の複合体構造を予測するマルチドメインモデリングシステム、(3) 粗視化分子動力学計算など利用したモデルリファインメントシステム、(4) GUI を活用したモデル評価支援システムから構成される multi-disciplinary な計算機システムである。ここでは、(1) 知識ベースモデリングにおいて、立体構造データベースと相互作用データベースの情報を統合し、超分子複合体構造モデルを予測する研究に焦点をあてて解説する。

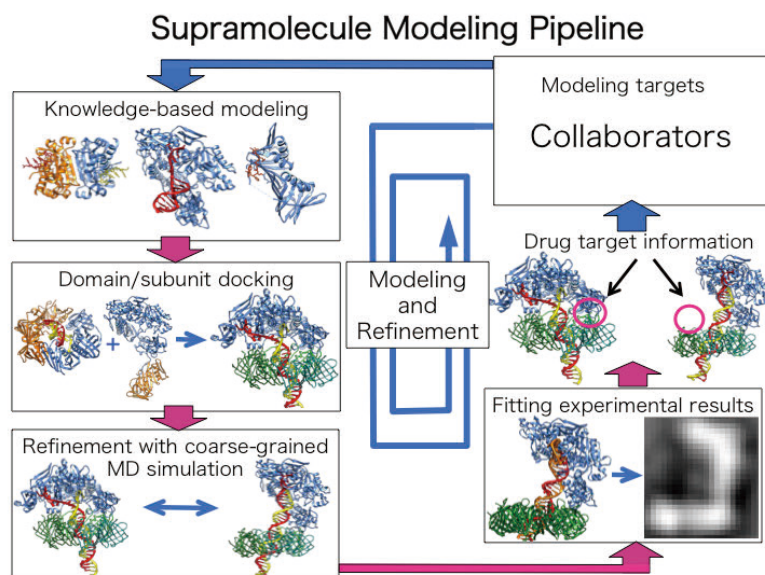


図1. 超分子モデリングパイプライン概念図

知識ベースモデリング(Knowledge-base modeling)、ドッキング(Domain/subunit docking)、リファイン(MD simulation)、実験データとの比較(Fitting experimental results)のインタラクティブな繰り返しにより複合体モデルが構築される。

2. タンパク質相互作用データと既知構造を用いた複合体モデリング法の開発

ヒトゲノムに代表されるように、ゲノムの DNA 配列情報はすでに多数の生物種について完成されている。一方、そこから発現するタンパク質やタンパク質複合体の立体構造情報について完

成されている生物種は、現時点で1つも存在しない。このゲノムと立体構造の完成度の差は、現在の生物学が抱える重大な情報ギャップの一つである。しかし、構造ゲノミクスの発展によって、すでに生体分子立体構造情報の蓄積は「決して小さくない」レベルに達している。

我々は、このギャップを計算科学的アプローチにより解決するために、実験データベースからの知識抽出と統合によるタンパク質複合体モデリング手法を開発している。巨大複合体構造を決定することは困難であっても、構造解析技術の発展により2量体や3量体のタンパク質複合体の立体構造は比較的多く決定され、タンパク質立体構造データベース PDB[1]に登録されている。また、実験的にタンパク質間相互作用を検出する技術も進歩しており、それらはタンパク質間相互作用データベース (IntAct[2], MINT[3], STRING[4]など) に集積され、公開されている。

そこで我々は、タンパク質間相互作用データベース IntAct とタンパク質立体構造データベース PDB を組み合わせて、タンパク質間相互作用ネットワークからタンパク質複合体を網羅的にモデリングする手法を開発した。

2.1 複合体モデリングに用いることができるデータ

複合体モデリングのストラテジーは極めて単純である。例えば図2に示すような3つのタンパク質 A、B、C からなる相互作用ネットワーク (C-A-B) があつた場合に、

- 1) タンパク質 A と B および A と C の複合体構造がそれぞれ明らかになっている
- 2) 2つの複合体において A が大きな構造変化を起こしていない
- 3) 2つの複合体構造を A で重ねあわせたときに B と C が衝突しない

これら3つの条件を満たせば、2つの複合体構造をタンパク質 A (重ね合わせに用いるタンパク質をアンカータンパク質と呼ぶ) で重ね合わせることで相互作用ネットワークに基づく3量体モデルを構築することができる。

この考え方をベースにして、複合体モデルを構築するアルゴリズムを考案した。このアルゴリズムで、アンカータンパク質による重ね合わせのあと、アンカータンパク質に隣接したタンパク質にアンカータンパク質を移動することで、次々と相互作用するタンパク質をつなぎ合わせて、より大きな複合体構造をモデリング事も可能である。この過程は、アンカーにできるタンパク質が無くなる、または新たに導入できるタンパク質候補が無くなるまで繰り返される。

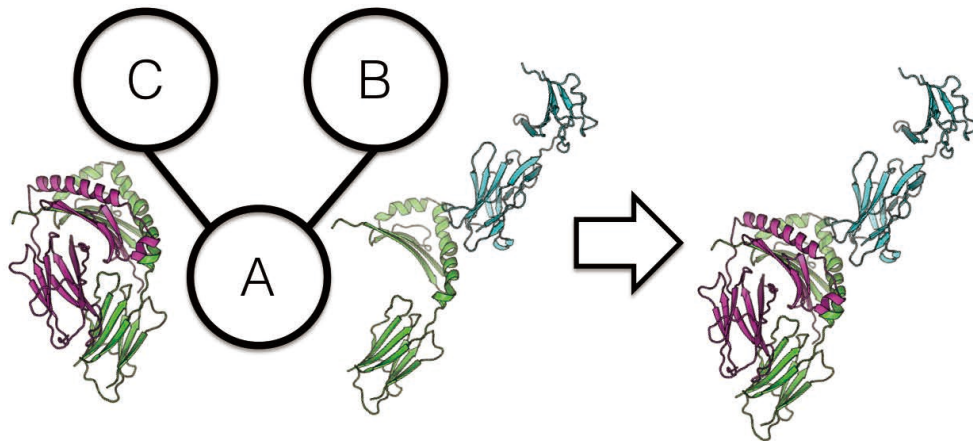


図 2.2 量体の集積による3量体構造のモデリング

図左は C(赤)-A(緑)類縁タンパク質と A(緑)-B(水色)類縁タンパク質の2量体であり、共通する A の重ね合わせにより、図右の3量体モデルが構築される。

モデリング可能な複合体は、タンパク質間相互作用ネットワークデータベースに登録されていて、相互作用するタンパク質ペアの複合体構造が実験的に解かれている必要がある。しかしながら、相互作用するタンパク質そのものの立体構造が決定されていることは稀であるので、ホモロジー(分子の進化的な類縁性。類縁分子は構造が保存される)に基づくモデリングを利用する。この場合、複合体構造をモデリングしたいタンパク質ペアとそれぞれ 25%以上のアミノ酸一致

度を持つタンパク質複合体の構造（テンプレート構造）が既知であれば、その複合体中のインターフェースは、目的の複合体と類似しており、モデリング可能であると期待できる。この条件を満たすデータは IntAct に収録されている相互作用全体の 8%、ヒトタンパク質の相互作用に限定すると 14%であった。ここでは、このヒト由来タンパク質の相互作用ネットワークについてモデリングを行った結果を解説する。

2.2 複合体モデリングによって新規に予測されたモデル

IntAct と PDB のデータを組み合わせて得られたモデリング可能な相互作用ネットワークには、2,884 個のヒトタンパク質と、それらの間の 5,455 の相互作用が含まれており、上に述べた方法によって 3,197 個の独立したヒトタンパク質複合体モデルが構築された(論文投稿中)。しかし、IntAct に登録された相互作用のうち 6%は、モデル構造上では直接相互作用せず、1 個以上の別のタンパク質を介してして相互作用していた。IntAct は様々な実験手法による相互作用データを収録しているが、免疫沈降法などによる実験では共沈降（あるタンパク質を抗体などで沈降した時に、同時に沈降するタンパク質を同定する）を指標とするため、間接的に相互作用する例が多数登録されていることに由来すると考えられる。

最も大きい複合体は 17 量体であり、proteasome の 26S サブユニットの一部であった。得られた複合体の平均サブユニット数は 3.3 で、55% (1,756/3,197) の複合体は 2 量体であった。2 量体が多い理由は、全ヒトタンパク質の 80%が、それぞれ 1 つしかインターフェースを持たないためである。

全複合体モデルに存在するインターフェースを解析した結果、3,959 個がユニークであり、このうち 21% (970/3,959) は IntAct に登録されていない新規の(モデリングにより予測された)インターフェースであった。また、複数のタンパク質と相互作用することができるオルタナティブインターフェースは全インターフェースの 35% (1,370/3,959) にものぼることが示唆された。オルタナティブインターフェースは実験的に同定することが比較的困難であり、同時には相互作用することのできないタンパク質ペアなどの重要な情報を提供する。

構築したモデルには、実験的に明らかにされている複合体構造をさらに拡張したものが含まれる。一例として cyclin A2-cyclin dependent kinase-CDK inhibitor 1-ubiquitin ligase complex の複合体モデルを図 3 に示した。

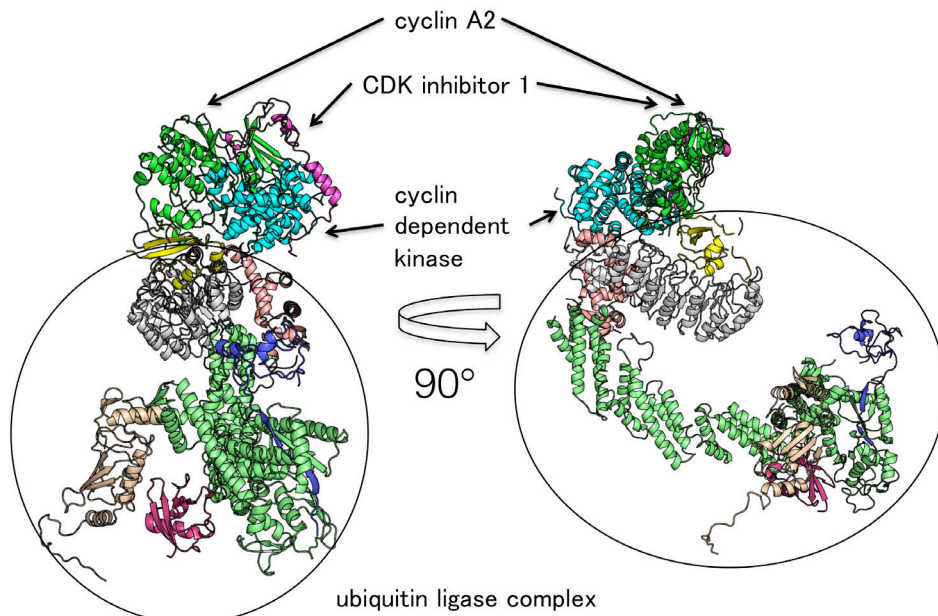


図 3. cyclin A2-cyclin dependent kinase-CDK inhibitor 1-ubiquitin ligase complex モデル
それぞれのタンパク質は異なる色で示されている。丸で囲まれた部分が ubiquitin ligase complex である。

Cyclin A2は細胞周期に関わるタンパク質で、cyclin dependent kinase (CDK)によってリン酸化されることで活性化し、細胞周期を進行させるトリガーになる。通常、CDKはCDK inhibitor 1が結合することにより不活性化されているが、CDK inhibitor 1は細胞周期依存的にubiquitin ligase complexによってユビキチン化されproteasomeによって分解される。これによってCDKが活性化し、続けてcyclin A2がリン酸化され細胞周期が進行する。この過程においては、ubiquitin ligase complexによってCDK inhibitorの天然変性領域(通常の状態では特定の立体構造を形成しないタンパク質領域)がユビキチン化を受けることが示唆されている。今回構築された図3の複合体モデルの部分構造は、これまでも結晶構造解析の組み合わせモデルとして報告されているが[5]、本手法により網羅的にデータベースからの知識を組み合わせることで、さらに拡張が可能であることが示された。

3. タンパク質複合体モデルへの疾患原因変異のマッピングとその解析

次世代シーケンサー技術の目覚ましい進展は、個人レベルでのゲノム配列解読を容易にし、これまで原因不明とされてきた様々な遺伝性疾患の責任遺伝子およびその変異の特定を可能とした。こうしたゲノムワイドな塩基配列の解析は、遺伝子上の変異と疾患の関連性を見つけ出す目的においては有用である。しかしその一方で、特定の遺伝子変異がどのようにして病気を引き起こすのかという、疾患の発症機序についての分子メカニズムはブラックボックスである場合が多い。また、こうした疾患原因変異情報の蓄積によって、同一遺伝子を責任遺伝子としながら、異なる表現型を持つ疾患が多く知られるようになってきており、遺伝子型と表現型との関連性を塩基配列レベルの解析のみから理解することは容易ではない。遺伝子機能を司るのは、多くの場合その遺伝子がコードするタンパク質である。我々は、疾患変異とタンパク質の立体構造、特に複合体構造との関係に着目して研究を進めている。

3.1 Mutation@A Glance

疾患の原因となる遺伝子上の変異をタンパク質立体構造上に投影(位置を特定すること。以下マッピングと呼ぶ)することは、変異が遺伝子機能にどのような影響を与えるかについて、原子レベルで解析する上で有効な手段である。我々は、こうした解析を容易とするためのデータベースMutation@A Glanceを開発し公開している(<http://harrier.nagahama-i-bio.ac.jp/mutation/>)。このデータベースは、ヒトの遺伝性疾患の原因変異および疾患とは関連性のない塩基バリエーション(SNV)について、塩基配列、アミノ酸配列、タンパク質立体構造およびその複合体構造の各階層についてマッピングしたデータを可視化し、遺伝子の階層横断的な解析を可能とする[6]。

Mutation@A Glanceを利用して、前節で解説した3,197個のヒトタンパク質複合体モデル上に疾患に関連するSNVによるアミノ酸変異をマッピングしたところ、複合体モデルに含まれるタンパク質領域にマッピングされたもののうち約20% (456/2,319)がインターフェースに存在した。このうち、モデリングによって新規に予測されたインターフェースに存在するものは、12% (54/456)に止まったが、単位面積あたりにマッピングされた数は、既知インターフェースと予測インターフェースでほぼ同等であった($10,000\text{\AA}^2$ あたり0.13個のSNVが存在する)。

解析した487種類の疾患のうち、26種類の疾患に関連したSNVが複合体モデルのインターフェース領域にマッピングされた。さらにこのうち11種類は、新規に予測したインターフェースにマッピングされることがわかった。このことから、複合体構造におけるインターフェース破壊は疾患原因として無視できない存在であり、また複合体モデル構築が疾患原因の解析に有効であることが示唆される。

