

新たなデータセットによる 長距離フラグメントリンキング手法の再評価

津嶋 佑旗¹ 柳澤 溪甫¹ 大上 雅史¹ 秋山 泰^{1,a)}

概要：フラグメントベース創薬の一手法であるフラグメントリンキングにおいて、既存の手法では長距離に対応することが難しい。そこで我々は第 67 回バイオ情報学研究会にて、フラグメント間の距離やドッキングスコアなどを複合した評価関数によるビームサーチで小型のリンカー要素の組み合わせを探索する手法を提案した。本研究では、5 種類の異なる標的タンパク質からなるデータセットと 30 種類のリンカー要素を新たに作成して手法の再評価を行った。改めて手法の汎用性を確認し、実行時パラメータの最適化と独自評価関数の効果の確認を行った。

1. 導入

1.1 Fragment-Based Drug Design (FBDD)

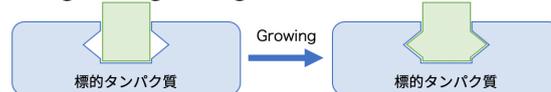
FBDD は、一般的な創薬で扱われる化合物群に比べて小型な、フラグメントと呼ばれる化合物をもとに薬剤分子を設計する方法論である。FBDD では、フラグメント単体の活性が一般的な薬剤に比べて大幅に低いことが多く、従来の実験手法では阻害能などの検出が難しいという課題があったが、高感度の実験を行う手法が開発されたこと、分子ドッキング計算による仮想的なフラグメントスクリーニングが可能になったことなどにより、FBDD は近年注目を集めつつある方法論となっている [1]。

実験的手法または計算的手法によって発見された活性をもつフラグメントからリード化合物を構成する基本方針として、Fragment growing, Fragment merging, Fragment linking の 3 つがある。これらの概念図を図 1 に示す。

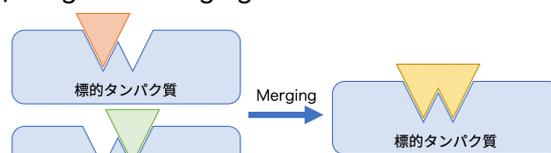
(a) Fragment growing

結合ポケットの一部を占めるフラグメントを発見した場合にとられる方法で、フラグメントの部分構造を改変したり化学修飾を行ったりして活性を向上させる。化学構造の改変は、まだ使用できていないポケットの残りの領域の物理化学的性質や既知の活性化合物の情報をもとに行う [1]。この手法では、単一の結合ポケットのみ利用できる。

(a) Fragment growing



(b) Fragment merging



(c) Fragment linking

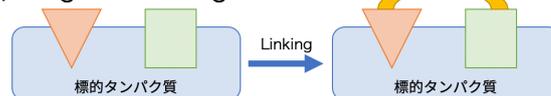


図 1 FBDD における基本方針の概念図

(b) Fragment merging

重なり合う複数のポケットを占めるフラグメントから共通部分を見つけてハイブリッド化する方法である [2], [3]。フラグメント同士だけでなくフラグメントと既知活性化合物の間でマージを行うこともできる。この手法では、隣接した位置にある結合ポケットを利用できる。

(c) Fragment linking

独立した異なる結合ポケットを占める複数のフラグメントの間に「リンカー」と呼ばれるつなぎの化学構造を導入することで単一の分子として構成し、全体として高い活性を得る方法である。Fragment linking に特有の利点として、離れた位置に存在する複数の結合ポケットを活用できることが挙げられる。また、リンクされた分子の結合自由

¹ 東京工業大学情報理工学系
Department of Computer Science, School of Computing,
Tokyo Institute of Technology, Meguro, Tokyo 152-8550,
Japan

^{a)} akiyama@c.titech.ac.jp

エネルギーは使用したフラグメントの結合自由エネルギーの和になるが、分子運動に制約が生じることでさらに強力に標的タンパク質と結合させることも可能であることも知られている [4].

1.2 Fragment linking の既存手法と課題

Fragment linking は Fragment growing や Fragment merging に比べて薬剤設計が難しいことが知られている。Bancet ら [5] によれば、FBDD によって設計された治験中または承認済み薬剤 46 種類のうち 41 種類は Fragment growing によって設計されている。Fragment growing では、従来の薬剤開発で広く用いられているヒット最適化やリード最適化の手法が利用できる一方、Fragment linking では未だに広く使用されるような汎用的手法の開発には至っていない。既存の Fragment linking 手法としては、単純で柔軟なリンカーを使用する方法が報告されている [6], [7] ほか、予め用意したリンカーの候補となる化学構造のライブラリから、フラグメントの相対位置情報を用いて検索する計算ツールとして、BiosolveIT の LeadIT や CCG MOE の linking/scaffold-replacement といったソフトウェアが開発されているが、これらの手法では、遠く離れた位置にあるフラグメント同士を接続することは難しい。

1.3 第 67 回研究会における提案手法

複数のポケットを組み合わせて活用できる Fragment linking の利点をより活用するため、離れた位置にあるフラグメント同士を接続する方法が必要である。そこで我々は、第 67 回バイオ情報学研究会にて、2つのフラグメント間をリンクする、長距離にも対応可能な手法を提案した [8].

