

学位論文（博士）

SBDD シミュレーションと実験値の結合親和性の相
関性ならびに新規 AI 技術の性能評価による計算化学
の医薬品候補化合物設計への有用性研究

2022 年 9 月

星薬科大学大学院 薬学研究科

総合薬科学専攻

薬品物理化学

半田 千彰

目次

論文リスト	．．．	3
略号リスト	．．．	4
本文		
1. 緒言	．．．	5
2. エストロゲン β 受容体とジアリールプロピオニトリル (DPN) および DPN 誘導体の等温滴定カロリメトリーとフラグメント分子軌道法による結合親和力の相関評価	．．．	8
2.1 研究背景	．．．	8
2.2 結果と考察	．．．	12
2.3 小括	．．．	23
3. 化合物の Ames 予測における Graph Convolutional Network の特徴評価	．．．	25
3.1. 研究背景	．．．	25
3.2. 結果と考察	．．．	27
3.3 小括	．．．	33
4. 結語	．．．	34
謝辞	．．．	35

引用文献	．．． 36
実験項	．．． 42
2章の実験項	．．． 42
3章の実験項	．．． 45
実験データ集	．．． 52

論文リスト

1. Chiaki Handa, Yuki Yamazaki, Shigeru Yonekubo, Noritaka Furuya, Takaki Momose, Tomonaga Ozawa, Takayuki Furuishi, Kaori Fukuzawa, Etsuo Yonemochi, Evaluating the correlation of binding affinities between isothermal titration calorimetry and fragment molecular orbital method of estrogen receptor beta with diarylpropionitrile (DPN) or DPN derivatives., The Journal of Steroid Biochemistry and Molecular Biology, Volume 222, September 2022, 106152 【2章】

<https://doi.org/10.1016/j.jsbmb.2022.106152>

2. 半田 千彰、小沢 知永、福澤 薫、米持 悦生、化合物の Ames 予測における Graph Convolutional Network の特徴評価、Journal of Computer Chemistry, Japan、20 卷 1 号 p. 1-9, (2021) 【3章】

<https://doi.org/10.2477/jccj.2020-0015>

略号リスト

本論文では、以下の略号を用いた。

AI	人工知能; Artificial Intelligence
CADD	Computer aided drug design
DPN	Diarylpropionitrile (試薬名)
EC	Enzyme commission number (酵素の体系的なナンバリング)
ER	エストロゲン受容体; Estrogen receptor
FMO	フラグメント分子軌道法; Fragment molecular orbital method
GCN	グラフ畳み込みネットワーク; Graph convolution network
GPU	Graphics processing unit (特定の計算を高速に行うハードウェア)
IFIE	フラグメント間の相互作用エネルギー; Inter Fragment Interaction Energy
ITC	等温滴定カロリメトリー; Isothermal titration calorimeter
LBD	リガンド結合領域; Ligand-binding domain
LBDD	Ligand-based drug design
PIEDA	Pair interaction energy decomposition analysis
QSAR	定量的構造活性相関; Quantitative structure-activity relationship
ROC	Receiver operatorating characteristic
ROC-AUC	ROC 曲線より下の面積。
SBDD	Structure-based drug design

1 緒言

医薬品の創製（創薬）は、非常に長いプロセスであり、一医薬品あたり 10～15 年程の期間を必要とし¹⁾、医薬品が市場に出るまでに約 26 億ドルのコストがかかっている²⁾。創薬研究では、適切な薬物ターゲット候補の特定から始まり、薬物ターゲットとしての薬理・薬効の妥当性検証、ヒット化合物からリード化合物の発見、リード化合物の最適化、前臨床試験や臨床試験といった多段階の研究が行われている³⁾。このように、新薬の創製には多大な投資と時間を費やしているにもかかわらず、臨床試験での成功率はわずか 13%である⁴⁾。このような環境下、製薬企業や研究グループによる検証においてコンピュータ支援による創薬（Computer aided drug design; CADD）は、創薬・開発プロセスの迅速化に有効であり、非臨床研究において CADD への期待は非常に大きい。特に創薬化学におけるコンピュータの利用は、ヒット化合物からリード化合物の発見、リード化合物の最適化の両場面での実績が多く、CADD の中でも計算化学という大きな一分野となっている⁵⁾。

計算化学が担う大きな役割の一つは、シミュレーションによる合理的薬物設計である。人間の目では確認できない、標的タンパク質とリガンド間の結合状態、その親和性、親和性の基となっている分子間相互作用をコンピュータによって評価・解析することである。このシミュレーション結果を人間が洞察し実際に合成する化合物を設計・選別し、より少ない合成化合物数で、臨床化合物を創製することに貢献している。近年では、コンピュータの処理性能の進展が目覚ましく、「富岳」⁶⁾をはじめとするスーパーコンピュータリソースや、並列処理、高度なプログラム、アルゴリズム、GPU（Graphics Processing Unit）の利用等により、より一層、創薬研究におけるリードの同定に活用されている⁷⁾。さらに、近年では、機械学習を代表とする人工知能（AI; Artificial Intelligence）技術の進歩が、創薬プロセスにおける医薬関連のビッグデータの分析、学習、説明に大きく役立っており、計算化学のもう一つの大きな役割となっている⁸⁾。