3.2 疾患の表現型と変異の立体構造的な特徴との相関

さらにこのデータベースを利用して、同一責任遺伝子上の異なる変異によって異なる種類の疾患を引き起こす約260例に着目し、変異とタンパク質立体構造上の特徴について解析を行った。変異が引き起こす遺伝性疾患の遺伝様式の違いによって分類したところ、常染色体優性型(AD)の遺伝様式を示す変異は、タンパク質分子表面のアミノ酸残基に起こる場合が統計的に有意に多く、逆に、常染色体劣性型(AR)の変異部位はタンパク質の内部のアミノ酸残基に有意に多い

という結果が得られた。我々を含む多くの研究グループによって、疾患との関連性のない SNV は、タンパク質立体構造上表面に多いことが示されているが[7, 8]、AD が表面に多いという我々の結果はどのように解釈できるだろうか？そこで、タンパク質分子間のインターフェースに着目して、ヒトタンパク質複合体立体構造モデルと変異との関係について解析を進めた。すると、AD の表現型となる変異は、AR や疾患との関連を無視した SNV 全体と比べて統計的に有意にインターフェースのアミノ酸残基であることが多いことが明らかとなった[9]。すなわち、(1) タンパク質内部に起きる変異は、タンパク質立体構造を壊すため、おそらく品質管理によって分解されることで、両方の対立遺伝子に変異が無い限り遺伝子異常が表現型として現れない。(2) タンパク質表面でインターフェースへの変異は、おそらくタンパク質立体構造は壊さないで、タンパク質の分子間ネットワークに対してドミナントネガティブ効果を与える、という遺伝子変異と疾患形態との関連性がタンパク質立体構造モデルを通した解析によって見えてきた。

3.3 構造モデルによる新規疾患関連変異の評価

これらの計算機手法やデータベースを利用した具体的な疾患関連変異の解析例として、京都大学大学院医学研究科(平家俊男教授ら)のグループとの共同研究による、アイカルディ・ゴーティエ症候群 (Aicardi-Goutières syndrome, AGS) 患者から同定された新規責任遺伝子 IFIH1 のミスセンス変異とタンパク質立体構造および機能との関連性の解析を紹介する。

AGS は発達遅滞、小頭症、頭蓋内石灰化を伴う遺伝性脳症であり、I 型インターフェロン産生の亢進によって誘導される自己炎症が疾患の原因と考えられている。責任遺伝子が未同定の AGS 患者 3 人のエクソーム解析を行ったところ、3 患者に共通して IFIH1 遺伝子にそれぞれ異なるミスセンス変異が見出された。IFIH1 遺伝子は細胞内の核酸受容体 MDA5 をコードしており、外来 2 本鎖 RNA と結合しオリゴマーを形成することで、そのシグナルが下流に伝わることで、I 型インターフェロン産生が亢進することが知られている。我々は MDA5 オリゴマー-RNA の複合体モデルを構築し、これら 3 種類の変異をマッピングしたところ、RNA 結合およびオリゴマー形成に直接影響を与える部位に位置していることがわかり、これらの変異が疾患を引き起こす変異であることが予測された (図 4)。患者由来の細胞を用いた実験によって、確かにこれらのミスセンス変異が MDA5 の機能を亢進していることが証明された [10]。

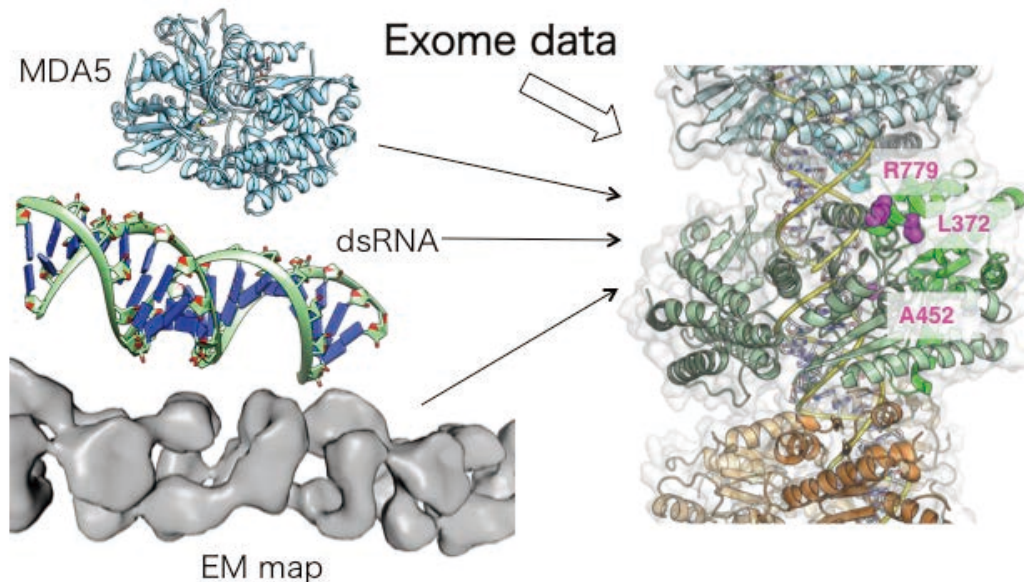


図 4. MDA5-RNA 複合体モデル上の AGS 原因アミノ酸変異

エクソーム解析によって同定された疾患関連変異 (図右) は、MDA5 構造、2 本鎖 RNA (dsRNA)、電顕マップなどから構築された複合体モデル上でインターフェースにマッピングされる。

4. おわりに

解説したタンパク質相互作用ネットワークを用いた複合体モデリングにより、実験的には未知のタンパク質インターフェースが予測可能であり、それらの予測は未解明の分子メカニズムの解析につながる。この方法は、IntAct や PDB などのデータベースに含まれるデータ(知識)量が増加すればするほど、構築可能なモデルの数も加速度的に増加するため、将来にわたって良好なポジティブフィードバックが形成されることが期待できる。

また、こうしたタンパク質立体構造および複合体モデルを用いた変異解析は、様々な遺伝性疾患における疾患メカニズムの理解につながるだけでなく、病気の治療に向けた創薬においても有用であると考えられる。例えば、先の AGS では、変異 MDA5 の機能の異常亢進が原因と考えられることから、モデル上で MDA5 のオリゴマー形成、あるいは 2 本鎖 RNA との結合を阻害する薬剤をデザインすることができれば、疾患の原因を抑えることができるかもしれない。複合体モデルはこのような薬剤ターゲットサイトの絞り込みに特に有用である。

タンパク質複合体モデル立体構造に基づく創薬はまだ研究段階であるが、我々は、実験データとバイオインフォマティクスを統合して詳細に解析することによって、近い将来実用段階に近づくと確信している。

謝辞

ここで解説した研究は、創薬等ライフサイエンス研究支援基盤事業「超分子モデリングパイプラインの構築」の支援を受けて、多くの共同研究者との協力により推進されたものであり、関係各位、また、研究協力にご同意いただいた患者さんご家族に感謝申し上げます。

参考文献

- [1] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- [2] Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., Hermjakob, H. The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–63 (2014).
- [3] Chatr-aryamontri, A., Ceol, A., Montecchi-Palazzi, L., Nardelli, G., Schneider, M. V., Castagnoli, L., Cesareni, G. MINT: the Molecular INTERaction database. *Nucleic Acids Res.* **35**, D572–574 (2007).
- [4] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L. J. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–815 (2013).
- [5] Hao, B., Zheng, N., Schulman, B. A., Wu, G., Miller, J. J., Pagano, M., Pavletich, N. P. Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. *Mol. Cell* **20**, 9–19 (2005).
- [6] Hijikata, A., Raju, R., Keerthikumar, S., Ramabadran, S., Balakrishnan, L., Ramadoss, S. K., Pandey, A., Mohan, S., Ohara, O. Mutation@A Glance: an integrative web application for analysing mutations from human genetic diseases. *DNA Res.* **17**, 197–208 (2010).
- [7] Wang, Z., Moulton, J. SNPs, protein structure, and disease. *Hum. Mutat.* **17**, 263–70 (2001).
- [8] Sunyaev, S., Ramensky, V., Bork, P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet.* **16**, 198–200 (2000).
- [9] 土方 敦司, 小原 収, 「ヒト遺伝性疾患の構造バイオインフォマティクス」 日本遺伝学会第 86 回大会 (2013)
- [10] Oda, H., Nakagawa, K., Abe, J., Awaya, T., Funabiki, M., Hijikata, A., Nishikomori, R., Funatsuka, M., Ohshima, Y., Sugawara, Y., Yasumi, T., Kato, H., Shirai, T., Ohara, O., Fujita, T., Heike, T. Aicardi-Goutieres syndrome caused by *IFIH1* mutations. *Am. J. Hum. Genet.*, **95**, 121–125 (2014).

//// Cutting Edge ////

天然変性タンパク質の機能予測法について

東京大学医科学研究所 Munmee Dutta, Ashwini Patil, 中井 謙太

1. はじめに

筆者の浅学ゆえの誤解であればお許しいただきたいが、タンパク質の構造活性相関を研究する立場では、タンパク質が一定の三次元構造[1]をとっていることは自明とされているのではないだろうか。しかし、かれこれ 20 年程前から、生理的条件下でも一定の三次元構造をとらないタンパク質が次々と同定されるようになってきた。それらは、天然変性タンパク質 (intrinsically disordered protein; IDP) と総称され、通常のタンパク質と比べて偏ったアミノ酸組成を持ち、また一定の立体構造をとり続けてはいないという性質そのものがしばしばその機能と結びついてることが特徴的である[2]。天然変性タンパク質はその構造全体が一定構造を持たないわけではなく、そのような部分は天然変性領域 (intrinsically disordered region; IDR) に局在しており、また天然変性領域をもつタンパク質は真核生物に多く見られることが知られている[3]。天然変性領域は複数のタンパク質と相互作用するタンパク質に多く見られる[4, 5]。このことは、異なる結合相手によって、変性領域がそのコンフォメーションを適合させていることによるものと考えられている[6]。最近のいくつかの報告によれば、天然変性タンパク質の中には、構造状態と変性状態間の秩序・無秩序転移を伴わず、柔軟な状態のままで相互作用することができるものもあるらしい[7]。

天然変性タンパク質は、この柔軟性によって、転写制御・細胞内のシグナル伝達や調節パルスウェイに関与している[8-10]。天然変性領域は、腫瘍抑制タンパク質 p53 や DNA 修復タンパク質 BRCA1 のように、多くの疾病関連タンパク質にも見られる。天然変性領域は PTEN タンパク質の例で見られるように、タンパク質や DNA、RNA と直接結合することで機能しているのかもしれない[11]。また、カルモジュリンの例で見られるように、異なる結合タンパク質と結合するときに固定構造ドメインの相対位置を調節するリンカーとして機能しているのかもしれない[12]。天然変性領域は、真核生物のタンパク質の約三分の一に存在すると推定され、その機能の重要性からみても、決して例外として無視すべき存在ではなく、何とかしてその (一次) 構造と機能との関係を理解して、機能予測法を開発することが望まれる。しかし、伝統的なタンパク質の機能予測法で用いられている方法の多く[13] は、固定構造を持たず、アミノ酸配列の進化的保存も明瞭でないことが多い天然変性タンパク質には適用不能と考えられる。そこで、本稿では、この困難な天然変性タンパク質の機能予測問題に対して行われている様々な試みを紹介する。まだまだ研究例も少なく、まとまった総説にはなっていないが、ここで紹介した内容が、少しでも構造活性相関研究者の参考になれば幸いである。

2. 天然変性タンパク質の機能予測

上述のように、天然変性領域は、荷電性残基や極性残基に富み、疎水性残基や芳香性残基に乏しいといった偏ったアミノ酸組成で特徴付けられる[14]。ここ数年間に発表された研究によれば、変性領域の一次構造は、機能予測に利用できる情報を含んでいる。それらの研究で発表された代表的なツールを表 1 に示し、以下にその内容を紹介する。

表 1. 天然変性タンパク質の機能予測や結合部位予測に利用できるツール

ツール	予測対象	ウェブサイト	文献
FFPred 2.0	機能	http://bioinf.cs.ucl.ac.uk/psipred/	[15]
IDD Navigator	機能	http://sysimm.ifrec.osaka-u.ac.jp/disorder/beta.php	[16]

SLiMPrints	モチーフ	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/slimprints.php	[17]
PepBindPred	結合部位	http://bioware.ucd.ie/~testing/biowareweb/Server_page/s/pepbindpred.php	[18]
ANCHOR	結合部位	http://anchor.enzim.hu/	[19]
MoRFPred	MoRF (注)	http://biomine-ws.ece.ualberta.ca/MoRFPred/	[20]
MFSPSSMPred	MoRF	http://webapp.yama.info.waseda.ac.jp/fang/MoRFs.php	[21]
MoRF _{CHiBi}	MoRF	http://www.chibi.ubc.ca/morf/	[22]
DISOPRED3	変性の有無、結合部位	http://bioinf.cs.ucl.ac.uk/disopred/	[23]

注：Molecular Recognition Features（詳しくは本文 2.2 を参照のこと）

2.1 配列情報に基づく機能予測

天然変性領域の配列がその機能に関する情報を担っているのは明らかである。Vucetic らは変性領域の異なるアミノ酸特性（彼らはフレーバーと呼んだ）と機能を関連付けた[24]。また Lobley らは変性領域の長さや一次構造上の位置情報を使って機能の予測を試みた[25]。この研究は FFPred2.0 において拡張され、同じタンパク質内の構造をもつドメインの特徴をいくつも加えることで、予測の精度向上がなされた[15]。

筆者らは最近、天然変性領域を持つタンパク質の機能が、その天然変性領域の化学的性質と相関づけられることを示した。たとえば、強い正電荷をもつ変性領域は転写因子に存在することが多く、極性残基に富んだ変性領域はプロテインキナーゼなどのシグナル伝達系タンパク質に多いことが観察された[26]。

我々の開発した、変性領域のアミノ酸組成だけを用いて機能予測を行う IDD Navigator というツールは、タンパク質の機能関連用語（ジーンオントロジー（GO）ターム）を統計的に有意に同定することができる[16]。タンパク質の機能の定義はいろいろなレベルでできて、たとえば多くの生物の遺伝子産物は、GO という用語群によって、体系的かつ階層的に記載（注釈付け）されている。ここではアミノ酸組成に基づいて、予測タンパク質と GO による注釈付きのタンパク質集合との距離を計算し、近いものの中で頻出する機能用語を対象タンパク質の予測機能であるとした。その結果、ランダムレベルと比べて有意な予測が可能であった。

さらに、最近の我々の研究では、予測精度を向上させるために、天然変性領域の様々な特徴、例えばアミノ酸組成、予測二次構造や疎水性、さらには進化的保存情報を用いて、単純ベイズ予測器を訓練し、変性領域を含むタンパク質の機能関連用語を予測した[27]。この方法はある程度の成功を収めた。しかしながら、得られた予測結果の精度は機能タイプによってかなりばらついた。すなわち、現段階ではどのような機能タンパク質にも有効な特徴集合を決めるのは難しく、異なる機能分類によって、異なる特徴集合を用いる必要があった。