提案手法では、まず、環構造を含む小型のリンカー要素数十種類からなるリンカーライブラリを用意する。それらのリンカー要素を、始点となるフラグメントから終点となるフラグメントに向かって、立体的な配置を探索しながら逐次接続する。また、リンカー要素を導入した中間構造は、標的タンパク質との間のドッキングスコアに基づいて選抜する。これは計算機科学的には、リンカー要素とその立体配置の組み合わせからなる木構造の探索になるが、その組み合わせは膨大であり全探索することは難しい。そこで、リンカー要素を1つ接続するごとに相互作用を評価、評価値の良い限られた数の構造に対して次のリンカー要素の導入を行うようにする。これはビームサーチである。木探索における幅優先探索では1ステップずつ全ての場合を列挙するが、ビームサーチでは各ステップのノードに対しヒューリスティックな評価を行い、評価値のよい順にビーム幅分のノードに限って次のステップの計算を行うことで、近似解法ではあるが計算量が削減できる。

提案手法による Fragment linking は大まかに図 2 に示す流れで行う。

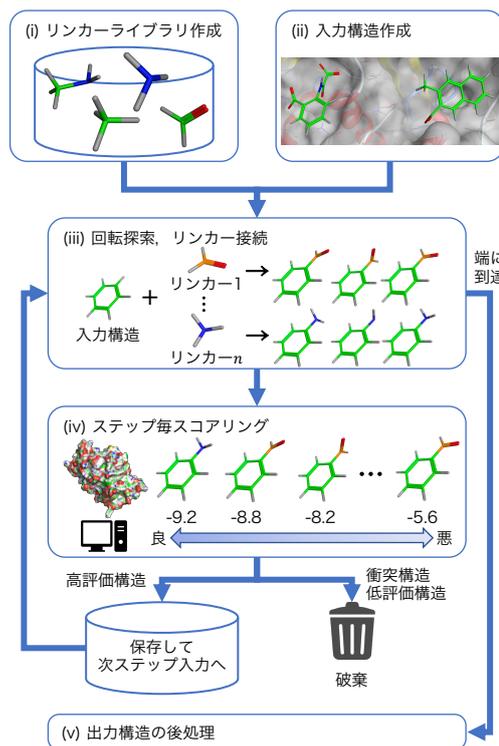


図 2 前回提案手法の流れ [8]

1.3.1 リンカー要素の準備 (図 2-(i))

提案手法では、リンカー要素として使用する化合物の立体構造が必要となるため、予め立体構造を用意する。リンカー要素は2つの分子の間の橋渡しを行うので、他の構造と接続を行う点であるアタッチメントポイント (以下 AP と略記する) を2つ持つ。

1.3.2 タンパク質とフラグメントの準備 (図 2-(ii))

標的となるタンパク質に2種類のフラグメントが結合した立体構造を用意する。この結合ポーズは、X線結晶構造解析などの実験的手法で求めたものや、ドッキングシミュレーションによって予想された構造が使用できる。

1.3.3 リンカー要素の回転探索 (図 2-(iii))

リンカー要素は原点を中心として、入力構造とリンカー要素の化学結合を中心とする軸 (主探索軸) と、それに直交する2つの軸 (副探索軸) の周りに一定の範囲で回転させた構造を列挙する。この回転探索の刻み幅と副探索軸周りの探索範囲はパラメータとして与える。回転操作を加えたリンカー要素は入力構造の AP に移動させて結合させ、リンカー要素が接続された中間構造を生成する。またこのとき、新たに追加したリンカー要素の AP と終点のフラグメントの AP が 1 Å 以内にある場合はその間を結合させて出力構造とする。

1.3.4 スコアリング (図 2-(iv))

1.3.3 で生成した中間構造から次のリンカー要素を導入する構造を選択するため、標的タンパク質との物理化学的な

表 1 設計した評価関数, s はドッキングスコア, w は分子量, d は残距離を表す.

番号	式	意味
0	$f = s$	Vina スコア (基準)
1	$f = \frac{s}{w}$	リガンド効率 (1)
2	$f = \frac{s}{\ln(w)}$	リガンド効率 (2)
3	$f = \frac{s}{d}$	距離効率
4	$f = \frac{\Delta s}{\Delta w}$	差分リガンド効率
5	$f = -\frac{s}{d}$	逆距離効率

相互作用の評価であるドッキングソフトウェア AutoDock Vina[9] のスコア関数を使用する. また, 中間構造の内部のエネルギーもこのスコア関数により計算することで内部衝突する構造の破棄も行う.