化学構造のデータベースからリード分子を同定するために採用されているコンピュー

手法には、ファーマコフォアモデリング、定量的構造活性相関 (QSAR)、分子ドッキング、量子化学、統計的学習法など多くの手法がある。これらの CADD 手法は、大きく主にタンパク質の立体構造を利用した SBDD (Structure-based drug design) とリガンドの情報を利用した LBDD (Ligand-based drug design) の 2 つのアプローチに分けられる。どちらのアプローチも創薬プロセスにおいて適切なリード分子の同定に広く利用されている。SBDD は、受容体とリガンドの分子間相互作用を理解するために、標的受容体とその活性部位の三次元複合体構造に依存し⁹⁾、LBDD は、ターゲットと相互作用するリガンドの情報に依存する¹⁰⁾。これらの計算化学手法には、多くの成功事例があり、創薬プロセスにおいて重要な役割を果たし続けている¹¹⁾。例えば、最近でも、新型コロナウイルス感染症 (COVID-19) に対する薬剤候補の選定において、SBDD と LBDD を組み合わせた薬剤探索の報告がある¹²⁾。

本研究では、このような現在の創薬環境を踏まえて、SBDD、LBDD、両計算化学アプローチに関する研究を行った。SBDD では、実験で求めたタンパク質と化合物の親和性とコンピュータを用いて計算で求めた親和性の相関についての研究を行った。エストロゲン β 受容体 (ER β) を研究題材として、一般的に利用されている ER β の標品化合物であるジアリールプロピオニトリル (DPN) およびその新規類縁体を合成し、それらと ER β との結合様式を X 線結晶構造解析により初めて明らかにした。また、これまで報告がなされていない DPN と ER β の物理化学的結合能を等温滴定カロリーメトリー (ITC) の測定により明らかにした。コンピュータによる親和性評価は、量子力学に基づくフラグメント分子軌道法 (FMO) によって結合エネルギーを求めて、それらの相関性を確認した。ITC と FMO の親和性の相関を解析した報告は、初めての研究事例である。

もう一方の LBDD では、AI 技術の一つである機械学習を用いた化合物の AMES 予測について研究を行った。AMES 予測は LBDD の分野で古くから研究されている題材であるが、グラフ畳み込みネットワーク (GCN)¹³⁾ というディープラーニング系の新規予測手法を用いて、既存の予測手法と比較してどのような場面において有用な手法であるか研究を行った。GCN と既存予測手法を比較した研究はいくつか存在するが、本研究のよ

うに GCN と既存手法の予測精度に明確に差がある場面を報告した事例はない。

以下、2章にて SBDD の研究について、3章で LBDD の研究について述べ、4章で両計算化学アプローチについて結果をまとめた。

2 エストロゲン β 受容体とジアリールプロピオニトリル (DPN) および DPN 誘導体の等温滴定カロリメトリーとフラグメント分子軌道法による結合親和力の相関評価

2.1 研究背景.

エストロゲン受容体 (ER) は、内因性リガンドにより転写活性化する分子であり、ER α と ER β という 2 つのサブタイプが存在する。これらは、細胞および組織によって異なる分布パターンを示す¹³⁾。両受容体の主な内因性リガンドは共通の 17 β -エストラジオール (E2) であり、E2 が受容体に結合すると受容体は活性化して、下流シグナルの転写が行われる¹⁵⁾。また、ER β アゴニストは、前立腺癌や心血管疾患のような病態の治療に対する可能性が期待されており、安全で有効な薬剤として選択的な ER β アゴニストを開発が試みられている¹⁶⁾¹⁷⁾。