2.2 結合部位の予測

天然変性領域中の機能的結合部位が予測できればそのタンパク質の機能同定に有用であるという意味で、結合部位予測も機能予測法の一つと捉えることができる。その先駆的なツールの一つが ANCHOR で、変性領域の残基が別のタンパク質と結合するときに折れたたまれるのに必要な能力を予測している[19, 28]。変性領域には、しばしばパートナータンパク質との結合に用いられる短い連続したモチーフ（short linear motif; SLiM）が存在する[29]。それらは通常長さが 3 から 10 残基で、パートナータンパク質との一時的で弱い結合に関与する。また、進化的に保存されることが多く、しばしば翻訳後修飾の標的となる[30]。

そのようなモチーフの例は、p53 の変性領域内にも存在する。p53 は N 末領域と C 末領域にそれぞれ大きな変性領域を持つ。C 末領域の 29 残基はいくつかのタンパク質との結合に関与し、結合時には変性状態から構造状態への転移を伴う（図 1）。この領域が結合に伴って形成する構造は転移を促す結合パートナーによって異なる。すなわち多くの場合はコイルドコイル コンフ

オメーションをとるが、S100B タンパク質と結合するときにはヘリックス コンフォメーションをとる (図 1B)。p53 が様々なターゲットタンパク質と結合するのに用いられるインターフェース部のアミノ酸配列アラインメントから見て取れるように、パートナーとの結合に用いられるのは大きく分けて 3 箇所の残基群であるようである (図 2)。

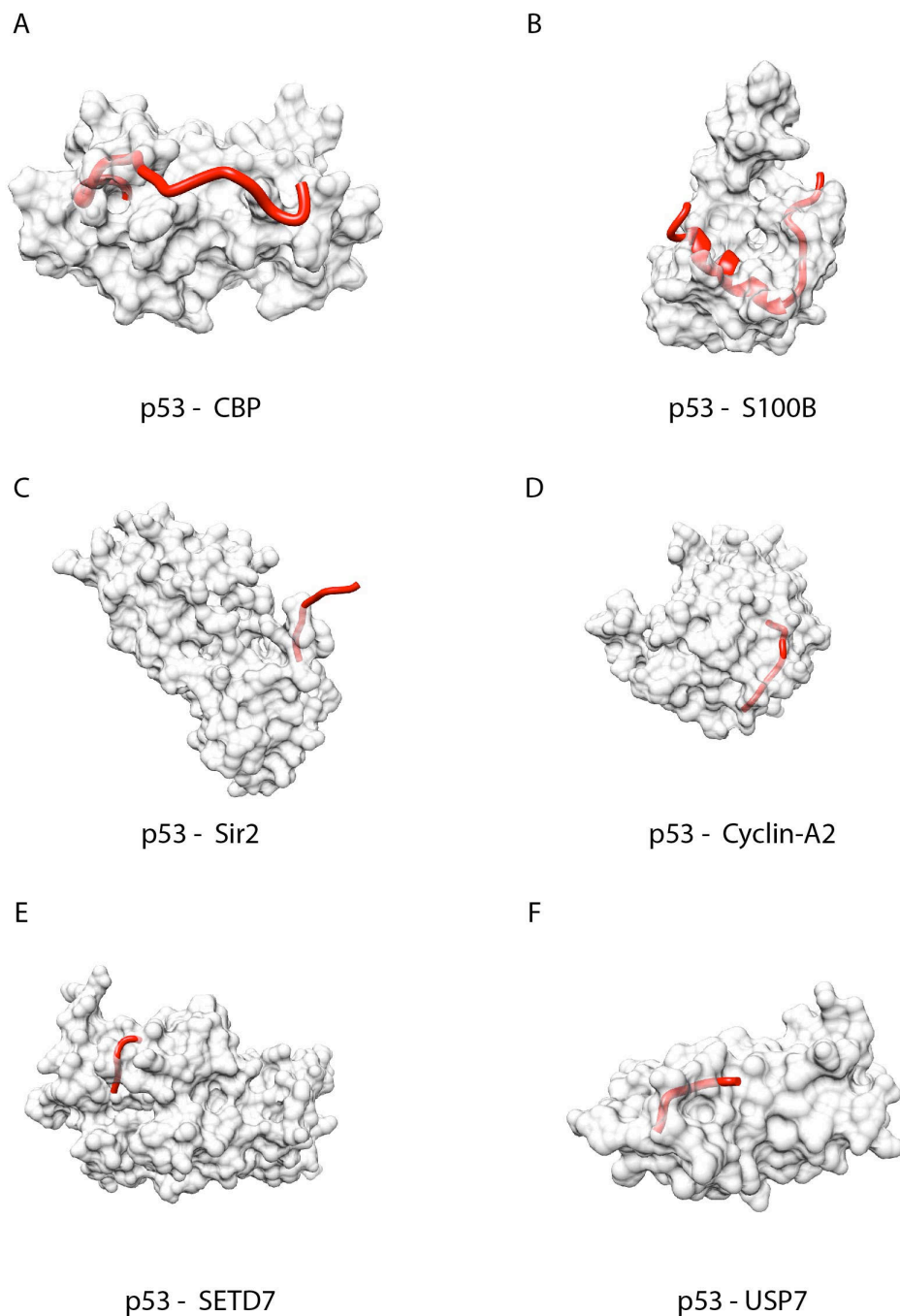


図 1. 様々なタンパク質と結合した p53 の C 末端領域。 p53 は赤いリボンとして示し、結合パートナーは灰色の空間充填モデルで示している。(A) CREB 結合タンパク質 CBP (1JSP: PDB 登録名), (B) S100B タンパク質 (1DT7), (C) NAD 依存デアセチラーゼ Sir2 (1MA3), (D) サイクリン A2 (1H26, CDK2 は非表示), (E) ヒストンリジン-メチルトランスフェラーゼ SETD7 (1XQH), (F) ユビキチン C 末端ヒドロラーゼ 7 USP7 (2FOJ)。各図のスケールは揃えられていない。

Interacting protein	Cellular tumor antigen p53 binding region
	361 GARAHSSHLKSKKGQSTSRHKKLMFKTEG 389
CREB-binding protein, CBP	-----SHLKS ^{KGQ} STSR ^{HK} LM ^{FK} ---
Protein S100B	-----SHLKS ^{KGQ} STSR ^{HK} LM ^{FK} TE-
NAD-dependent deacetylase, Sir2	-----KKGQSTSR ^{HK} LM ^{FK} TEG
Cyclin-A2	-----ST ^{SR} HKKLM ^{FK} ---
Histone-lysine N-methyltransferase, SETD7	-----L ^{KS} SKKGQST-----
Ubiquitin carboxyl-terminal hydrolase 7, USP7	GARAHSS-----

図 2. p53 の C 末端ドメイン領域のアラインメント。それぞれのターゲットタンパク質との結合に係る残基のみを示している。相互作用残基は赤で表示している。

SLiMPrints は、上述の変性領域中のモチーフを配列の保存に基づいて予測するツールである [17]。しかし、それらのモチーフの出現頻度を考えると、それらのすべてに明瞭な保存は期待できず、従ってその予測は容易ではない。そこで、変性領域の配列、予測二次構造、そしてドッキングスコアを組み合わせることで、変性領域中の結合部位に対応するペプチドを予測する PepBindPred というツールも開発されている [18]。

天然変性領域のパートナーとの結合の特徴付けには、短い配列モチーフばかりでなく、分子認識特性 (molecular recognition feature; MoRF) も用いられる。MoRF とは天然変性領域において、ターゲットタンパク質の認識や結合に際して、変性状態から構造状態への転移を受ける配列上連続した領域のことである [31]。MoRFPred は MoRF を予測するために、変性領域の折れたたみ能力を、アミノ酸の特性、物理化学的特性、その他 B 因子や溶媒露出度などの性質を用いている [20]。MFSPSSMpred は、MoRF と周辺の配列の位置特異的スコア行列を使って MoRF を同定する [21]。もう一つ最近発表されたツール MoRF_{Chibi} は、やはり MoRF とその周囲の配列の特異性を利用している [22]。

変性タンパク質の予測ツールである DISOPRED3 は今や変性領域だけでなく、その中のタンパク質結合部位まで予測するが、それには変性領域の進化プロファイルと配列的特徴を用いている [23]。

以上のように、変性領域中の機能部位や、ある程度まではその機能までを予測するための方法やツールがいくつか利用可能である。

3. 考察

現在、利用可能なタンパク質の機能予測法の多くは、機能部位が固定構造ドメインにあるものを対象としている。これは、秩序だった構造ドメインの方が、特徴付けも容易で、アミノ酸配列レベルの進化的保存度が高く、配列や構造の相同性を用いた機能の予測がしやすいことを考えると、当然であろう。逆に言えば、固定した三次元構造を持たず、アミノ酸配列も保存されていない天然変性タンパク質に、これまで開発されてきた機能予測の技術を応用することは困難であるし、そもそもどのようにして機能の特異性を担保しているのかが不思議でもある。

しかしながら、天然変性領域は、これをもつタンパク質の機能とも関連付けられる、独特のアミノ酸組成を持っている。多くの変性領域はまた、機能部位として働く、短く連続したモチーフをもっている。変性領域中で折れたたまれる能力を持っていそうな短い領域を検索することはある程度可能であり、そのような領域は、結合部位やおそらく機能部位の候補となる。これらの特徴はすべて、変性領域やもとのタンパク質の機能を推定することに利用できる。ただし、天然変性領域の配列的特徴と機能の関連性に関する研究は、有望ではあるが、機能の詳細を予測できるレベルにはない。同様に、今や変性領域が一時的もしくは安定的にターゲットタンパク質と結合する候補部位を同定することができるが、その部位に結合する相手のタンパク質 (群) を予測したり、結合によってどのような機能が生じるのかを予測したりすることは難しい。結合相手の

予測には、大規模タンパク質間相互作用実験の結果を併用することが有用であろう。また、機能既知の天然変性領域内の機能部位を記載する努力がなされており、そのようなデータが十分に蓄積すれば、既知部位との類似性を頼りに予測を行うのに役立つことであろう。今後、総合的な機能予測能を高めていくためには、変性ドメインのまわりの通常構造領域の情報（当然、得られる情報量が多い）を取り入れていくことも有望と思われる。とにかく、予測能力向上には、いろいろな特徴を組み合わせていく必要があるだろう。

4. 結語

天然変性タンパク質は細胞内で重要な機能を果たしている。従って、その機能を同定するために、そのアミノ酸配列から情報を抽出する研究は重要である。この研究分野の最近の進歩は著しいが、現在の実験的、情報科学的方法論がまだまだ十分な解析を行うには未熟なため、いくつかの課題が残されている。天然変性タンパク質に関する新しい予測法を開発したり、実験情報を収集したりする研究には、情報研究者と実験研究者のより一層の協力と参画が求められている。その意味で、本稿が構造活性相関を研究される皆様にとって、この分野に少しでも興味をもっていただく一助となれば幸いである。

謝辞

本稿執筆の機会をくださった日本薬学会構造活性相関部会に謝意を表したい。

参考文献

- [1] Rose PW, Prlić A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, et al: The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* 2015, **43**:D345-356.
- [2] Wright PE, Dyson HJ: Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 1999, **293**:321-331.
- [3] Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT: Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004, **337**:635-645.
- [4] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN: Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS Journal* 2005, **272**:5129-5148.
- [5] Patil A, Nakamura H: Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Letters* 2006, **580**:2041-2045.
- [6] Dyson HJ, Wright PE: Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 2002, **12**:54-60.
- [7] Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD: Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 2010, **18**:494-506.
- [8] Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK: Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 2002, **323**:573-584.
- [9] Fong JH, Shoemaker BA, Panchenko AR: Intrinsic protein disorder in human pathways. *Mol Biosyst* 2012, **8**:320-326.
- [10] Wright PE, Dyson HJ: Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 2015, **16**:18-29.
- [11] Malaney P, Pathak RR, Xue B, Uversky VN, Davé V: Intrinsic disorder in PTEN and its interactome confers structural plasticity and functional versatility. *Sci Rep* 2013, **3**:2035.
- [12] Patil A, Kinoshita K, Nakamura H: Hub promiscuity in protein-protein interaction networks. *International Journal of Molecular Sciences* 2010, **11**:1930-1943.
- [13] Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor

- K, Ben-Hur A, et al: A large-scale evaluation of computational protein function prediction. *Nat Methods* 2013, **10**:221-227.
- [14] Bracken C, Iakoucheva LM, Romero PR, Dunker AK: Combining prediction, computation and experiment for the characterization of protein disorder. *Curr Opin Struct Biol* 2004, **14**:570-576.
- [15] Minneci F, Piovesan D, Cozzetto D, Jones DT: FFPred 2.0: improved homology-independent prediction of gene ontology terms for eukaryotic protein sequences. *PLoS One* 2013, **8**:e63754.
- [16] Patil A, Teraguchi S, Dinh H, Nakai K, Standley DM: Functional annotation of intrinsically disordered domains by their amino acid content using IDD navigator. *Pac Symp Biocomput* 2012:164-175.
- [17] Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ, Edwards RJ: SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res* 2012, **40**:10628-10641.
- [18] Khan W, Duffy F, Pollastri G, Shields DC, Mooney C: Predicting binding within disordered protein regions to structurally characterised peptide-binding domains. *PLoS One* 2013, **8**:e72838.
- [19] Dosztányi Z, Mészáros B, Simon I: ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 2009, **25**:2745-2746.
- [20] Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L: MoRFPred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 2012, **28**:i75-83.
- [21] Fang C, Noguchi T, Tominaga D, Yamana H: MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. *BMC Bioinformatics* 2013, **14**:300.
- [22] Malhis N, Gsponer J: Computational identification of MoRFs in protein sequences. *Bioinformatics* Epub 2015 Jan 30. (doi: 10.1093/bioinformatics/btv060)
- [23] Jones DT, Cozzetto D: DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 2015, **31**:857-863.
- [24] Vucetic S, Brown CJ, Dunker AK, Obradovic Z: Flavors of protein disorder. *Proteins* 2003, **52**:573-584.
- [25] Lobley A, Swindells MB, Orengo CA, Jones DT: Inferring function using patterns of native disorder in proteins. *PLoS Comput Biol* 2007, **3**:e162.
- [26] Moesa HA, Wakabayashi S, Nakai K, Patil A: Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Mol Biosyst* 2012, **8**:3262-3273.
- [27] Sharma A, Dehzangi A, Lyons J, Imoto S, Miyano S, Nakai K, Patil A: Evaluation of sequence features from intrinsically disordered regions for the estimation of protein function. *PLoS One* 2014, **9**:e89890.
- [28] Mészáros B, Simon I, Dosztányi Z: Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 2009, **5**:e1000376.
- [29] Nguyen Ba AN, Yeh BJ, van Dyk D, Davidson AR, Andrews BJ, Weiss EL, Moses AM: Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci Signal* 2012, **5**:rs1.
- [30] Tompa P, Davey NE, Gibson TJ, Babu MM: a million peptide motifs for the molecular biologist. *Molecular Cell* 2014, **55**:161-169.
- [31] Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN: Analysis of molecular recognition features (MoRFs). *J Mol Biol* 2006, **362**:1043-1059.