さらに, このスコアリングにおいて, ドッキングスコアに対してさらに工夫を加えることを考える. 分子量や他方のフラグメントまでの残距離などを使用し, より優れた構造を出力できる可能性を探るため, 表 1 に示す 6 種類の評価関数を設計した. これらの評価関数によって得られる出力構造の差異については実験で検証する.

リガンド効率

創薬においてはリード化合物の分子量を小さくすることは重要であると指摘されていることから, 分子量が小さいわりに強い活性をもつ化合物を選択する指標として, 重原子数あたりの結合自由エネルギーを表すリガンド効率が提案され [10], 広く用いられている. この定義を元に, 分子量あたりのドッキングスコアを表す値を評価関数とすることで, 分子量を抑えつつより良いドッキングスコアの構造が得られることを期待する. 評価関数 1 と 2 は, 分子量に対する評価値の変化を変化させている. また, 評価関数 4 は, リンカー導入前後の分子量変化とドッキングスコア変化を用いることで, 追加したリンカー要素に絞って評価することを期待する.

距離効率と逆距離効率

Vina スコアをリンクすべき残りきよりで除算した値を距離効率と定義し, 評価関数 3 として使用する. Vina スコアは負であり, 残り距離は正なので, この評価関数を小さくするには残り距離をできるだけ短く, ドッキングスコアをできるだけ小さくする必要がある. 他方フラグメントまで回り道を避けてできるだけ短い距離でリンクすることが期待される.

一方, 評価関数 5 では, 評価関数 3 の -1 倍の値を使用する. この関数を用いた場合, 残り距離をできるだけ減らさずにドッキングスコアを小さくする必要がある. これにより, 小型でかつドッキングスコアが改善するリンカーを選択し, 全体のドッキングスコアが改善することを期待した.

1.3.5 出力構造の最終処理 (図 2-(v))

1.3.3 でリンク成功と判断された構造は, 2 つのフラグメ

ントの間がリンクされた状態で出力される. これらは全ての候補構造のリンクが終了するまで一時的に保管した後, まとめてドッキングスコアを計算する. さらに, この時点で得られている構造は, 予めリンカーライブラリに登録された立体構造のつなぎ合わせであり, 結合長さや結合角などが最適ではない可能性がある. そのため, 標的タンパク質と出力構造の複合体に対し, 標的タンパク質の全原子を固定した状態で分子力学法 (MM) による局所エネルギー最小化を行い, 分子内歪みを除去する. 最後に, MM 最小化後の構造について再度ドッキングスコアの計算を行う.

2. データセット作成

本報告の主題は, 提案手法に対して新たに複数のタンパク質を標的としたフラグメントデータセットとリンカーライブラリを構築し, 提案手法の再検証を行うことである. ここでは, 標的タンパク質・フラグメント複合体構造とリンカーライブラリそれぞれの収集方法を説明する.

2.1 標的タンパク質・フラグメント複合体構造の収集

2.1.1 フラグメント結合ポーズの決定方法

標的タンパク質に結合する 2 つのフラグメントを元に, 提案手法でリンキングを行い, 生成された分子のドッキングスコアを評価する. しかし, フラグメント単体で標的タンパク質に結合した構造が実験的に決定されている例は少なく, 2 つのフラグメントが同時に結合している構造を入手することも極めて難しい. そこで本研究では, 標的タンパク質と大型のリガンドが結合した状態の実験的に決定された立体構造を用意し, 当該リガンドをフラグメントに分割することで擬似的に標的タンパク質と 2 つのフラグメントの結合ポーズを作成, これを例題として利用してリンキングの評価実験を行う.

2.1.2 データセットの規模

本研究は提案手法が実際に複数の標的タンパク質について, *in silico* での分子設計ができることを示すことまでが目的である. また, 提案手法が有する様々なパラメータの影響を確認し, 実用的な動作設定を確認する必要がある. そこで, 実験用データセットとしては, 十分な確認作業を可能とするために異なる標的の少数の構造を用いることにする. ここでは, 5 つの標的タンパク質からなるデータセットを作成した.