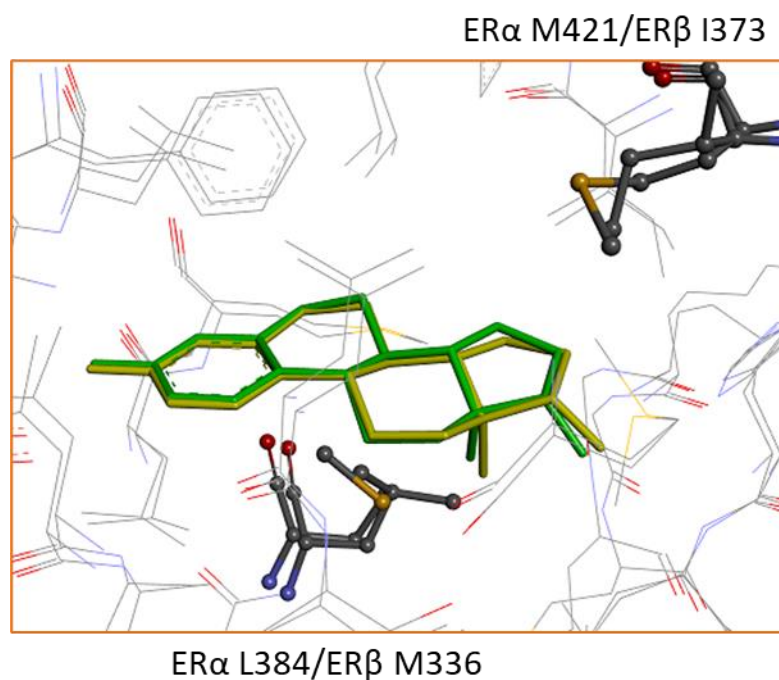


図 1: エストロゲン受容体 α と β のエストラジオール (黄色; ER α 、緑色; ER β) の重ね合わせ図。

ER の両サブタイプとも、DNA 結合とリガンド結合の 2 つのドメインを含んでいる（以降、リガンド結合ドメインを LBD と表記する）。サブタイプ間ではアミノ酸配列の一致度が 90%を超える極度に類似する集団も存在するが、ER β と ER α では配列一致度は約 58%である。このような配列の相同性であるにもかかわらず、リガンド結合ポケットのアミノ酸構成の違いはほとんどないことが知られており、ER α と ER β の LBD の X 線複合体構造から、結合したリガンドの近傍には 2 つのアミノ酸の違いのみがあることが明らかにされている。ER α の Leu384 は ER β の Met336 に、ER α の Met421 は ER β の Ile373 にそれぞれ置き換わっている（図 1）¹⁸⁾¹⁹⁾²⁰⁾。E2 は両方の受容体に対してほぼ同様の親和性を示しており、その原因はリガンド結合ポケットに違いが少ないことが主な原因であると考えられる。

それにも関わらず、現在、いくつかの ER β 選択的なアゴニストリガンドが報告されている。例えば、ERB-041 は Wyeth 社によって開発された選択的な ER β アゴニストの一つであり²¹⁾、LY3201 は Eli Lilly 社によって神経保護作用が報告された高選択的なアゴニストである²²⁾。さらに、DPN (diarylpropionitrile) は、最も広く利用されている ER β 選択的なアゴニストの一つであり、ER β の薬理作用を明らかにするために多くの研究で利用されてきた。DPN は不斉中心を持っているため、多くの研究ではラセミ体が利用されている。しかし現在までに、DPN と ER β の複合体構造は報告されておらず、その結合様式は不明で、どちらのキラル化合物が生物活性エナンチオマー（ユートマー）であるかは明らかでない²³⁾²⁴⁾。

先行研究として、Sun らは、分子モデリングにより、ER β に対して S-DPN がエネルギー的に有利であることを発見し、その相互作用から ER のサブタイプ選択性を説明できることを報告した²⁵⁾。さらに、彼らのグループは、DPN のキラル合成法を開発し、*in vitro* でそれらのリガンド結合親和性を測定したが、その結果では、キラル選択性はわずかであるうえに（3-4 倍）、予想と異なって R-DPN がより親和性が高かった²⁶⁾。この現象について、彼らは、細胞ベースのアッセイにおける受容体結合親和性では、結合力以外の様々な要素が影響するために予想と異なったのではないかと考察している。

より直接的な結合能の測定法として、近年、創薬現場において、タンパク質とリガンドの結合自由エネルギーを等温滴定型カロリメトリー（ITC）で測定し、創薬候補化合物の同定に用いられている²⁷⁾²⁸⁾。ITCでは、反応のカップリングや基質のタグ付けなしにタンパク質-リガンド相互作用の熱力学特性を測定でき、結合自由エネルギーだけではなく、エンタルピー項およびエントロピー項を取り出すことで、リガンドとそのターゲットが結合する時の駆動力も観察できる。創薬候補化合物探索の初期段階では、タンパク質とリガンド間の水素結合などのエンタルピーの寄与が高い相互作用によって親和性を高めるのが一般的である。一方、エントロピーの増大は、非極性表面から水分子が脱溶媒和することがエントロピー的に有利である疎水性効果に基づくと一般的に考えられている。新薬候補の探索の終盤では、化学構造の骨格は定まっておらず新たな水素結合を得ることが困難なため、タンパク質-リガンド親和性は主にエントロピー的な寄与によって増強される。例えば、Bielaらは、トロンビン阻害剤の2骨格において、炭素で構成された置換基を付けるとエントロピー的な寄与が大きくなることを報告している²⁹⁾。このようにITCは、創薬の初期段階での化合物デザインに有用であるが、ER β とDPNの結合親和性をITCで測定した結果はいまだに報告されていない。