//// Cutting Edge ////

PoSSuM: ポケット類似性情報に基づく合理的薬剤設計支援に向けて

株式会社レベルファイブ 池田 和由
産業技術総合研究所 富井 健太郎

1. はじめに

タンパク質の立体構造情報の近年の飛躍的な増加に伴い、基質結合部位構造の比較が、生物学的等価体の発見や既存医薬品の再利用（ドラッグリポジショニング）、副作用（オフ・ターゲット）予測などに結びつく例が報告されてきている[1]。こうしたことから、タンパク質立体構造データベース（PDB）中に存在する大量の既知および未知基質結合部位構造を網羅的に比較・分類することは、タンパク質と薬剤の結合に関する重要な特徴発見を一層促すものと期待される。しかし大量の結合部位を正攻法で網羅的に比較することは、部位間の類似度計算に多大な時間を要するため、これまで実現困難であった。そこでわれわれは、結合部位の構造情報の大胆な粗視化と高速なソートアルゴリズムを組み合わせることで、100 万のオーダーの結合部位の網羅的比較を可能にし、その結果をまとめたデータベース PoSSuM (Pocket Similarity Search using Multi-Sketches) を開発・公開した[2, 3]。2011 年にリリースした PoSSuM は、PDB のエントリー増大などに伴い、現在では 550 万の既知および未知結合部位の網羅的比較から、4,900 万の類似結合部位ペアを収載したデータベースとなっている[4]。最新の更新では、ドラッグリポジショニング、副作用予測などを意識した新たなデータベース PoSSuMds をリリースしたので、これらデータベースの概要について紹介する。

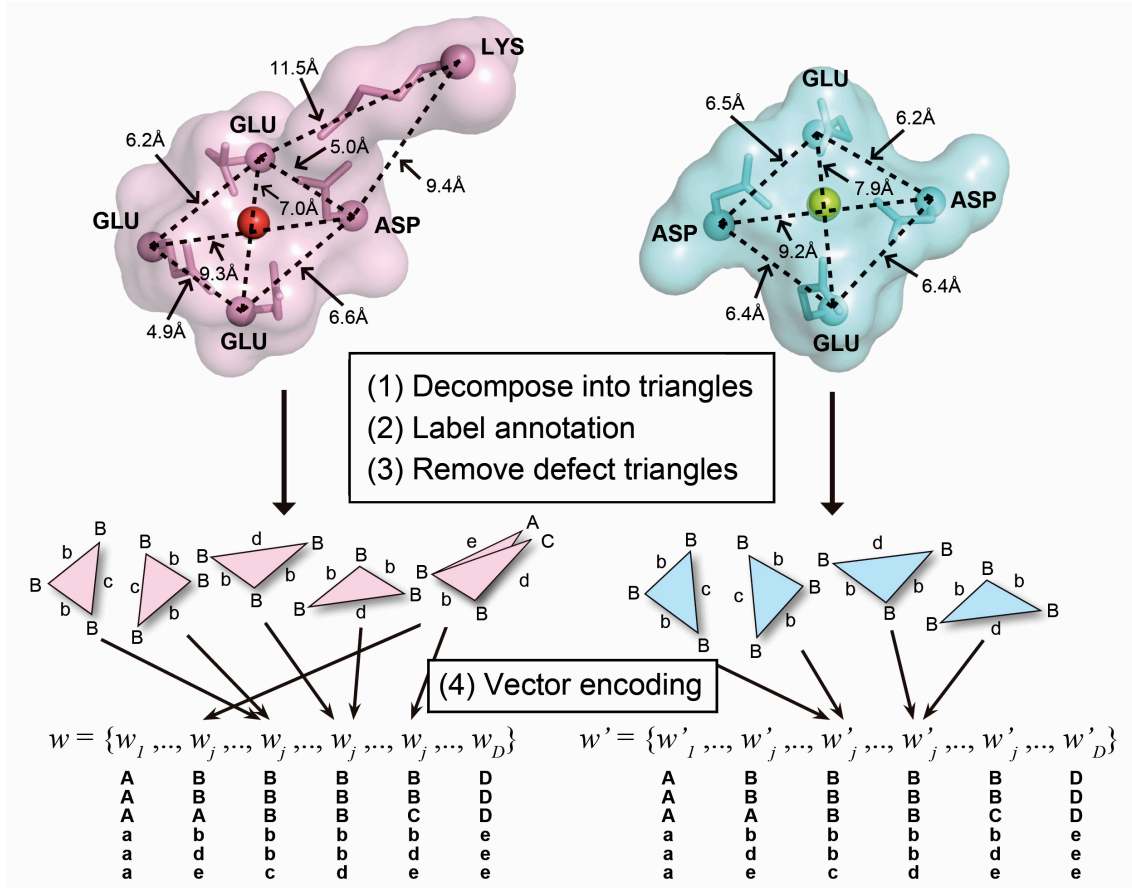
2. 網羅的比較法と PoSSuM/PoSSuMds データベース

2.1 最近傍探索アルゴリズムを適用した全ペア類似度検索

先に触れたように、例えば通常良く用いられる構造の重ね合わせ法などによる大量の結合部位の網羅的比較は、部位間の類似度計算に多大な時間を要するため一般に困難である。この問題の克服にあたり、われわれは「最近傍探索アルゴリズム」を適用した。比較あるいは類似性検索を行いたい場合、対象の近傍データは重要であるが、対象から非常に離れた（非類似）データはそもそも比較を行う意味があまりない。そこで、高次元空間上の局所的な（あるいは近傍の）性質が良く保存されるような低次元への射影である局所性鋭敏型ハッシュ（locality sensitive hashing）を利用した。手順は以下の通りである。まず、SiteAlign を開発した Rognan らのグループの先行研究[5]に倣い、基質結合部位をその構成残基（ここでは C_{α} 原子で代表）を頂点とする三角形に分割し、各残基が備えている様々な性質（基質との結合に重要と考えられる物理化学的性質や二次構造情報など）と C_{α} 原子間距離に基づき 1,540 種類に分類された三角形の出現頻度ベクトルに変換する（図 1）。Rognan らのグループは、これらのベクトルのなす角、即ちコサイン距離を結合部位間の類似度としたが、この総当たりの計算は一般に多大な時間を要することは既に述べたとおりである。そこで、われわれは更なる近似表現を用いた。つまり、この 1,540 次元空間上で個々の基質結合部位に対応する各点を、高次元空間での局所的な性質が良く保存されるような低次元への射影である局所性鋭敏型ハッシュを用い（具体的には高次元空間で任意の超平面のどちら側に位置しているかにより 0 か 1 の符号を割り当てるランダム射影を繰り返し）、32 次元のビットベクトル（=2 値の数値列）に変換する。直感的に述べると、類似した結合部位、つまり高次元空間で近傍にある二点は、類似の数値列として表現される。低次元での近似表現の利用によって高速なソートアルゴリズムの一つである複合ソート法[6]の利用が可能となり、生成された二値の数値列の限定されたハミング距離での近傍探索が行われ、性質の良く似た基質結合部位ペアの高速かつ効率的な検出が可能となる。これにより、PDB に登録されている既知の基

質結合部位のみならず、タンパク質の立体構造情報から計算[7]される潜在的な基質結合部位を加えた 100 万のオーダーの基質結合部位の網羅的比較が実時間で可能となる。

(A)



(B)

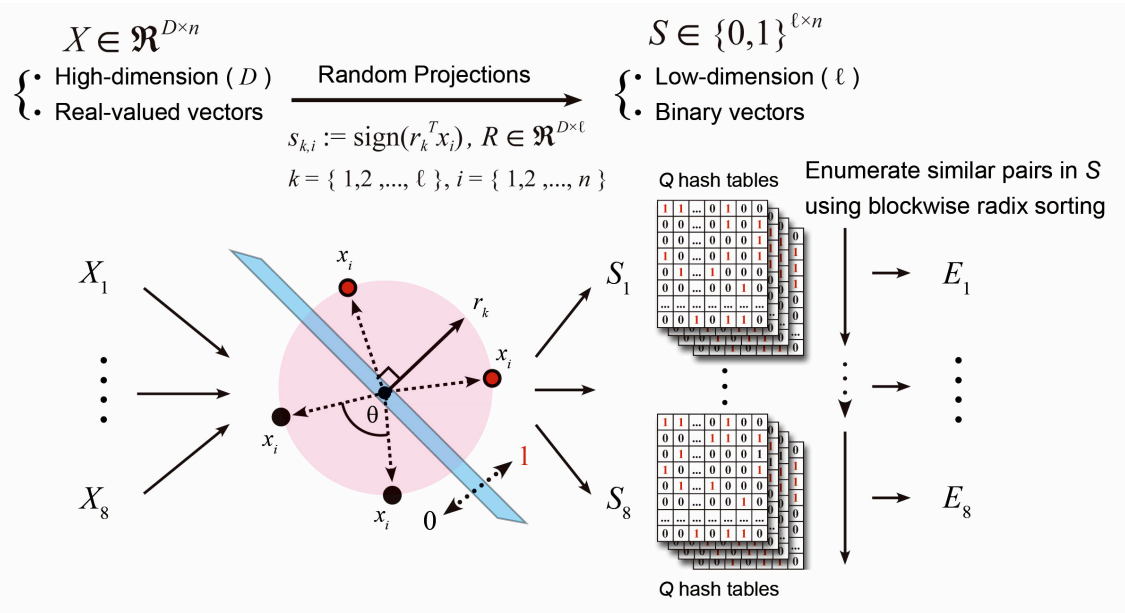


図 1. PoSSuM データベース構築に用いた基質結合部位構造情報の粗視化の概念図
(前ページ)

- A) 構成残基の C_{α} 原子を頂点とする三角形への分割と、各残基の性質および C_{α} 原子間距離に基づき分類された三角形の出現頻度ベクトルへの変換。各頂点のラベルは、基質との結合に重要と考えられる（荷電残基であるか否か、芳香族環をもつか否かなどの）アミノ酸の物理化学的性質や残基の二次構造情報などに応じて A~D の 4 種類が割り当てられ、 C_{α} 原子間距離のラベルは、一辺 13.6Å までの距離が（最少の距離区分を除き）2.2Å 刻みで a~e の 5 種類に分類される。例えば、三頂点が (A, A, A) のラベルをもち、三辺が (a, a, a) の距離区分にある三角形が一つ観測される時、 $w_1=1$ となる。こうした手順に従い、任意の基質結合部位は $D(=1,540)$ 次元のベクトルとして表される。文献[2]の図 2 を改変。
- B) 局所性鋭敏型ハッシュを用いたビットベクトルへの変換。1,540 次元空間上で個々の基質結合部位に対応する各点 x_i ($i = 1, 2, \dots, n$; ここで n は部位数) を、ランダム射影により ℓ ($=32$) 次元のビットベクトルに変換する。現在の PoSSuM では、アミノ酸の多様な性質のうちこういった性質に着目するかに応じて、8 種類の異なる三角形の分類法 ($X_1, X_2, \dots, X_8$) を用いている。文献[2]の図 6 を改変。

2.2 PoSSuM の概要

現在 PoSSuM (<http://possum.cbrc.jp/PoSSuM/>) には、PDB から抽出された 30 万余の低分子化合物の結合部位とタンパク質の立体構造情報からポケット検出プログラム Ghecom[7]により計算された 520 万超の基質結合部位候補の合計 550 万部位の網羅的比較から検索された、約 4,900 万の類似結合部位ペアのデータが収載されている。前節で紹介した方法で計算された網羅的比較の結果が SQL データベースとしてまとめられており、“SearchK”と称するモードでは、既知結合部位情報（PDB ID と 3 文字表記の基質名）を入力として、その部位と閾値以上の類似度をもつ既知および推定された結合部位のリストを取得可能である。このリストには、タンパク質の立体構造分類データベースである SCOP や CATH、EC 番号および GO タームに関する情報が付加されており、部位ペアの属性の異同確認が可能である。また、部位相互の実際の構造の重ね合わせも閲覧可能である。これとは逆に、未知結合部位の類似部位を知りたい場合、“SearchP”というモードで、対象タンパク質の PDB ID を入力として“SearchK”の場合と同様の情報を取得可能である。文献[2]では、こうしたデータに含まれる未知結合部位と既知結合部位の類似性に基づいた、構造ゲノムプロジェクトなどで決定された機能未知タンパク質に対する基質および結合モードの予測例が示されている。なお“SearchP”では、知りたいタンパク質の立体構造が PDB にな

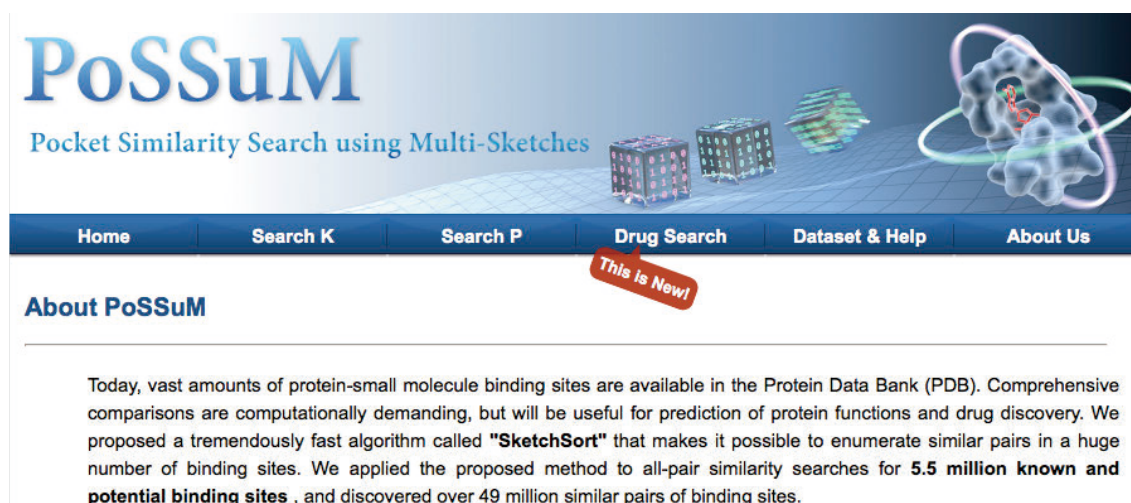


図 2. PoSSuM のトップページ (<http://possum.cbrc.jp/PoSSuM/>)