2.1.3 データ収集方法

まず, 標的タンパク質-大型リガンド複合体構造の収集について説明する. データ収集元として, 実験的に結合親和性が決定されているリガンドとタンパク質の複合体構造を収録したデータベースである PDBbind [11] を採用した. 2021 年 9 月時点で最新版であった PDBbind v2020 を取得し, そのうち, 一定の基準を満たした構造が収められている refined set を使用した. 次に, ペプチド, 核酸, 糖類,

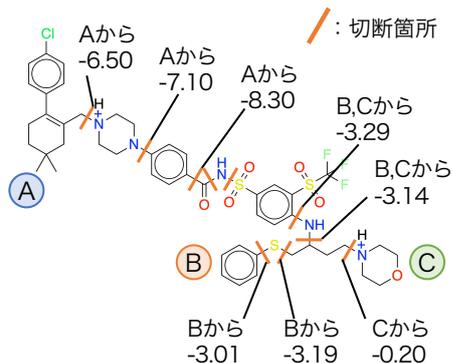


図 3 大型リガンドのフラグメント分割とスコアリング

ピロリン酸を含む化合物をリガンドとする複合体を除去した。次に、リガンドの重原子数の降順にソートし、それぞれの標的タンパク質ごとに最も重原子数が多い構造を抽出する。最後に、抽出された構造について、リガンドが複数のタンパク質分子に近接している構造、タンパク質の表面以外に結合している構造、リガンド周辺に金属原子を含む構造を排除し、その上で重原子数が多い構造を5件採用した。この5件については、CCG MOEのQuickPrep機能をデフォルト設定で実施し、中性条件での水素原子付加、電荷設定とリガンド周辺の分子力学法によるエネルギー最小化計算を行った。

次に、収集した「標的タンパク質-大型リガンド複合体構造」から「標的タンパク質-2フラグメント結合構造」を作成する方法を説明する。

手順 1 BRICS によるフラグメント分割

PDB登録構造のリガンドを、フラグメント分割手法の一つであるBRICS [12]により、図3のように分割する。BRICSの実装は、Pythonで利用できる計算化学ライブラリであるRDKitによる実装を使用した。

手順 2 Vina ドッキングスコアの計算

分割したフラグメントを、元の化合物で端にあたる部分から1つずつ接続していき、それらの構造のドッキングスコアを全て計算する。この作業は、分割したフラグメントのうち全ての端点から開始する。図3の例では、A, B, Cそれぞれから行うが、ある程度接続するとBとCの両方に繋がる構造も発生する。ドッキングスコアはAutoDock Vinaのスコア関数を使用する。

手順 3 フラグメントとして利用する箇所の決定

手順2でドッキングスコアを計算した構造のうち、フラグメントのドッキングスコアが最悪でも-3 kcal/mol確保でき、なおかつ10 Å程度のフラグメント間距離が確保できるように両側のフラグメント部位を決定し、残りの部分を破棄する。なお、フラグメント間距離は、リンカーを接続して繋ぐべき距離であり、両フラグメントのBRICS切断点間の距離として定義する。

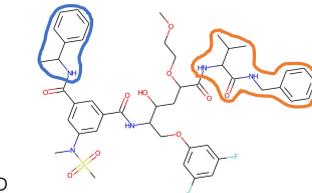
表 2 作成した標的データセット一覧

PDB ID	標的タンパク質	分子量	活性値
1NNY	タンパク質チロシンホスファターゼ 1B	720	22 nM (K _i)
3I25	ベータセクレターゼ 1	896	3.1 nM (K _i)
3QKD	Bcl-xL	839	4.2 nM (K _d)
4LVT	Bcl-2	977	0.04 nM (K _i)
5IZF	cAMP 依存型タンパク質キナーゼ	800	0.18 nM (K _d)

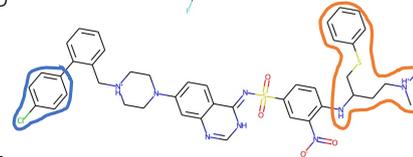
標的1NNY



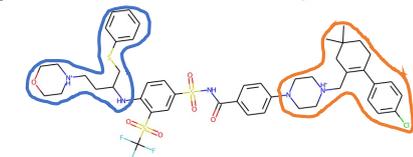
標的3I25



標的3QKD



標的4LVT



標的5IZF

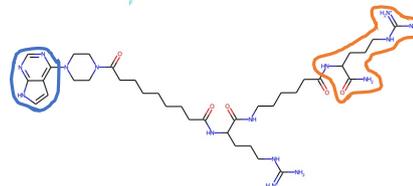


図 4 各構造のリガンドと分割したフラグメント (囲み部分)

表 3 フラグメントの Vina スコアとフラグメント間距離

PDB ID	Vina スコア (kcal/mol)		フラグメント間距離 (Å)
	フラグメント 1	フラグメント 2	
1NNY	-4.75	-5.37	11.43
3I25	-4.26	-4.15	10.00
3QKD	-4.37	-3.48	15.39
4LVT	-3.28	-6.87	10.10
5IZF	-4.65	-3.39	16.00