これに対して、ERとそのリガンドの結合親和性の理論的な詳細評価が報告されている。その一つがフラグメント分子軌道法（FMO）による相互作用エネルギーの推定であり³⁰⁾³¹⁾³²⁾、例えば福澤らは、FMOで算出した結合エネルギーが、いくつかのER α リガンドの*in vitro*活性と良い相関を示し、さらにリガンドとアミノ酸残基とのフラグメント間相互作用エネルギー（IFIE）が、リガンド結合ポケット周辺の分子間相互作用の評価に有用であることを報告している³³⁾³⁴⁾。FMO法は、タンパク質とそのリガンドとの複合体のような大規模分子系の電子状態を、第一原理分子軌道法に基づいて計算可能とする代表的な方法の一つである。分子軌道法は経験力場を用いた評価方法とは異なり、あらかじめパラメータを設定する必要がないため、タンパク質とリガンドの相互作用エネルギーを電子レベルで理論的に調べることができる。しかし、FMO計算にはタンパク質とリガンドの複合体の正確な原子座標が必要であり、ER β との複合体構造の報告がな

い DPN では FMO による精密な評価を行うことは困難であった。

このように、結合親和性の評価には、理論的アプローチと実験的アプローチそれぞれについて有望な評価方法が存在しており、それらを有効に組み合わせることで、結合性に関する理解がより一層深まると期待できる。実際に Kawasaki らは、ペルオキシソーム増殖剤活性化受容体 α とそのモジュレータの結合親和性を ITC により測定し、FMO 計算を用いてモジュレータと周囲のアミノ酸残基の相互作用を考察している³⁵⁾。また、Drobot らは、ITC と FMO を用いて、カルモジュリンに高い親和性を示す金属イオンの結合機構を調べている³⁶⁾。しかし、両グループとも分子機構の解明を目的とした考察の一部に FMO 計算を用いるに留まっており、ITC と FMO の定量的な相関関係に踏み込んだ例は未だに報告されていない。本研究では、ER β と DPN との結合性の解明と創薬応用を目的として、DPN およびフェノール官能基を塩素原子で修飾した DPN 誘導体のキラル化合物を合成し、それぞれの ER β との共結晶構造について X 線結晶構造解析を実施し、得られた構造に対して FMO 計算によって結合親和性を評価した。また、ITC を用いて ER β との結合親和性を測定し、FMO 結果との相関を解析した。なお、本研究で用いた化合物について図 2 に示す。

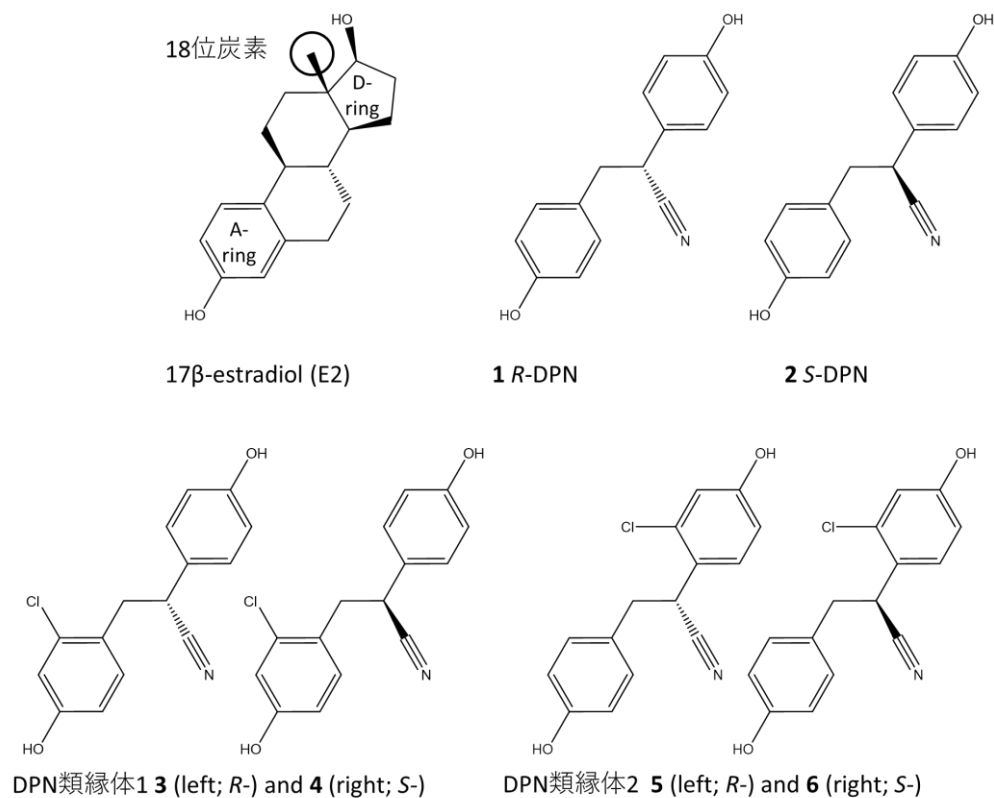


図 2：内因性リガンドエストラジオール、DPN、および DPN 類縁体の化学構造

2.2 結果と考察

2.2.1 X 線複合体構造

X 線結晶解析により、**2** (S-DPN) と DPN 誘導体 (**3-6**) の 5 化合物と ER β との結合様式を明らかにした (図 3-6)。各複合体構造の解像度は、それぞれ 1.54 Å、2.08 Å、1.92 Å、1.89 Å、2.16 Åであった。化合物 **4** のみ同じ空間を占める 4 つのコンフォマーの電子密度が観測され、他の化合物は全て単一の電子密度が観測された。(化合物コンフォメーションを **4-Oc1**, **4-Oc2**, **4-Oc3**, **4-Oc4** と表記する)。いずれの複合体構造