い場合でも、PDB フォーマットの構造情報を入力すると、その情報を基に Ghecom を用いて計算されたポケットについて、同様の情報を取得可能である。

2.3 PoSSuMds: 低分子薬への注目とケミカルゲノムデータへのリンクの意義

前節で紹介した結果を踏まえ、また、冒頭に述べたようなトレンド[1]から、われわれは、ドラッグリポジショニングやオフ・ターゲット予測などへの応用を意識し、PDB の低分子医薬品結合部位に焦点をあてたデータベース PoSSuMds (http://possum.cbrc.jp/PoSSuM/drug_search/) の開発と公開を行った[4]。低分子医薬品は、医薬品全体の売り上げ中の半分以上を占めている重要な医薬品分子であり、一般的に分子量が小さく（1970 年代以降の低分子薬の平均分子量は 391）、膜透過性が高いものも多く、生体内の動態性質も良く、患者負担が低い経口投与にし易い性質がある。近年、ChEMBL[8]、PubChem[9]といったオープンケミカルゲノム情報の増加は目覚ましい。ChEMBL の最新版（2015 年 2 月リリース version 20）では、トータルの生物活性値数が 1,352 万件、アッセイ数が 114 万件、このうちタンパク質をターゲットとするものが全体の 56% を占める。化合物 1 分子あたりのターゲット数はおおよそ 4 個であり、複数のターゲットに結合活性をもつものも多く、5 個以上のターゲットに結合する化合物は全体の 27% 存在する。

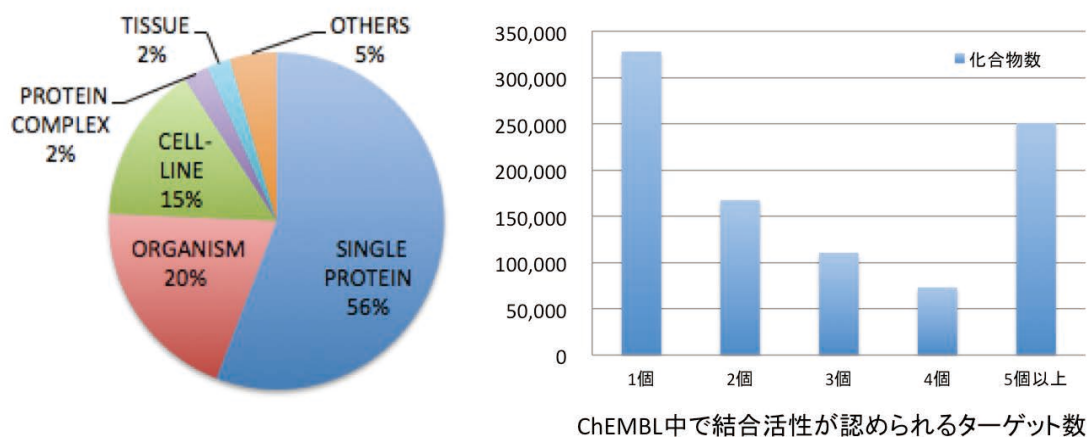


図 3. ChEMBL 中のターゲットに存在するタンパク質の割合（左図）と複数のターゲットに結合する化合物数（右図）

ChEMBL 中のタンパク質と基質の結合相互作用を観察すると、進化上あるいは全体構造をみても関係の推定できないタンパク質に、同一（あるいは非常に類似した）化合物が結合している例が多くあることに気づく。これらの低分子活性情報は、タンパク質の基質結合部位の情報と関連付けることによって、タンパク質機能とのより直接的な議論が可能になり、低分子が結合可能なターゲットの発見やドラッグリポジショニングへの応用が期待される。ドラッグリポジショニングでは、すでに安全性、動態物性が臨床段階で検証済みであり、過去の医薬品開発情報が豊富に蓄積されている低分子経口医薬品がその中心である。

今回われわれは、PoSSuM データベースに登録されている結合低分子（既知基質）についての [A]ChEMBL の化合物情報へのリンク、および [B] ChEMBL のアッセイデータへのリンクを行った[4]。手順は、以下の通りである。[A]において、まず PoSSuM から既知基質を抽出し、その基質構造と ChEMBL 中の化合物とのマッチングを InChIKey を用いて行った。ChEMBL では、化合物構造の登録 Standard InChI (<http://www.iupac.org/home/publications/e-resources/inchi.html>) を用いているため、そのハッシュである InChIKey を使用することで高速な構造検索が可能である（現在は UniChem (<https://www.ebi.ac.uk/unichem/>) を使うとより容易であることを付言しておく）。2014 年 8 月現在、PoSSuM 中の 30 万余の既知基質結合部位のうち、低分子経口医薬品の結

合部位は、0.86%の2,595個であり、ユニークな低分子薬数は194個であった。また、PoSSuM中でこれらの結合部位と類似するポケットは26,509個検出された。

複数のターゲットに結合活性を有する低分子薬の存在や、同一化合物の多様なタンパク質との結合例については既に述べた通りである。こうした現象を理解するために、また、低分子薬が、その結合部位に類似した他の結合部位にも親和性を示す可能性があることを支持する一つのエビデンスとなりえると考え、ChEMBLにおいてクエリの基質とターゲットのタンパク質との生物活性情報の有無を調べた。ここで、[B]におけるPoSSuMデータからChEMBLのアッセイ情報へのリンクは、PoSSuM中のタンパク質とChEMBLのターゲット（単一タンパク質）をUniprot IDでマッチングを行った。なお、今回われわれは、医薬基盤研が開発した統合データウェアハウスTargetMine (<http://targetmine.nibio.go.jp>) [10]を使用した。この結果、PoSSuM中に存在する上記26,509個の既知低分子経口薬の類似結合サイトで、クエリ化合物と生物活性があるものが存在することがわかった。次節にその例を示す。

3. 結合部位の類似性と ChEMBL アッセイデータの対応例

ここでは文献[4]でとりあげた例について紹介する。慢性骨髄性白血病などの治療薬であるイマチニブ（商品名：グリベック）はチロシンキナーゼ阻害薬であり、イマチニブとチロシンキナーゼの共結晶構造も決定されている。PoSSuMでは、BCR-ABL1チロシンキナーゼのイマチニブ結合部位とtype IIホスファチジルイノシトール4-リン酸5-キナーゼ（PIP5K）のポケットの類似性が検出された（図4）。チロシンキナーゼはタンパク質リン酸化酵素（EC#: 2.7.10.2）であり、PIP5Kはリン脂質リン酸化酵素（EC#: 2.7.1.149）である。両者は、酵素反応が異なり、また異なるドメイン構造（PfamではPF07714とPF01504）およびフォールド（CATHではPhosphorylase Kinase; domain 1 & Transferase (Phosphotransferase); domain 1とPhosphatidylinositol Phosphate Kinase II Beta）を持つが、PoSSuMでは類似性を検出した。実際に、2つの結合部位のRMSDは2.71Åであり、相互作用する位置のアミノ酸残基（K271とK152、286Eと158D）も

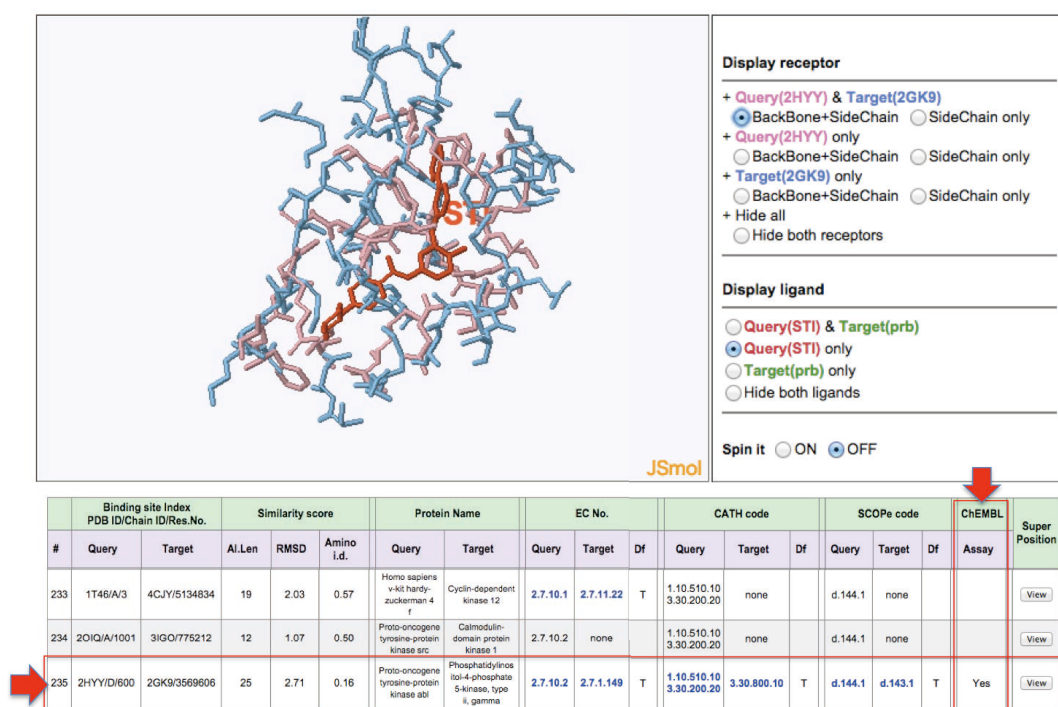


図 4. PoSSuM 中で低分子薬（STI）と類似検出されたポケット
ChEMBL に生物活性情報が存在する（Assay=Yes、下図）

保存されており、これは PoSSuM が、ポケットの形状のみならず、残基の物理化学的特徴などを考慮して類似性を検出できていることを示している。PIP5K (PDB ID: 2GK9) の結晶構造には基質は存在しないが、前節で述べた ChEMBL のアッセイ情報 (イマチニブと PIP5K の複合体解離定数 K_d 値 380nM) が存在し、PoSSuM で検出されたイマチニブ結合部位と PIP5K のポケットの類似性が実際の活性予測に有益であることを示唆するものであった。

4. 結語

近年、ドラッグリポジショニングやオフ・ターゲット予測などの創薬研究において、基質結合部位構造の比較に基づく手法の存在感が、構造情報の蓄積とも相まって、増してきている。しかしながら、われわれの手法も含め、スクリーニング効率にはまだ改善の余地があると考えられる。われわれの手法の特長の一つである、基質結合部位が備えている様々な性質 (幾何学的性質、アミノ酸の属性、二次構造など) の特徴抽出法を修正することで、計算機によるスクリーニング効率の更なる改善が可能かどうか検証を進めたいと考えている。一方、アッセイやそのターゲットのアノテーションに関しても一層の工夫が必要である。現状 ChEMBL の活性結合情報は、タンパク質レベルでの基質相互作用が主である。これらをドメインレベル [11] (さらに結合部位レベル) でのアノテーションや活性値との関連付けすることで、類似結合部位の基質情報からより高活性をもつ化合物構造との比較も可能になる。また PoSSuM に収載されているデータは、既知および未知結合部位の網羅的比較から得られるものであることから、ポリファーマコロジーの視点とも合致するため [12]、この方面の利用可能性も検証したいと考えている。今後益々タンパク質の立体構造データの増加が見込まれる状況を鑑みても、計算機を用いた結合部位の網羅的比較は、創薬研究においてもより一層その重要性を増すものと考えられる。

謝辞

本研究推進の要となった伊東純一博士、ご協力頂いた田部井靖生博士、清水佳奈主任研究員、東北大学山田和範先生、東京大学津田宏治先生、医薬基盤研究所水口賢司先生および関係の皆様感謝申し上げます。本研究の一部は、JSPS 科研費、最先端研究開発支援プログラム、厚生労働省科研費および文部科学省創薬等支援技術基盤プラットフォームの助成を受けたものです。

参考文献

- [1] Konc, J. and Janežič, D. Binding site comparison for function prediction and pharmaceutical discovery, *Curr. Opin. Struct. Biol.*, **25**, 34-39 (2014).
- [2] Ito, J. *et al.* PDB-scale analysis of known and putative ligand-binding sites with structural sketches, *Proteins*, **80**, 747-763 (2012).
- [3] Ito, J. *et al.* PoSSuM: a database of similar protein-ligand binding and putative pockets, *Nucleic Acids Res.*, **40**, D541-D548 (2012).
- [4] Ito, J. *et al.* PoSSuM v.2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs, *Nucleic Acids Res.*, **43**, D392-D398 (2015).
- [5] Weill, N. and Rognan, D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites, *J. Chem. Inf. Model.*, **50**, 123-135 (2010).
- [6] Tabei, Y. *et al.* Single versus multiple sorting for all pairs similarity search, In Sugiyama, M. and Yang, Q. ed. *Proceedings of 2nd Asian Conference on Machine Learning (ACML2010)*, *JMLR Workshop and Conference Proceedings*, **13**, 145-160 (2010).
- [7] Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins*, **78**, 1195-1211 (2010).
- [8] Bento, A. P. *et al.* The ChEMBL bioactivity database: an update, *Nucleic Acids Res.*, **42**, D1083-D1090 (2014).
- [9] Wang, Y. *et al.* PubChem BioAssay: 2014 update. *Nucleic Acids Res.*, **42**, D1075-D1082 (2014).
- [10] Chen, Y.-A. *et al.* TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery, *PLoS ONE*, **6**, e17844, (2011).

- [11] Kruger, F. A. *et al.* PPDMS—a resource for mapping small molecule bioactivities from ChEMBL to Pfam-A protein domains. *Bioinformatics*, **31**, 776-778 (2015).
- [12] Duran-Frigola, M. *et al.* Structural systems pharmacology: the role of 3D structures in next-generation drug development, *Chem. Biol.*, **20**, 674-684 (2013).