2.1.4 作成したデータセット

表2に、作成した5件のデータセットのPDB ID、標的タンパク質名、分子量、実験活性値を示す。活性値はPDBbindに登録されている値を使用する。図4に、各構造のPDB登録リガンドの構造式と分割したフラグメント部分を示す。表3に、決定したフラグメントのVinaドッキングスコア、フラグメント間距離を示す。

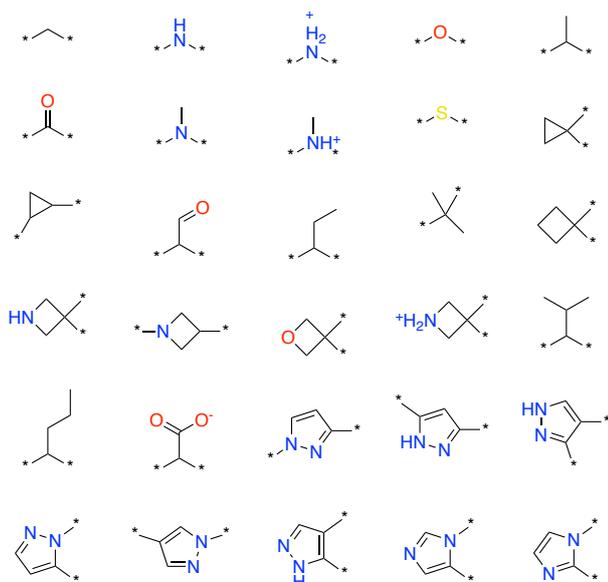


図 5 作成した 30 種類のリンカー要素。*は AP を示す。

2.2 リンカー構成要素の収集

本研究の提案手法は、基本的な化学構造からなる少数のリンカー構成要素を使用する。計算資源や実行時間を考慮し、ここでは、分子量が小さい 30 種類の化学構造からなるライブラリを構築する。

データ収集元として、キシダ化学株式会社が提供するビルディングブロックの化学構造を利用した。同社のビルディングブロックには 5,735 個の化合物が含まれる。次に、収集した化合物をさらに小型の基本的な化学構造に分解するため、BRICS によりフラグメント化した。得られたフラグメントのうち重複する構造を削除した上で、2つ以上の切断点をもつ構造に絞り込んだ。この切断点を AP として使用する。最後に、フラグメントを分子量の昇順に並び替え、先頭から 1 つずつ、予め決定した要素数（本研究では 30 種類）に達するまでリンカーライブラリに加えた。ただし、加えようとしているリンカー要素がすでにライブラリ内にあるリンカー要素の組み合わせで表現可能な場合は追加は行わずに破棄した。

生成したリンカー要素の一覧を図 5 に示す。30 種類のリンカーを、分子量の昇順に表示している。

3. 実験

3.1 実験 1：実行時パラメータの検討

本提案手法にはビームサーチ幅など複数の実行時パラメータが存在するが、それらの設定次第で、リンキングが成功するかや、出力される構造の特徴、実行所要時間が変化する。そこで、実験 1 として実行時のパラメータ変化と出力の関係について調べる。

3.1.1 本提案手法のパラメータとその特性

本提案手法の探索に関するパラメータは表 4 に示す 4 種

表 4 提案手法の探索に関するパラメータ一覧

パラメータ名	内容
副探索範囲	2つの副探索軸まわりの回転探索を許容範囲(±°)
ビームサーチ幅	あるステップの出力構造のうち、評価値上位何件を次のステップに与えるか
主探索刻み	主探索軸まわりの回転探索の刻み幅(°)
副探索刻み	2つの副探索軸まわりの回転探索の刻み幅(°)

表 5 実験 1 で使用するパラメータ

パラメータ名	設定値
副探索範囲	5, 10, 15, 20*, 25*
ビームサーチ幅	30, 45, 60
主探索刻み	10
副探索刻み	5

*はビームサーチ幅 30 の場合のみ

類である。この中で、特に重要なのは副探索範囲とビームサーチ幅である。副探索範囲の設定次第で、フラグメントとリンカー要素の間またはリンカー要素同士の間の化学結合の結合角の許容範囲が変わることにより、化学的な立体構造の探索が制約される。また、ビームサーチ幅は、木探索の近似解法であるビームサーチの性能を左右するものであり、この設定次第で探索される計算空間が制約される。

3.1.2 探索するパラメータと実験設定

この実験では、化学的に重要なパラメータである副探索範囲と、計算機科学的な探索上重要なパラメータであるビームサーチ幅が、それぞれ出力構造に与える影響を調査する。そのため、表 5 に示すパラメータを設定して観察する。ビームサーチ幅 30 の場合は副探索範囲 5~25 の 5 通り、ビームサーチ幅 45 と 60 の場合は 5~15 の各 3 通りの合計 13 通りの組み合わせである。