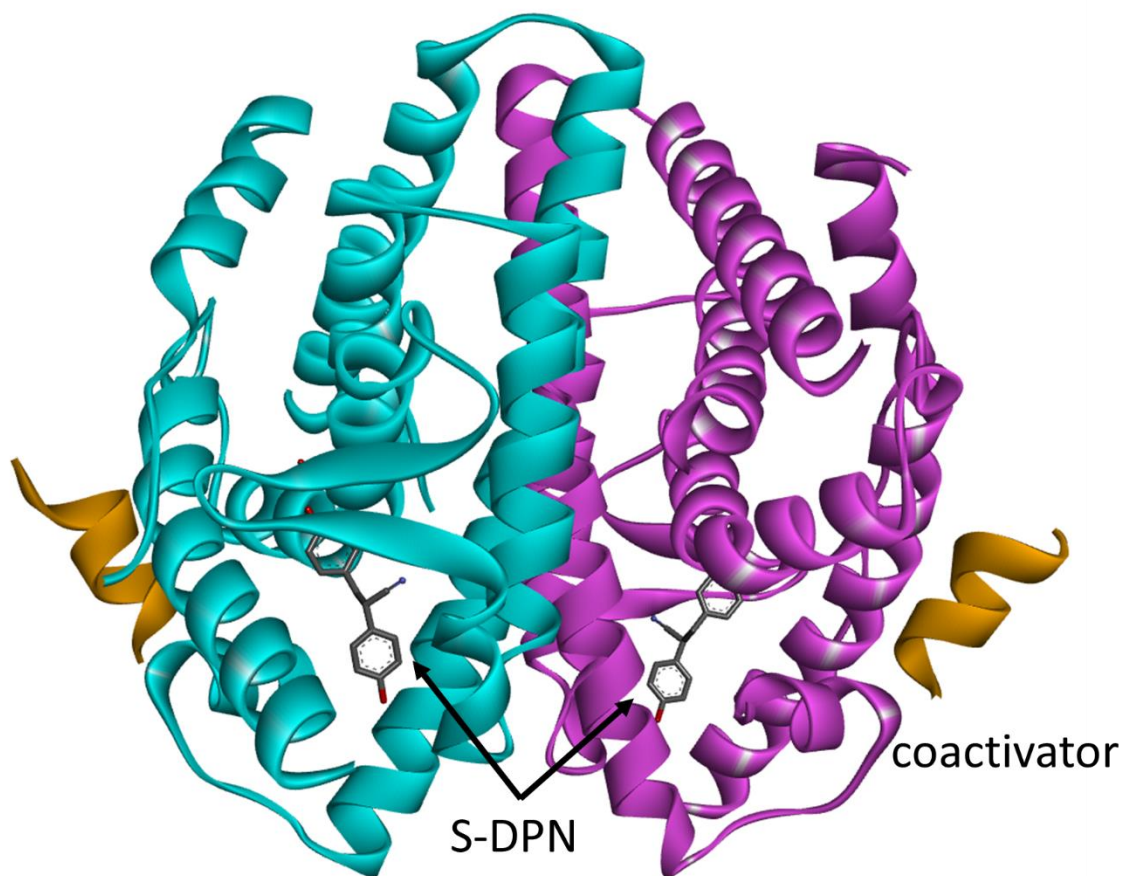


図 3 : ERβ と S-DPN の複合体構造

も受容体タンパク質と補酵素、各化合物を一組とした二量体として観測されたが（図 3）、タンパク質の C α 原子による重ね合わせ後の距離の二乗平均平方根(RMSD)は最大 0.710Å（最小 0.671Å）であり、それぞれのコンフォメーションの差は小さかった。全ての化合物は E2 と同じ結合ポケットに存在していた。E2 の A 環（以下、A 環）と周辺残基との相互作用に水が関与していることは知られているが、多くの既知の X 線結晶構造解析結果と同様に本研究でも同様の水分子が観測された（図 4）。但し、**2** の B 鎖、および、**3** の A 鎖においては、水分子の数がそれぞれ 2 個と 0 個であったが、既知の報告でもまれに水分子の数が異なる例の報告があり、一般的には 1 個の水分子が保存されていると考えられる²²⁾。