//// SAR Presentation Award ////

< SAR Presentation Award について >

「SAR Presentation Award」は、構造活性相関シンポジウムにおける若手研究者の発表を奨励し、構造活性相関研究の発展を促進するため、2010 年度に創設された。当初は応募制として審査対象講演の募集を行った。2012 年度からは、正式名称を「構造活性相関シンポジウム優秀発表賞」（英語表記 SAR Presentation Award）と定め、構造活性相関シンポジウムにおける 45 歳以下の発表者（登壇者）による全ての一般講演（口頭発表）を選考対象とすることとした。

< 2014 年度選考結果について >

第 42 回構造活性相関シンポジウム「優秀発表賞」受賞者の選考については、審査員から提出頂いた審査票の集計結果をもとに、去る 1 月 7 日（水）に薬学会館（渋谷区）で開催された平成 26 年度第 4 回常任幹事会において、別途定めた評価基準にもとづいて協議が行われ、選考対象講演のうち、評点が上位の 2 名を受賞候補者として選出した。

2014 年度構造活性相関シンポジウム優秀発表賞

寺師 玄記（北里大学 薬学部）

長門石 暁（東京大学大学院 工学系研究科）

受賞者の発表は本ニュースレターにおける誌上発表とし、受賞者には別途部会長名で授賞通知を送付するとともに受賞の諾否を確認し、表彰状と副賞（図書券）を贈呈した。なお、審査にあたっての観点と評価基準を以下に付す。

審査の観点と評価基準：

（：大変優れている 4:優れている 3:良い 2:改善が必要である 1:大いに改善が必要である）

- a) 講演要旨：講演要旨は発表内容を反映して適切に作成されているか。
- b) 講演資料：スライドは専門領域の異なる参加者にも分りやすく、見易く、かつ発表時間に見合って適切に作成されているか。
- c) 研究のねらい：研究の背景と目的、先行研究との関係、研究の新規性あるいは有用性が明確になっているか。
- d) 論理構成の合理性：研究方法が適切であるか、適切な文献資料、データに基づいて議論が進められているか。考察・結論は妥当か。
- e) 質疑応答：質問等に対し、的確な応答・議論がなされたか。活発な討論がなされたか。
- f) 総合評価：a)~e)を総合的に評価する。

2014 年度審査員（50 音順、敬称略）：

赤松 美紀、飯島 洋、伊藤 篤史、岡島 伸之、岡田 孝、小田切 優樹、粕谷 敦、狩野 敦、川下 理日人、合田 浩明、高橋 由雅、竹田一志鷹 真由子、田上 宇乃、谷村 隆次、田宮 実、中川 好秋、服部 一成、永田 尚也、広川 貴次、広野 修一、藤原 英明、本間 光貴、前田 美紀、丸岡 博、宮本 秀一、横山 祐作

< 受賞者コメント >

KO09

氏名 寺師 玄記（てらし げんき）
 所属 北里大学薬学部
 演題 残基間相互作用に基づく新たなタンパク質構造アライメント法の開発

このたびは、構造活性相関シンポジウム優秀発表賞という評価をいただき、審査にあたられた先生方ならびに日本薬学会構造活性相関部会の先生方に厚く御礼申し上げます。また、本研究のご指導を賜りました竹田一志鷹真由子教授に心より御礼申し上げます。

本研究で発表した手法は、タンパク質構造間の隠れた類縁性を検知し高い精度の構造アライメントを計算する手法であります。従来の三次元空間座標を用いたタンパク質構造アライメント法や、アミノ酸配列検索では見出だせない類縁性を見つける事が可能となりました。また、本研究の手法は大きなタンパク質構造の変化に影響を受けない新しい構造アライメント法であります。タンパク質構造アライメント法は、Homology modeling 法やタンパク質構造分類、Protein-protein docking などに幅広く適用されております。本研究の手法が、これらの研究を一段階引き上げるものと考えております。今回頂いた賞を励みにして、今後も一層の研究に励んでいきたいと考えております。この度は本当にありがとうございました。

本研究は、本研究は日本学術振興会科学研究費補助金・若手研究 B の助成を受けて行われました。この場を借りてお礼申し上げます。

KO12

氏名 長門石 暁（ながといし さとる）
 所属 東京大学大学院 工学系研究科
 演題 *Staphylococcus aureus* 莢膜合成酵素 CapF と低分子化合物の物理化学的な相互作用解析

この度は、第 42 回構造活性相関シンポジウム優秀発表賞という大変名誉ある賞をいただき、まことにありがとうございました。身に余る光栄と心から感謝申し上げます。

本研究は、薬剤設計のための構造活性相関 (SAR) として、“等温滴定型カロリメーター (ITC)” という測定技術を駆使し、熱量変化から熱力学的パラメータを得て、蛋白質と低分子化合物間の相互作用に関する定量的な議論を試みております。本アプローチは、*in silico* 解析でないことはもちろんのこと、複合体の構造情報もない条件での SAR に挑んでおります。にもかかわらず、審査員の先生方に高く評価いただきましたことは、*in vitro* での物理化学的解析を SAR に活用していくことへの大きな期待と感じております。日本薬学会構造活性相関部会の先生方に厚く御礼申し上げます。熱量評価に関しましては、今後、創薬分野においてより実用的なレベルへと高めて参りたいと考えております。これからも私が所属しております津本研究室共々、何卒よろしくお願い申し上げます。

最後に、本研究を進める上で御指導を賜りました津本浩平教授と Jose Manuel Martinez Caaveiro 主幹研究員に改めて感謝申し上げます。

< 受賞講演要旨 >

次頁以降に、優秀発表賞受賞講演の講演要旨を掲載します。

KO09

残基間相互作用に基づく新たなタンパク質構造アライメント法の開発

(北里大薬¹)

○寺師玄記¹、竹田一志鷹真由子¹

1. 序論

遺伝情報の伝播の過程において、タンパク質の構造と機能はアミノ酸配列上の保存性よりも強く保存・伝播することが明らかとなっている。そのためタンパク質の構造比較の研究はアミノ酸配列上の類似性が見られないタンパク質間の類縁度を研究するのに必須の手法である。他にもホモロジーモデリング法や、リガンド結合部位の解析に「タンパク質を重ねあわせる」構造比較が行われる。タンパク質の構造比較に用いられている構造アライメントアルゴリズムのほとんどは、より多くの対応したタンパク質原子の座標が三次元座標上で重ね合わされるようにするものである。すなわち、Coverage が大きく、構造上のずれ Deviation が小さい構造アライメントを探索する手法である。

一方、タンパク質は相互作用相手や環境下に応じて動的に構造を変化させることが知られている。タンパク質構造の変化には、相互作用部位周辺のループが動くものから、ドメイン構造の相対配置が大きく変化するものまで様々である。機能や遺伝学的類縁関係が明らかとなっているタンパク質間で構造変化がある場合、従来の構造アライメント法は正確なアライメントを計算できず構造類似性を検出することができない。現在主流の構造アライメント法の精度は、タンパク質の構造変化に大きく影響を受けている。

我々はこの問題を解決するにあたり、三次元座標の比較では検出が不可能なタンパク質構造の「構造類似」に着目した。すなわち、三次元座標上は構造が類似していないが、アミノ酸残基間のネットワークは類似しているならば、類縁関係が存在するという考え方である。本研究で我々は、アミノ酸残基間のネットワークの類似性に着目し、タンパク質の動的構造変化に影響されない新しい構造類似指標とそれに基づく構造アライメント法の開発を行った。

2. 方法

2.1 残基間相互作用に基づくタンパク質構造類似性

タンパク質の立体構造をアミノ酸残基間の相互作用のネットワークとして記述するため、まず各アミノ酸残基の形状をポリゴンに変換した(図1)。ポリゴンの面(surface)のデータから、各残基の相互作用面積を計算した。

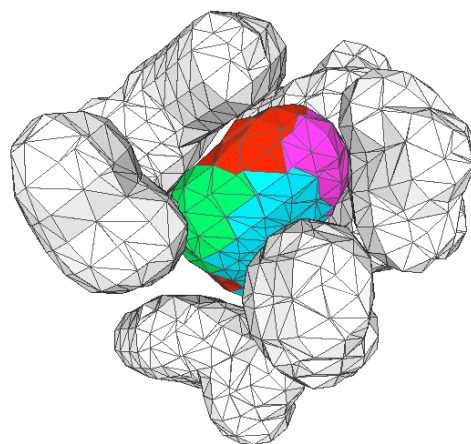


図1 ポリゴン表現によるアミノ酸残基の相互作用解析

本研究では、ここで計算された相互作用面積を残基間のネットワークの強さとし、より多くのネットワークが2つのタンパク質分子間で類似しているほど、類似したタンパク質であると定義した。構造アライメントから計算される構造類似度 S は次の式(1)で計算される。

$$S = \sum_i^L \sum_j^L \theta(i, j) \quad (1)$$

L は構造アライメントの長さ、 i, j はそれぞれアライメント中同じ位置にアライメントされた残基ペアを意味している。したがって、タンパク質 A と B の構造アライメントにおいて、 $i = (i_A, i_B)$ 、 $j = (j_A, j_B)$ である。 i は A の残基 i_A 番目と B の残

基 i_B 番目がアライメントされていることを意味している。 θ は、アライメント位置 i, j 間のネットワークの類似度関数である。数式 (1) の意味するところは、「構造アライメント中、より多くの類似したネットワークが存在するならば、2つのタンパク質構造は類似している。」である。本研究では、式 (1) の値が最も高くなるように Iterative Dynamic Programming 法を実行しアライメント計算を行った。

2.2 構造類似度の標準化

式 (1) の構造類似度は、計算するタンパク質構造の大きさに依存するため、標準化を行った。

$$S' = \frac{S}{(T_A)^{pow}} + \frac{S}{(T_B)^{pow}} \quad (2)$$

S は式(1)のスコアであり、 T_A と T_B は、タンパク質 A 、 B 内に存在する残基間相互作用面積の総和である。パラメーター pow を 0~1 の間に設定すると、 S' は次の2つの性質を示す：① S の値が高いほど、 S' の値が高くなる。② $\frac{S}{T_A}$, $\frac{S}{T_B}$ の値が高いほど S' の値が高くなる。

この2つの性質は、タンパク質間の類縁性を評価する上で有用である。より多くの残基間ネットワークが類似しているならば、強い類縁関係が示唆されるため、たとえ、高い $\frac{S}{T_A}$, $\frac{S}{T_B}$ を有する構造アライメントであっても、小さな分子同士の S' は低く補正される。また、ほぼ同じ S を持つ2つの構造アライメントを比較したとき、より高い $\frac{S}{T_A}$, $\frac{S}{T_B}$ をもつ構造アライメントの方で、より強い類縁関係が示唆される。

3.結果および考察

前述したアルゴリズムを C 言語で実装し、コンピュータプログラム CA-align を開発した。ポリゴン化から構造アライメント計算までのすべての処理は一般的な Linux OS を搭載した計算クラスター 1 台で計算可能である。本研究では CA-align と代表的な構造アライメントプログラム

(TM-align[1], FATCAT[2], GR-Align[3])の比較を行った。本研究では、次の3つの視点から構造アライメント法の性能の検証を行った：① 類縁関係が明らかとなっているタンパク質同士について、profile-profile アライメントとどれだけ一致しているか。② 構造アライメントからマルチプルアライメントを構築し、整合性が取れているか。③ 類縁関係が確認されているドメイン分類をどれだけ再現できるか。以下にその詳細を記述する。

3.1 Profile-profile Alignment (HHalign)との比較

計算された構造アライメントと HMM-HMM アライメントプログラム (HHalign[4]) から計算されたアライメントとの一致度を構造アライメントの精度とした。HMM-HMM アライメントは、三次元構造に依存しない配列アライメントであり、十分に E-value が低いアライメントは非常に高い精度を示す。ドメインデータベース(SCOPe Release 2.03, Oct 2013 [5])から作成された 3,660 個のペアに対するアライメントの結果を図2に示す。検証で使った HHalign のアライメントは E-value が 10^{-10} 以下であり、高い類縁関係が示されている。図2にあるように CA-align のアライメントは、最も精度の高い結果を示した。

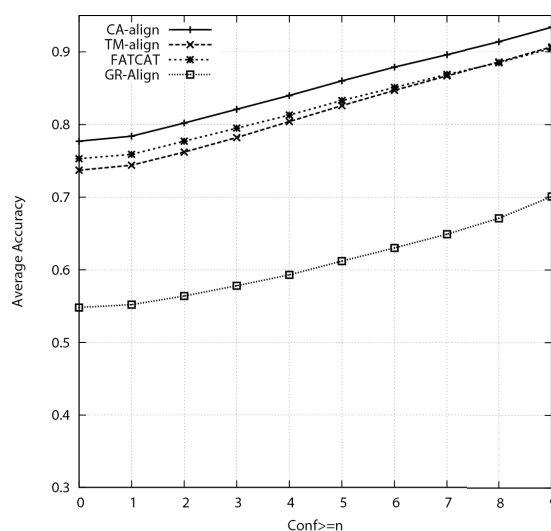


図2 アライメント精度の比較 (縦軸：平均精度、横軸：HHalign の信頼値 Confidence value)

3.2 マルチプルアライメントの整合性

例として、タンパク質 A , B , C について構造アライメントを3つ計算した場合(A - B , B - C , C - A)、3組のアライメントに整合性があるほど精度が

高いアライメントであると考えられる。例えば、A-BにおいてAの*i*番目 (A_i) とBの*j*番目 (B_j) がアライメントされており、B-Cにおいて B_j とCの*k*番目 (C_k) がアライメントされている場合、C-Aでは、 C_k と A_i がアライメントされているならばその位置で整合性[6]があると言える。

SCOPE から1セットに3組のアライメントが含まれる7,384セットの検証データを作成し検証を行った(図3)。横軸には、アライメント長に対する整合性のあるアライメント残基数の割合の平均値(Rate)、縦軸に整合性のあるアライメント残基数のCoverageを示す。CA-alignは最も高いCoverageを示し、HHalignの次に高いRateを示した。

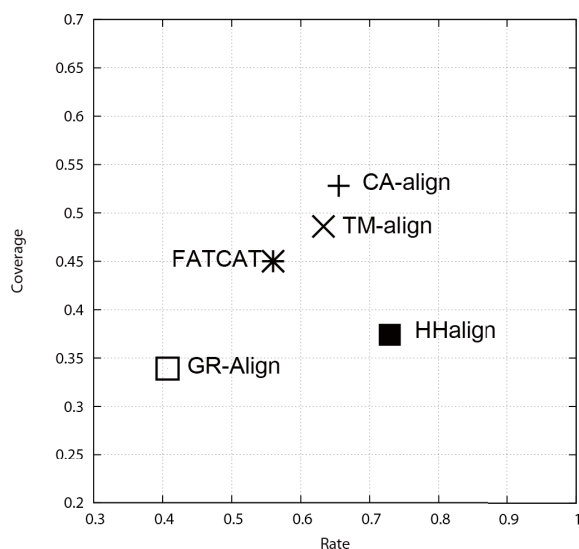


図3 マルチプルアライメントの整合性 (横軸: Rate, 縦軸: Coverage)

3.3 ドメインデータベースの分類との比較

SCOPEにおいて、Evolutionary relationshipがあるとされる分類の Superfamily を基準とし、構造類似度の分類性能を比較した。SCOPEからランダムに選んだ500ドメインについて総当りの構造アライメント計算をし、ROC解析を行った。図4に示したように、横軸 False positive rate に対して縦軸 Recall をプロットしたところ、CA-alignは最も高いArea Under Curve(AUC)値を示した。加えて、CA-alignはタンパク質構造情報のみでHHalignよりもずっと高い分類性能を示している。これはアミノ酸配列上類縁関係が検出できないものでも、superfamilyを検出できていることを意味する。

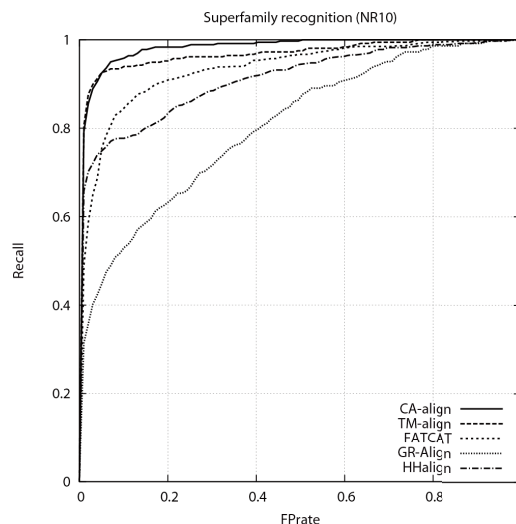


図4 Superfamily 分類の ROC curve

3.4 適用例1 (Calmodulin-like protein)