3.1.3 評価方法

提案手法の実際のユースケースとして、出力構造のうちスコアの良い上位何件かを検討材料として分子設計することが考えられる。そのため、ドッキングスコアの良い順に並べて a 番目の構造のドッキングスコアを Top- a スコアと定義する。本提案手法では数十~数百程度の出力構造が得られる場合が多い。この実験では、上位 10 構造に着目して Top-10 スコアを評価に用いる。

3.2 実験 2：中間構造の評価関数の検討

提案手法の説明の中で表 1 に示したように、本研究では、出力構造のドッキングスコアだけでなく、分子量やリンク残距離を考慮するために複数の評価関数を設計した。実験 2 ではこれらの評価関数ごとに得られる最終構造を比較し、各評価関数が意図通りの効果を発揮するか否かを検証し、また評価関数による出力構造への影響を確認する。

3.2.1 実験設定

この実験では、表 1 に示す 6 通りの中間構造の評価関数

表 6 実験 2 で使用するパラメータ

パラメータ	設定値
副探索範囲	10
ビームサーチ幅	60
主探索刻み	10
副探索刻み	5

を使用し、評価関数ごとに結果を集計する。実験 1 の結果を踏まえ、表 6 に示すパラメータで実験を行う。

3.2.2 評価方法

実験 1 と同様に Top-10 スコアを用いて評価する。また、出力される分子の特徴を捉えるため、分子量とドッキングスコアの分布の観察を行う。

3.3 提案手法の実装

入力ファイルの前処理、リンカーの回転、接続操作および分子力学法計算は、CCG 社の統合化学計算ソフトウェア MOE のスクリプト言語環境 SVL により実装し、バージョン 2019.0102 にて実行した。AutoDock Vina のスコア関数は、バージョン 1.1.2 に基づいて久保田ら [13] により再実装されたものを使用した。その他の計算は Python で実装し、バージョン 3.7.7 にて実行した。また、スコアリングについては、予備実験から計算時間の大部分を占めることがわかってきたため、並列処理ライブラリ Joblib を利用して、データ並列によるノード内並列実装を行った。

3.4 実験環境

実験 1, 2 とともに、TSUBAME3.0 を利用して計算を実施した。計算は全て、1 ノードを占有利用する資源タイプ f_node を 1 つ利用した。当該設定では、物理 28 コアの CPU, 235 GB の DRAM, 2 TB の NVMe 接続の SSD を利用できる。

4. 結果

4.1 実験 1：実行時パラメータの検討

出力構造のうち Vina スコア上位から 10 番目の構造のスコアである Top-10 スコアを表 7 に示す。ただし、出力構造数が 10 件に満たない場合は算定不能のため NA と表記し、各標的ごとに最もスコアが良い箇所を赤色で示している。この結果より、5 標的中 3 標的で、ビームサーチ幅 60, 副探索範囲 10 の設定で最良であり、また標的 5IZF でも 2 番目に良かった。そのため、実験 2 はこの設定を利用することとした。

4.2 実験 2：中間構造の評価関数の検討

出力の Top-10 スコアを表 8 に示す。この結果より、評価関数 0 (Vina スコア) または評価関数 2 (リガンド効率 (2)) の場合に Top-10 スコアが良かった。また、Vina スコアと分子量の分布の代表として、標的 4LVT の結果を図 6

表 7 実験 1 の Top-10 スコア (kcal/mol)。出力構造数が 10 件に満たない場合は NA と表記。標的ごとに最良の箇所を赤色・下線付きで表記。

ビームサーチ幅	副探索範囲	1NNY	3I25	3QKD	4LVT	5IZF
30	5	NA	-9.24	-8.40	-8.73	NA
	10	-9.72	-9.61	-8.91	-8.62	-9.59
	15	-9.75	NA	-8.90	-8.58	-9.31
	20	-10.00	NA	-8.85	-8.62	-9.30
45	25	-10.22	NA	<u>-8.94</u>	-8.53	-9.36
	5	-8.47	-9.36	-8.88	-8.78	-9.78
	10	-10.07	-9.62	-8.91	-8.82	-10.02
60	15	-9.80	-9.18	-8.89	-8.69	-9.09
	5	-8.85	-9.27	-8.88	-8.77	-9.84
	10	<u>-10.23</u>	<u>-9.69</u>	-8.81	<u>-8.88</u>	-10.14
	15	-9.80	-9.64	-8.90	-8.74	<u>-10.28</u>

表 8 ビームサーチ幅 60 のときの Top-10 スコア (kcal/mol)。出力構造数が 10 件に満たない場合は NA と表記。標的ごとに最良の箇所を赤色・下線付きで表記。