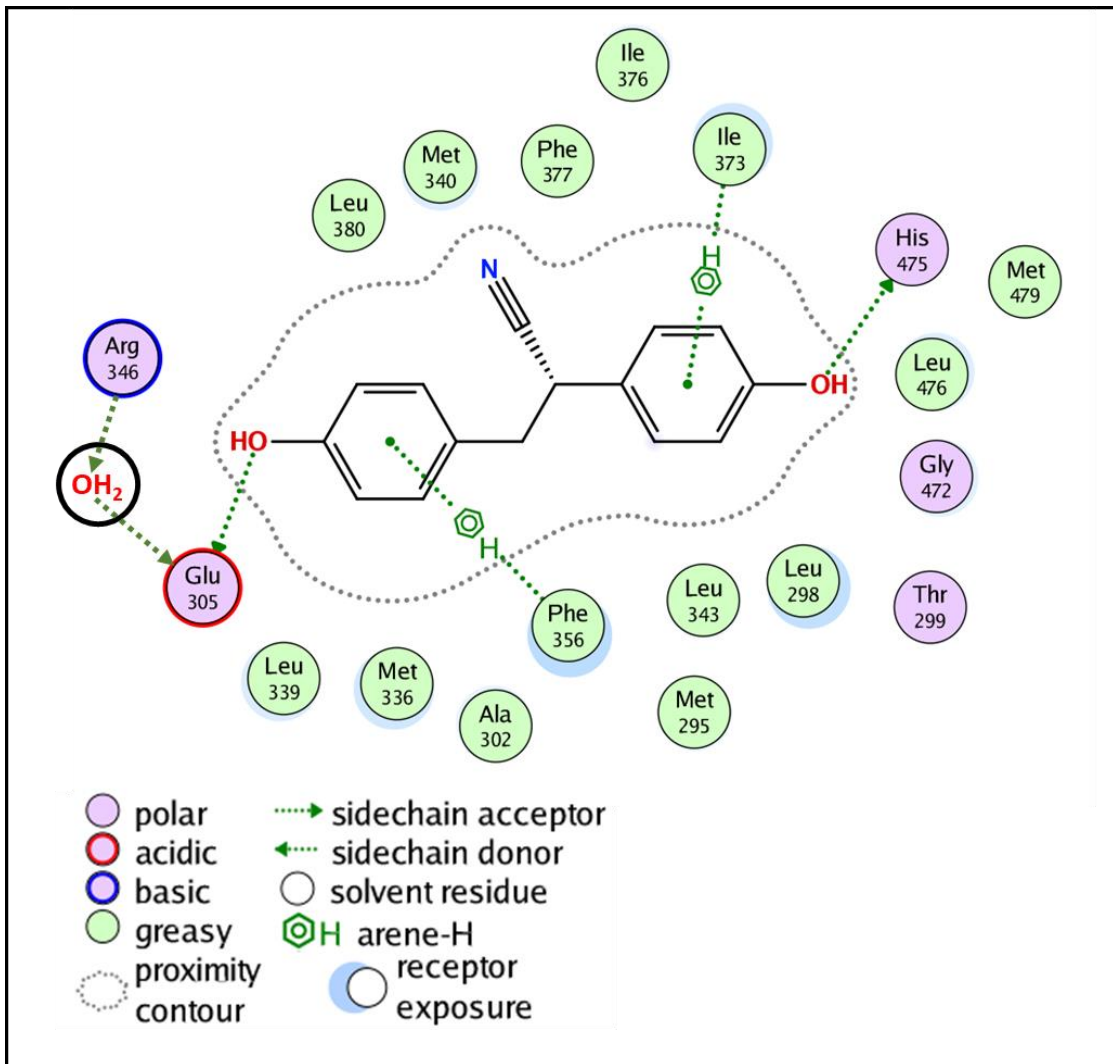


図4：ERβとS-DPNの相互作用様式の概念図（MOEのligand interaction機能による描画）

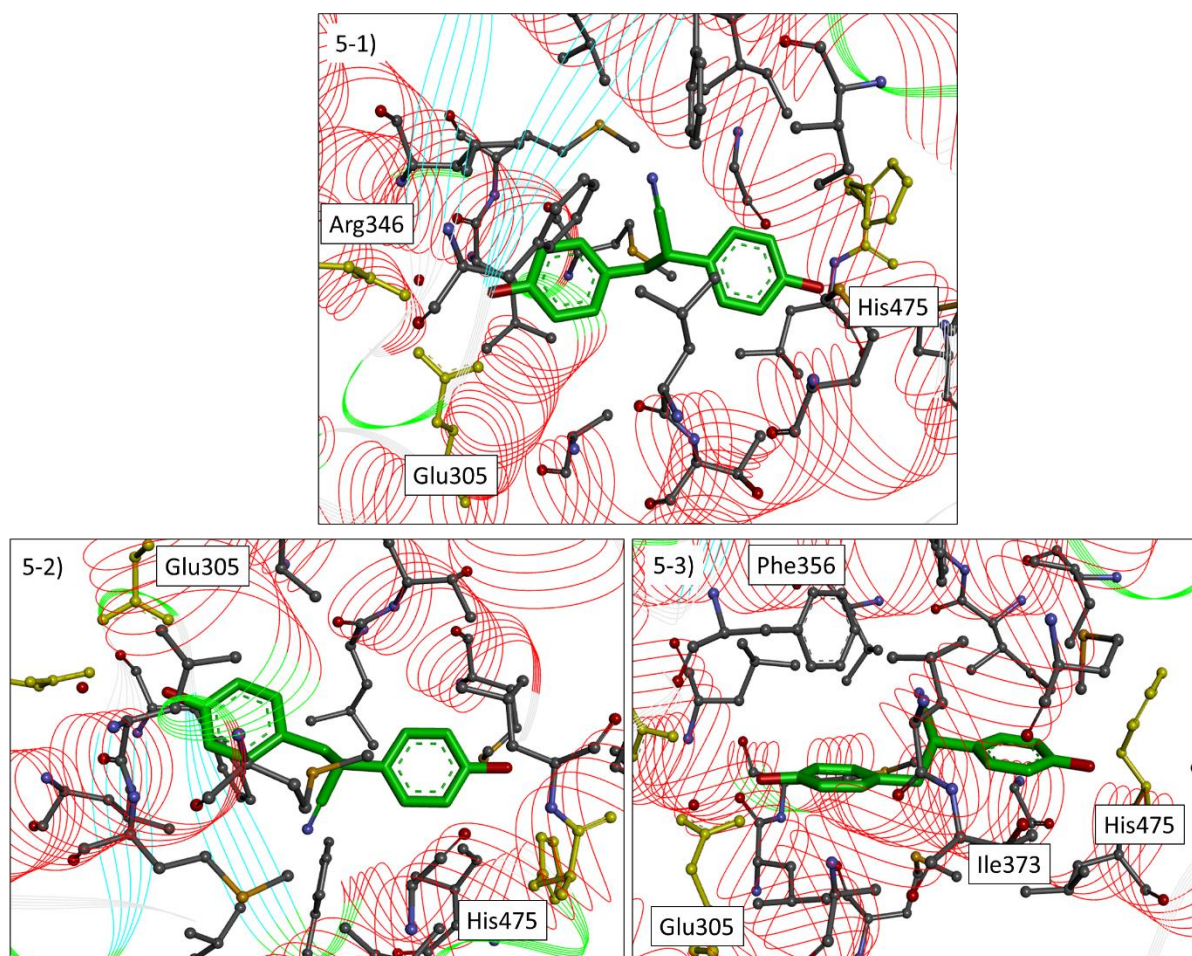


図 5 : ERβ と S-DPN の周辺アミノ酸の図。表中のアミノ酸ラベル Glu305、Phe356、His475、Ile373 については、他のアミノ酸を隠さない限り、そのアミノ酸の種類と残基番号を表示した。5-3 は 5-1 を手前方向に約 90°回転させた図、5-2 は 5-3 をさらに約 90°回転させた図である。

図 2 に示す通り、すべての化合物は 2 つのフェノール基を持っており、そのうちの 1 つのフェノール基が A 環側の位置に位置していた。その水酸基は A 環の親水基と同様に Glu305, Arg346 および水と水素結合ネットワークを構築していた。もう一つのフェノール基は E2 の D 環（以下 D 環）に近い位置にあり、その水酸基は D 環側の親水性アミノ酸である His475 と水素結合していた。さらに、A 環側のフェノール基は Phe356 と CH/π（図では arene-H）相互作用が、D 環側のフェノール基は Ile373 と CH/π 相互作用が観察された。2 つのフェノール基の先のリンカーの位置は以下の 2 パターンであった。