異なるコンフォメーションを持つ Calmodulin-like protein の構造比較を行った(図5)。使用したタンパク質構造は PDBID 1ncx (オープンダンベル型)、2sas (クローズ型) である。CA-alignによってアライメントされた領域をCartoonで表示し、アライメントされた場所で共通して存在する残基間ネットワークを破線で示す。図5で示したように、三次元空間上構造が類似していなくても CA-align はほぼ全ての領域でアライメントを検出している。また、CA-alignが計算したアライメントでは共通した残基間ネットワークが多く存在していることが分かる。

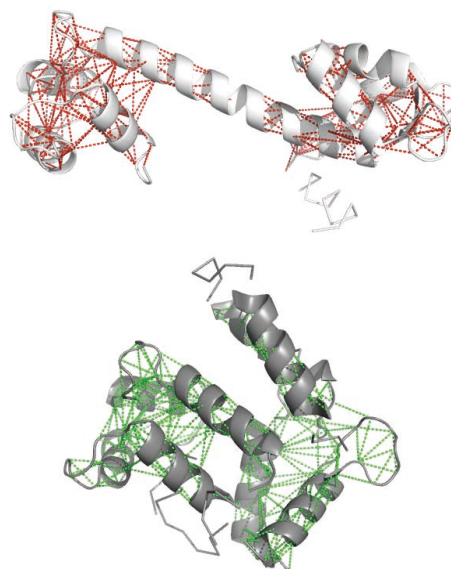


図5 CA-align の構造アライメント: 1ncx (white)と 2sas (gray)

比較対象として TM-align の結果(図 6)を示す。TM-align は C 末端部分のカルシウム結合ドメインのみでアライメントを計算できている。

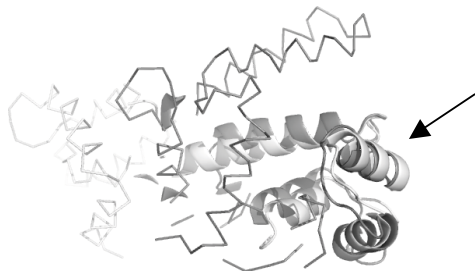


図 6 TM-align の構造アライメント : 1ncx (white)と 2sas (gray) 矢印は C 末端側

3.5 適用例 2 (HIV-1 reverse transcriptase)

同様に HIV-1 reverse transcriptase の阻害剤結合型 (SCOPID: d2rkia2), DNA 結合型 (SCOPID: d1rtdd_) の構造アライメントを計算した (図 7, 8)。CA-align の Alignment length=414 に対し、TM-align の Alignment length=253 であった。

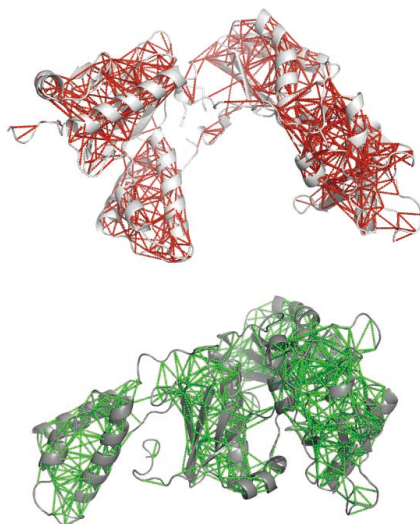


図 7 CA-align の構造アライメント : d2rkia2 (white)と d1rtdd_ (gray)

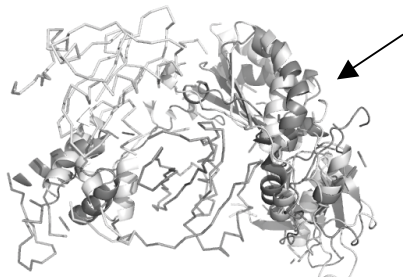


図 8 TM-align の構造アライメント : d2rkia2 (white)と d1rtdd_ (gray) 矢印は N 末端側

4 結論

本研究で開発された CA-align は、タンパク質の構造変化に依存することなく構造類似性を検出できる構造アライメント法である。

検証の結果、大きな構造変化を含まないドメインデータベースの構造に対して従来の構造アライメント法よりも高いアライメント精度を示した。また検証結果から CA-align でマルチプルアライメントを構築するならば最も高い Coverage を持つアライメントが計算できることが示された。さらにドメインデータベースの分類を最も精度よく再現できることから、機能未知のタンパク質の機能推定に適用可能であることが示された。

適用例 1, 2 で示したように、CA-align が計算する構造アライメントは単にアミノ酸残基の対応関係を明らかにするだけではない。2つのタンパク質構造を比較することで、共通して存在する残基間相互作用ネットワークを明らかにする。CA-align が提供するこれらの情報は、ホモロジーモデリング法における拘束条件の設定や、タンパク質フォールディングの解析、タンパク質間相互作用解析に大きく貢献すると期待される。

謝辞

本研究は JSPS 科研費若手研究 B 26730151 の助成を受けたものです。

参考文献

- [1] Zhang Y, and Skolnick J, Nucleic Acids Res, 33, 2302-9 (2005).
- [2] Ye Y, and Godzik A, Bioinformatics, 19 Suppl 2, ii246-55 (2003).
- [3] Malod-Dognin N, and Przulj N, Bioinformatics, 30, 1259-65 (2014).
- [4] Remmert M, Biegert A, Hauser A, and Soding J, Nat Methods, 9, 173-5 (2012).
- [5] Fox NK, Brenner SE, and Chandonia JM, Nucleic Acids Res, 42, D304-9 (2014).
- [6] Sadowski MI, and Taylor WR, Bioinformatics, 28, 1209-15 (2012).

KO12

Staphylococcus aureus 莢膜合成酵素 CapF と 低分子化合物の物理化学的な相互作用解析

(東大院・工¹、東大・OCDD²、東大院・新領域³、東大・医科研⁴)

○長門石 暁^{1,2}、中納 広一郎³、宮房 孝光³、

Caaveiro Jose¹、津本 浩平^{1,2,3,4}

【背景】

黄色ブドウ球菌はグラム陽性の球菌で、世界中に幅広く存在し、ヒトをはじめとした哺乳類に生息する常在菌である。しかし免疫力の低下した患者や乳幼児に感染すると重篤な疾病を引き起こす。さらに黄色ブドウ球菌は各種抗生物質に対して容易に耐性を獲得するため、深刻な社会問題にもなっている。さらに、口腔や皮膚などに感染し炎症を起こすなど、日常生活においても関わることの多い病原菌である。したがって黄色ブドウ球菌に対する薬剤、抗菌剤の開発は社会的意義が強く、かつ急務である。

黄色ブドウ球菌の病原因子の一つである莢膜は糖の重合体であり、黄色ブドウ球菌の表面を覆い、食食から免れることにより黄色ブドウ球菌の病原性を高めている。この莢膜合成に必須の酵素である CapF は、CapE と共同的に働き、NADPH を用いて基質を還元することにより莢膜の合成の一端を担っている。また、近年 CapF の構造が明らかとなり、構造に基づく CapF の酵素メカニズムが明らかになりつつある [1]。

そこで本研究では、CapF の機能を阻害することで黄色ブドウ球菌の莢膜合成を阻害する抗菌剤の探索を行うことを目的とし、CapF に対する化合物スクリーニングを行った。

化合物スクリーニングには、近年注目が集まっている Fragment-Based Drug Discovery (FBDD) 法を選択した。FBDD は 250 Da レベルの低分子サイズの化合物を用いるため、ケミカルスペースの網

羅性が高く、したがって少ない化合物数で効率よく結合する化合物を選抜することができるという利点があると期待されている。その一方で、結合する化合物の親和性は低いことが予測されるため、低親和性化合物を効率的に評価する実験系を構築する必要がある。そこで本研究では、物理化学的解析技術を用いて、CapF に対し特異的に結合する化合物の選抜を試みた。

【目的】

CapF に特異的に結合し、その酵素活性機能を抑制する化合物を探索することを目的とした。創薬オープンイノベーションセンター(OCDD)から供与を受けたフラグメントライブラリーから候補化合物を選抜するために、表面プラズモン共鳴法 (SPR) を用いた相互作用測定と、分光学的手法を用いた酵素活性阻害測定とによる一次スクリーニングを行った。選抜した化合物において、再現性と非特異結合判断するために、再度 SPR を用いた濃度依存的な結合実験を行った。濃度依存的に結合量が上昇した候補化合物に関し、詳細な物理化学的解析を行った。

【結果と考察】

1) 1 次スクリーニング

フラグメントライブラリーを用いて、CapF に結合する低分子化合物を SPR によって選抜した。その際、結合レスポンスの大きさと共に、会合速度と解離速度の速いセンサグラムのものをヒット

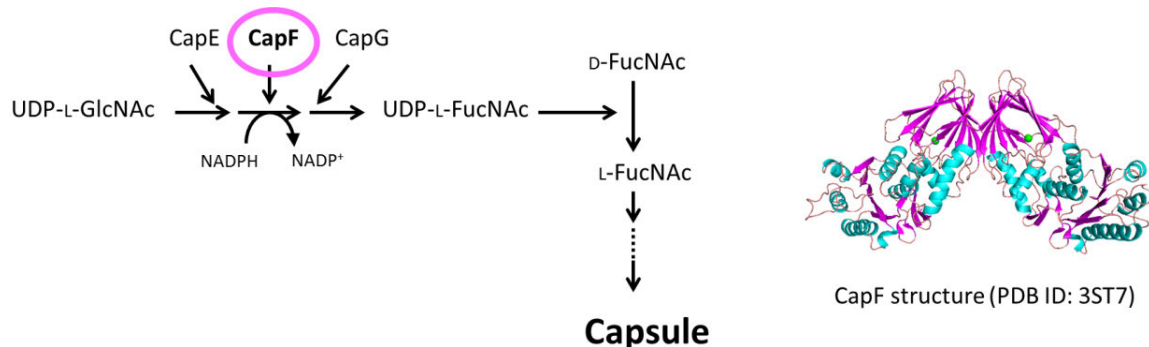


Figure 1. 黄色ブドウ球菌由来の莢膜合成酵素 CapF による反応プロセス

候補とし、会合速度や解離速度の遅いものを排除した(Figure 2A, B)。一方で酵素活性阻害測定として、NADPH の消費量を吸光度測定によりスクリーニングを実行した(Figure 2C)。

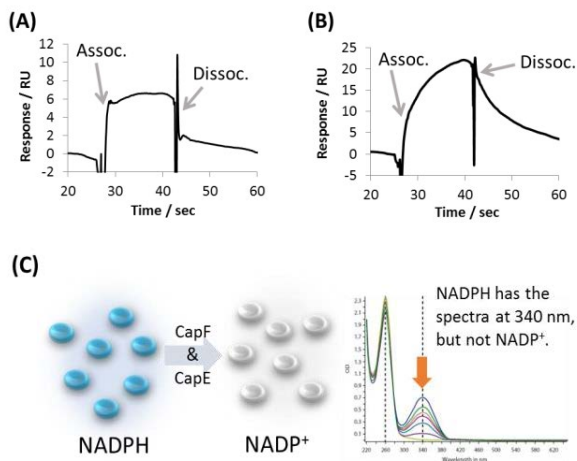


Figure 2. 1 次スクリーニングにおける (A, B) SPR のセンサグラム, (C) 340 nm に吸収スペクトルをもつ NADPH の消費量変化

2) 候補化合物のカウンターアッセイ

選抜した 29 種の化合物について、1 次スクリーニングにおけるデータの再現性、非特異性を評価するために、SPR を用いた濃度依存的な結合実験を行った。その結果、29 種化合物のうち 6 種の化合物において濃度依存性があり、特異的な結合レスポンスが得られた。これら 6 種の化合物に関し、SPR を用いて解離定数解析を行った結果、1 種の化合物(Compound 1)において、 $K_D=514 \mu\text{M}$ が得られた。また、この Compound 1 において酵素活性阻害測定を行ったところ、 $0 \mu\text{M}$ から $300 \mu\text{M}$ まで濃度領域において、NADPH の消費が濃度依存的に阻害されていることが分かった。

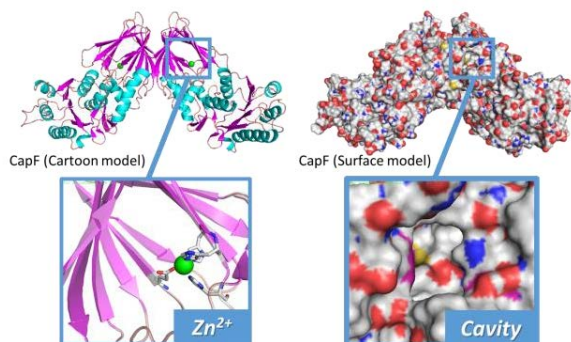


Figure 3. CapF の亜鉛イオン結合部位とその周辺構造

3) ヒット化合物の物理化学的解析

Compound 1 に関して、結合部位、結合様式に対する詳細な理解を目指し等温滴定型熱量測定 ITC を用いた結合解析を行った。その結果、Compound 1 は CapF に対し発熱を伴い結合することが分った(Figure 4A)。また、Compound 1 に類似した化合物を用いて ITC 評価を行ったところ、ヒドロキシル基が発熱反応に重要な役割を果たしていることが明らかとなった。このヒドロキシル基の存在は、CapF に結合している亜鉛イオンを認識していることが示唆された (Figure 4B, C)。さらに、数種の構造類似体を用いた物理化学的な相互作用解析による構造活性相関より、Compound 1 は CapF に対して特異的に結合していることが示唆された。

【結論】

本研究において、CapF に対する阻害剤スクリーニングにより、CapF に特異的結合し、かつ酵素活性を阻害する化合物を得ることに成功した。ヒット化合物は、亜鉛イオンを認識し、発熱反応を有する非共有結合にて CapF と相互作用していることが明らかとなった。

参考文献】

[1] Miyafusa, T. *et al.*, *Biochem. J.*, 443, 671-680, 2012.