評価関数	1NNY	3I25	3QKD	4LVT	5IZF
0	<u>-10.23</u>	-9.69	-8.81	-8.88	<u>-10.14</u>
1	-8.94	-9.41	NA	-7.20	NA
2	-9.72	<u>-9.78</u>	<u>-9.06</u>	<u>-8.91</u>	-9.97
3	-9.10	-8.62	-8.66	-8.42	-8.38
4	-9.40	-8.82	-8.03	-7.41	NA
5	-8.31	-9.19	25.99	-7.08	-8.10
PDB 登録 リガンド	-9.07	-10.86	-10.40	-10.32	-7.06

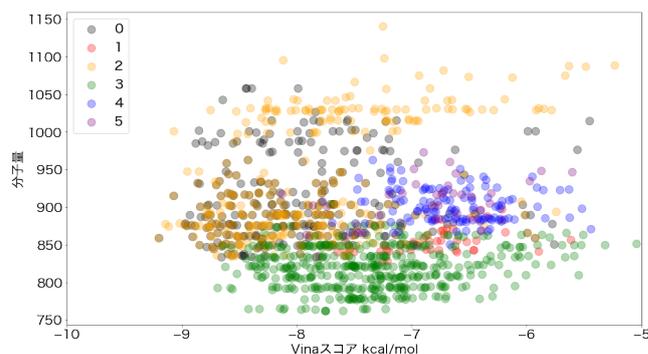


図 6 評価関数 0~5 により得られた出力構造の Vina スコアと分子量の散布図 (標的 4LVT)

に示す。

5. 考察

図 6 の分布から、評価関数 0 (Vina スコア)、評価関数 2 (リガンド効率 (2)) は図の左寄りに分布しており、Vina スコアが良い傾向がある。一方、評価関数 1 (リガンド効率 (1)) や評価関数 3 (距離効率) は図の下側に分布し、分子量が小さい構造を出力する傾向にある。紙面の都合上掲

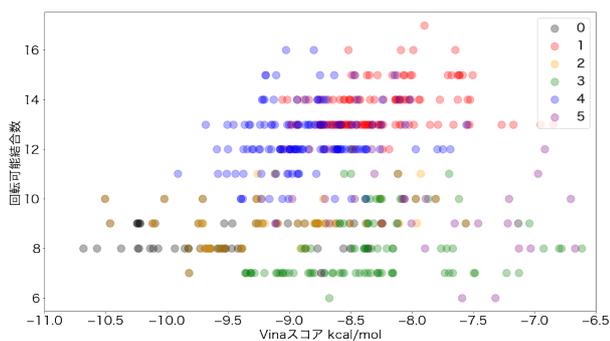


図 7 評価関数 0~5 により得られた出力構造の Vina スコアと回転可能結合数 (標的 1NNY)

載は省略するが、その他 4 標的の実験でも同様の傾向がみられた。また、評価関数 0 では Vina スコアが良く、評価関数 3 では分子量が小さくなる傾向は前回の研究会報告で得た結論とも一致する。評価関数 1 と 4 では、出力構造が得られない場合があったが、これは 1 原子や 2 原子からなるリンカーが選ばれやすくなるために 10 ステップではリンクングを完了できなかったことによる。これらの結果を踏まえると、出力構造に求める特性に応じて適切な評価関数に切り替えることが望ましいと考えられる。

提案手法は、単純なリンカー要素を用いる既存手法に比べて柔軟性の低い構造を出力できることを目指したことから、出力構造の柔軟性について検討する。図 7 に、実験 2 で得られた出力構造の Vina スコアと回転可能結合数の分布をの代表として標的 1NNY の例を示す。評価関数 1 (リガンド効率 (1)) の赤色や評価関数 4 (差分リガンド効率) の青色の点は散布図の上側に分布しており、回転可能結合数が多いことがわかる。これらの関数では特に小型のリンカー要素が選ばれやすく、1 原子や 2 原子からなるリンカー要素が多用される傾向にあることが原因である。また、評価関数 3 (距離効率) の緑色の点は全ての図で下側に分布し、回転可能結合数が少ない。これは、距離効率を使用した場合、環などの大型のリンカーが選択されやすく、また使用されるリンカー要素数自体も少なくなることで原因と考えられる。

6. まとめ

6.1 本研究の結論

長距離への対応が難しいという Fragment linking の課題を解決するべく第 67 回バイオ情報学研究会で報告した提案手法について、新たに 5 種類の標的タンパク質-フラグメント複合体と 30 種類のリンカー要素を収集し、改めて実行時パラメータの検証設定と独自評価関数の効果を確認した。実行時パラメータ探索では、異なる標的を用いた評価により以前の研究で使用したよりも優れたパラメータを発見した。評価関数による出力構造への影響の検討では、評価関数 0 (Vina スコア) や評価関数 2 (リガンド効率 (2))