【謝辞】

本研究のスクリーニングにおいて、東京大学 OCDD のご協力をいただいた。また本研究は文部科学省 創薬等支援技術基盤プラットフォーム事業 (H24-28)と日本学術振興会 科学研究費補助金基盤研究(A) (H25-27)の支援のもと遂行された。

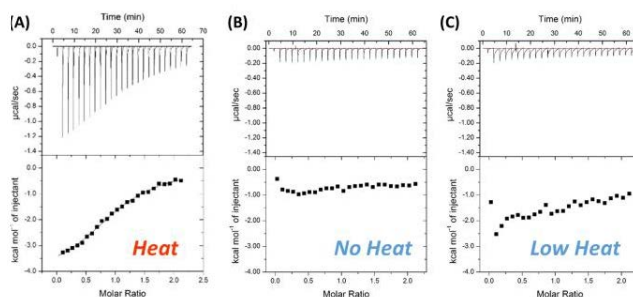


Figure 4. ヒット化合物の ITC 測定。 (A) CapF、 (B) EDTA 存在における CapF、 (C) 亜鉛イオン

//// Activities ////

第 42 回構造活性相関シンポジウム開催報告

第 40 回構造活性相関シンポジウム実行委員長
熊本大学薬学部 今井 輝子

銀杏の色づく季節に、熊本市において第 42 回構造活性相関シンポジウム（会場：くまもと森都心プラザ、期間：2014 年 11 月 13 日（木）～ 11 月 14 日（金）、主催：日本薬学会構造活性相関部会、後援：日本化学会・日本農芸化学会・日本分析化学会・日本農薬学会、有機合成化学協会）を開催しました。昨年度に引き続き、口頭発表時間の 20 分への一本化及び「構造活性相関シンポジウム優秀発表賞」（SAR Presentation Award）対象者拡大の 2 点を継続いたしました。九州での単独開催にも関わらず、特別講演 1 件、招待講演 3 件、依頼講演 1 件、口頭発表 13 件、ポスター発表 29 件となり、例年に匹敵する講演数になりました。参加者も 146 名（招待者を含む）に達し、盛会のうちに終えることができました。これもひとえに、参加いただきました皆様と、実行委員の皆様、ならびに構造活性相関部会幹事の先生方のご支援の賜と存じます。主催頂きました日本薬学会構造活性相関部会はじめ、後援頂きました学協会に感謝致しますと同時に、ご講演頂きました先生方にお礼申し上げます。また、開催資金のご援助を頂きました、日本薬学会ならびに関係企業の方々に深くお礼申し上げます。

横井毅先生（名古屋大学大学院医学研究科）には「薬物肝障害の多様性への対応と予測」の演題で特別講演をしていただきました。また、池谷鉄兵先生（首都大学東京 理工学研究科）には「生きた細胞内での蛋白質の立体構造とダイナミクス」、Dr. Kun-Yi Hsin（沖縄科学技術大学院大学）には「Molecular docking simulations in the era of network pharmacology」、山縣ゆり子先生（熊本大学大学院生命科学研究部）には「X 線結晶構造解析による酵素反応機構の解明」と題して、それぞれのご専門の立場から貴重なご講演をいただきました。

次年度の第 43 回構造活性相関シンポジウムは、第 10 回薬物の分子設計と開発に関する日中合同シンポジウムとの併催が企画されており、新潟薬科大学の石黒 正路先生のお世話で、新潟日報メディアシップで 2015 年 9 月 28 日（月）～ 9 月 29 日（火）に開催される予定です（日中合同シンポジウムは 2015 年 9 月 26 日（土）～ 9 月 29 日（火）開催）。皆様のご参加、ご講演、ご討論により、より活発な討論会になりますよう、宜しく願い申し上げます。



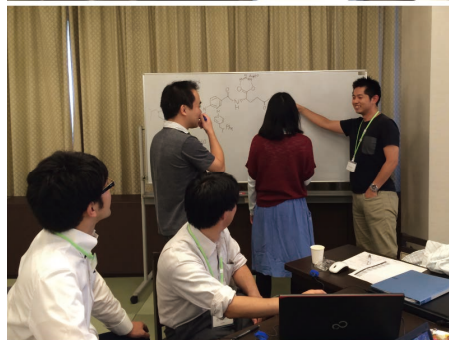
//// Activities ////

第1回 QSAR 解析ハンズオンセミナー開催報告

第1回 QSAR 解析ハンズオンセミナー
実行委員長 山下 富義

2014年9月18、19日の両日、ニューウェルシティ湯河原にて第1回 QSAR 解析ハンズオンセミナーが開催されました。本セミナーは、QSAR 解析を適正に使用し、創薬研究に役立つ基本的能力を養うことを目的として、構造活性相関部会が新たに企画・開催した少人数のチュートリアルセミナーです。初めての開催ということもあって、実行委員会を中心に、企画素案の作成から、セッションの概要、資料の作成、セミナーの実施に至るまで、約2年かけて進めてきました。これまでに部会で開催してきた集会は構造活性フォーラム、構造活性相関シンポジウムの2つでしたので、まずは会告案内を周知できるのかと懸念しながらのスタートでした。チュートリアルセミナーという性格上、募集人数は30名と限定しましたが、それでも正直不安が残る数でした。実際に募集が始まると、皆様から多数問い合わせをいただき、締切を延長することなく予定人数の30名に到達しました。この30名はかなりバラエティーに富んだ背景をお持ちで、一般/学生、製薬会社/農薬会社/アカデミア、分子設計/合成化学/薬理・動態・毒性、男性/女性でさまざまな形で分類できました。本チュートリアルセミナーの売りは何といってもグループワークでしたので、このバックグラウンドに基づいて各グループ5名で計6グループに分かれ、各セッションに臨んでいただきました。初日午後はQSAR 解析演習、2日目午前は分子シミュレーション演習、2日目午後には分子設計グループワークを行いました。QSAR 解析演習では、セッションリーダーの赤松美紀先生（京都大学大学院農学研究科）が、まず、クラシック QSAR の物理化学的背景、用いるパラメーターの意義等について概説されました。その後、各グループに課題として与え QSAR 式の構築に取り組んでいただき、さらに発表・討論による意見交換を通じて QSAR 式の作成の手順や注意点を学んでいただきました。分子シミュレーション演習では、理化学研究所の本間光貴先生にリーダーとなっていただき、妥当なドッキング計算を行う上で考慮すべきこと等を概説いただいた後、タンパク質初期構造の選択、妥当な複合体構造の選択についてグループ討議をしていただき、その内容について意見交換を行いました。2日目午後の分子設計グループワークは、中外製薬の大田雅照先生がリードして下さり、QSAR 式の構築とそれに基づく次の合成展開、タンパク質/リガンド複合体構造の解釈とそれに基づく合成展開について、各グループでの討議、その結果について発表・討論を行っていただきました。各グループからは予想以上にバラエティーのあるアイデアが紹介され、なかにはユニークな意見もあり、全体討論は大いに盛り上がりしました。1日半というタイトなスケジュールではありましたが、少人数セミナーということもあって活発な意見交換ができ、お互いの交流という点でも非常に意味のある会合であったと思います。終了後、参加者の皆様からアンケートにご回答いただきましたが、各セッションに対し高い評価をいただき、次回開催を要望する声も多数いただきました。

初めての試みで不安が一杯でしたが非常に好評をいただきましたので、本企画は継続して行うことになりました。先般の常任幹事会では3年に1回の開催と決定しましたので、次回は2017年ということになります。是非、この企画を心にお留めおきいただき、次回開催にも奮ってご参加を賜りますようお願い申し上げます。



//// Activities ////

<会告>

構造活性フォーラム 2015

「様々なリード創製手段において創薬に貢献するインシリコ技術」

主催： 日本薬学会構造活性関連部会

協賛： 日本化学会、有機合成化学協会、日本農芸化学会、日本農薬学会、日本分析化学会

会期： 2015 年 6 月 12 日（金）

会場： 日本薬学会長井記念ホール（〒150-0002 東京都渋谷区渋谷 2-12-15）

フォーラムホームページ： <http://www.qsarj.org/forum2015>（4 月 1 日公開）

開催趣旨： 企業創薬での経験から、インシリコ技術の特性は目的に応じて対応できるという柔軟性とその結果として得られる技術の多様性であると感じている。また、この柔軟性のゆえに、FBDD、SBDD、virtual screening、HTS など様々な創薬アプローチにおいてインシリコ技術が創薬に必須の技術として貢献できているのではないかと感じている。本フォーラムでは、最新技術動向の把握というより、様々な場面で利用されているインシリコ技術を網羅的に提示・俯瞰し、創薬に貢献するインシリコ技術に共通する特性を考えることにより、さらなるインシリコ技術開発、さらに多様な場面での応用と貢献の深化を目的とした。本フォーラムが、企業インシリコ担当者の実践のヒントに、メディシナルケミストにとって利用可能なインシリコ技術の理解に、アカデミアの企業創薬の理解と技術開発のヒントの一助になれば幸いである。

プログラム：

講演 1. 「リード創製における社内外データの活用」

服部 一成（塩野義製薬）

講演 2. 「キナーゼ創薬の基盤技術 –FBDD と複数ターゲット HTS のデータ解析–」

黒野 昌邦（小野薬品工業）

講演 3. 「New approaches to old targets through fragment based drug discovery」

Mark Hixon (Takeda California Inc.)

講演 4. 「バーチャルスクリーニングによるリード探索 –メディシナルケミスト視点からの化合物選択–」

田中 大輔（大日本住友製薬）

講演 5. 「*In silico* technologies in drug discovery: examples at Chugai」

大田 雅照（中外製薬）

パネルディスカッション「創薬とインシリコ技術」

参加登録および申込締切日： 当日参加受付はありません。5 月 22 日（金）までに、フォーラムホームページから事前参加登録をお願いいたします。参加人数が上限に達しましたら、参加登録を打ち切らせていただきますのでご了承ください。参加登録後、5 月 31 日までに銀行口座振り込みをお願いいたします。なお、お振込みの際には「ご所属とご氏名」を振込人として下さい。領収書は当日受付デスクでお渡しいたします。

参加費：（薬学会会員・非会員ともに）一般 5,000 円、 学生無料

懇親会： 17:15～ 長井記念ホール前ロビーにて

懇親会費：（薬学会会員・非会員ともに）一般 4,000 円、 学生 1,000 円

問合先： 構造活性フォーラム 2015 実行委員会 大田 雅照（実行委員長）

〒412-8513 御殿場市駒門 1-135 中外製薬(株)研究本部 創薬化学研究部 内

Tel: 0550-87-6722 Fax: 0550-87-5326 E-mail: forum2015@qsarj.org

//// Activities ////

< 会告 >

第 43 回構造活性相関シンポジウム・

第 10 回薬物の分子設計と開発に関する日中合同シンポジウム 共同開催

主催： 日本薬学会構造活性相関部会

後援： 有機合成化学協会、日本分析化学会、日本農芸化学会、日本化学会、日本農薬学会

協賛： 新潟薬科大学

会期： 2015 年 9 月 27 日（日）から 2015 年 9 月 29 日（火）

会場： 新潟日報メディアシップ 日報ホール
（〒950-0088 新潟県新潟市中央区万代 3-1-1 <http://niigata-mediaship.jp/>）

日程：

- 9 月 27 日（日） 第 10 回薬物の分子設計と開発に関する日中合同シンポジウム
（Welcome Party）
- 9 月 28 日（月） 第 43 回構造活性相関シンポジウム・第 10 回薬物の分子設計と開発に関する
日中合同シンポジウム合同セッション
（懇親会）
- 9 月 29 日（火） 第 43 回構造活性相関シンポジウム
（エキスカッション）

招待講演/特別講演：

- 9 月 27 日（日） 招待講演（3 件の予定）
- 9 月 28 日（月） 招待講演 午前
舩屋圭一先生（PeptiDream Inc. COO）
特別講演 夕方
嶋田一夫先生（東京大学大学院薬学研究科 教授）
- 9 月 29 日（火） 招待講演 午前
小林拓也先生（京都大学大学院医学研究科 准教授）
招待講演 午後
福澤 薫先生（日本大学松戸歯学部 助教）

発表登録・参加申込など：

発表申込締切日：6 月 30 日
講演要旨締日：8 月 7 日
参加申込締日：9 月 18 日

参加申込方法などの詳細に関しては決まりしだい順次 シンポジウムホームページにてお知らせします（http://www.nupals.ac.jp/~sar_symp2015/）。

参加登録費： [一般]予約 10,000 円、[学生]予約 5,000 円（両シンポジウム共通）※

※:両シンポジウム、もしくは、どちらかのシンポジウムに参加する場合でも参加登録料は 10,000 円とさせていただきます。

Welcome Party 参加費：

[一般] 7,000 円、[学生] 3,000 円

懇親会参加費：

[一般] 7,000 円、[学生] 3,000 円

エキスカーション（佐渡を予定）参加費：

[一般、学生とも] 10,000 円

問合先：

〒956-8603 新潟県新潟市秋葉区東島 265-1 新潟薬科大学

「第 43 回構造活性相関シンポジウム・第 10 回薬物の分子設計と開発に関する
日中合同シンポジウム 実行委員会」 石黒正路

e-mail: sar_symp2015@nupals.ac.jp

URL: http://www.nupals.ac.jp/~sar_symp2015/

構造活性相関部会の沿革と趣旨

1970年代の前半、医薬品を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と分子設計に対する新しい方法論が世界的に台頭してきた。このような情勢に呼応するとともに、研究者の交流と情報交換、研究発表と方法論の普及の場を提供することを目的に設立されたのが本部会の前身の構造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980年からは年1回の「構造活性相関シンポジウム」が関係諸学会の共催の下で定期的に行われるようになった。

1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。構造活性相関懇話会は1995年にその名称を同研究会に改め、シンポジウム開催の実務担当グループとしての役割を果たすこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の構造活性相関に関する学術・研究の振興と推進に向けて活動している。現在それぞれ年1回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR Newsを年2回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。

本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。

(http://bukai.pharm.or.jp/bukai_kozo/index.html)

編集後記

日本薬学会構造活性相関部会誌 SAR News 第28号をお届けいたします。今号の Perspective/Retrospective は、長浜バイオ大学の土方先生、辻先生、白井先生に超分子タンパク質複合体構造について解説いただきました。Cutting Edge では東大医科研の Dutta 先生、Patil 先生、中井先生から、タンパク質の天然変性領域について解説いただきました。レベルファイブの池田先生と産総研の富井先生には、タンパク質のリガンド結合部位の類似性の研究法をご紹介いただきました。タンパク質の構造、機能、柔軟性とそれらの間にある秩序について多面的な切り口があることに、この分野の発展の余地を大いに感じました。ご寄稿いただいた先生方には、大変お忙しい中でのご執筆、心よりお礼申し上げます。この SAR News が今後とも構造活性相関研究の先端情報と展望を会員の皆様にご提供できることを、編集委員一同願っております。

長らく本誌の編集委員をお務めくださいました田辺三菱製薬の福島千晶さんが退任されました。今までのご貢献に深謝いたします。本号から大正製薬の小田晃司さんが編集委員として加わりました。よろしく願います。(編集委員会)

SAR News No.28 平成 27 年 4 月 14 日

発行: 日本薬学会 構造活性相関部会長 高木 達也

SAR News 編集委員会
(委員長) 飯島 洋
小田 晃司
粕谷 敦
竹田-志鷹 真由子
幸 瞳

*本誌の全ての記事、図表等の無断複写・転載を禁じます。