を用いると出力構造の Vina スコアが良くなる傾向があり、評価関数 3 (距離効率) を用いると出力構造の分子量が小さくなるだけでなく、回転可能結合数が減少し硬い構造が得やすくなることもわかり、出力構造に期待する性質によって評価関数を適切に選択する必要があると考えられる。

6.2 今後の課題

6.2.1 評価関数のさらなる改良

本研究で比較した関数は Vina スコアのほか、分子量、リンク残距離を考慮していた。回転可能結合数は実験後の考察において使用したが、これも評価関数に組み込むことが考えられる。また、例えば Vina スコアを分子量とリンク残距離の積で除算するなど、複数の要素を組み合わせた評価関数を試すことで、複数の目的を同時に達成できる可能性がある。

6.2.2 リンカーライブラリの改良

公知の化合物ライブラリを元にフラグメント分割を行い、分子量が小さい順に選択することでリンカーライブラリを構築したが、この方法では六員環や縮合環のような、重要と思われる部分構造を利用できなかった。また、ピラゾールやイミダゾールのような複素環では AP の位置関係の違いによって多くの構造がライブラリ内に存在するようになる。このような問題があるため、既知の薬剤の傾向などの情報に基づいてリンカー要素を選抜したり、ヘテロ原子数を制限したりすることでリンカーライブラリを改良できる可能性がある。

謝辞 本研究の一部は、JSPS 科研費 (20H04280, 20K19917)、文部科学省地域イノベーション・エコシステム形成プログラム、AMED 創薬等先端技術支援基盤プラットフォーム (BINDS) の支援を受けて行われた。本研究は、東京工業大学のスーパーコンピュータ TSUBAME3.0 を利用して実施した。

参考文献

- [1] Bian, Y. and Xie, X. Q. S.: Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications, *AAPS Journal*, Vol. 20, No. 3, 59 (2018).
- [2] Schade, M. *et al.*: Highly Selective Sub-Nanomolar Cathepsin S Inhibitors by Merging Fragment Binders with Nitrile Inhibitors, *Journal of Medicinal Chemistry*, Vol. 63, No. 20, pp. 11801–11808 (2020).
- [3] Simon, S. C. *et al.*: Discovery of WD Repeat-Containing Protein 5 (WDR5)-MYC Inhibitors Using Fragment-Based Methods and Structure-Based Design, *Journal of Medicinal Chemistry*, Vol. 63, No. 8, pp. 4315–4333 (2020).
- [4] Ichihara, O. *et al.*: Compound Design by Fragment-Linking, *Molecular Informatics*, Vol. 30, No. 4, pp. 298–306 (2011).
- [5] Bancet, A. *et al.*: Fragment Linking Strategies for Structure-Based Drug Design, *Journal of Medicinal Chemistry*, Vol. 63, No. 20, pp. 11420–11435 (2020).

- [6] Möbitz, H. *et al.*: Discovery of Potent, Selective, and Structurally Novel Dot1L Inhibitors by a Fragment Linking Approach, *ACS Medicinal Chemistry Letters*, Vol. 8, No. 3, pp. 338–343 (2017).
- [7] Rossi Sebastiano, M. *et al.*: Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs beyond the Rule of 5, *Journal of Medicinal Chemistry*, Vol. 61, No. 9, pp. 4189–4202 (2018).
- [8] 津嶋佑旗ほか: タンパク質表面との結合親和性を考慮した長距離フラグメントリンキング手法の開発, 情報処理学会研究報告, Vol. 2020-BIO-67, No. 7, pp. 1–8 (2021).
- [9] Trott, O. and Olson, A. J.: AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *Journal of Computational Chemistry*, Vol. 31, No. 2, pp. 455–461 (2009).
- [10] Hopkins, A. L. *et al.*: Ligand efficiency: a useful metric for lead selection, *Drug Discovery Today*, Vol. 9, No. 10, pp. 430–431 (2004).
- [11] Wang, R. *et al.*: The PDBbind Database: Collection of Binding Affinities for Protein – Ligand Complexes with Known Three-Dimensional Structures, *Journal of Medicinal Chemistry*, Vol. 47, No. 12, pp. 2977–2980 (2004).
- [12] Degen, J. *et al.*: On the Art of Compiling and Using ‘Drug-Like’ Chemical Fragment Spaces, *ChemMedChem*, Vol. 3, No. 10, pp. 1503–1507 (2008).
- [13] 久保田陸人ほか: 共通な部分構造の再利用による高速なタンパク質リガンドドッキング手法の開発, 情報処理学会研究報告, Vol. 2020-BIO-61, No. 4, pp. 1–8 (2020).