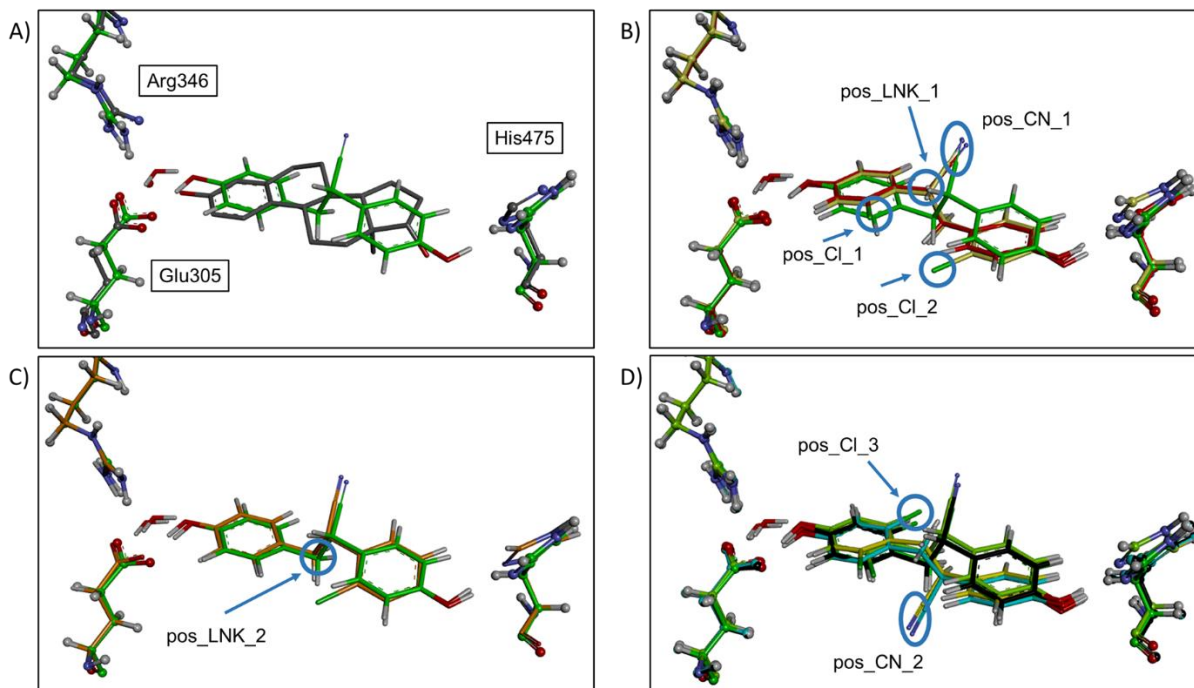


図 6 : ER β のアミノ酸 (Glu305、Arg346、His475) と水分子を重ね合わせ図。化合物 **2-6** の座標は、各 FMO 計算のための水素原子も表示している。A) エストラジオール (原子色 PDB: 5TOA⁴³) と **2** (S-DPN) (緑色の炭素)、B) **2** と **3** (赤色炭素) および **5** (青色炭素)、C) **6** (オレンジ色の炭素) と **2**。D) **4** (黒 (Oc1)、水色 (Oc2)、黄色 (Oc3)、薄緑 (Oc4)) と **2**。

- ・ pos_LNK_1: A 環側のリンカー炭素は、E2 の 18 位炭素方向にあり (ステロイド骨格の A 平面側)、もう一つのリンカー炭素は逆側 (B 平面側) に位置しているパターン。
- ・ pos_LNK_2: 上記と逆に、A 環側のリンカー炭素が B 平面側、D 環側の炭素が A 平面側に位置しているパターン。

R-エナンチオマーである **3** と **5** は、リンカーの位置は pos_LNK_1 であった。S-エナンチオマーである化合物 **2**、**4** (-Oc1、-Oc4)、**6** では pos_LNK_2 が観察された。(図 6-B, C)

シアノ基の位置の電子密度は、R-または S-エナンチオマーに限らず、すべての化合物において近い位置で存在が観察された。その近傍のアミノ酸は、Met340, Phe377, Leu380 の疎水性アミノ酸に囲まれた位置 (図 6-B の pos_CN_1) であった。しかし、**4** の占有リガンドのうち 2 つのシアノ基は異なる位置にあり (図 6-D の pos_CN_2)、

近傍のアミノ酸は pos_CN_1 と同様に疎水性アミノ酸 (Met336、Leu339、Leu476) であった。これらの観察されたリガンドのリンカーは、他の S-体とは異なり、R-体と同じ pos_LNK_1 のパターンを有していた。一方、R-体では (pos_LNK_1 + pos_CN_1) 以外の組み合わせは観察されなかった。

DPN の誘導体における塩素原子の位置は、D 環側に位置する座標 (pos_Cl_2)、A 環側に位置する座標 (pos_Cl_1 または_3) の 3 パターンが観察された。塩素原子の位置については共通の傾向は観察されなかったため、シアノ基の位置が優先され、それによって塩素原子が A-環側にあるか D-環側にあるか決定されると考えた。

以前の Sun らの報告²⁴⁾では、化合物 R-DPN と S-DPN のドッキングシミュレーションにより、R-体と S-体でシアノ基の位置が異なることが示されている。エナンチオマー化合物では、一方が活性を持つユートマーで、もう一方は活性が非常に弱いディストマーであることが多く³⁷⁾³⁸⁾、彼らは ERβ において S-DPN が経験力場による評価でエネルギー的に有利であったことから、S-DPN がユートマーであり、R-DPN がディストマーと結論付けている²⁵⁾。しかし、彼らの予想に反して、R-DPN と S-DPN の細胞での結合親和性の差は非常に小さいことを報告している²⁶⁾。本研究では、残念ながら R-DPN の構造解析は成功しておらず、R-DPN の結合様式は明らかにならなかった。一方、DPN 誘導体の R-エナンチオマーは、塩素原子の位置が異なる以外はほぼ同じ結合様式であった (図 6-B)。この結果は、R-DPN の 3 つの官能基 (2 つのフェノールとシアノ基) も S-DPN と同様の位置にあることを示唆しており、細胞アッセイにおける結合親和性が R-体と S-体で同程度の値を示した理由であると考えられる。

2.2.2 ITC 測定結果

ITC 測定の結果を表 1 に示す。DPN のエナンチオマーである **1** と **2** の結合定数を比較すると、両分子ともほぼ同じ値 (約 1.3 倍) であった。この実験結果は、R-と S-DPN

表 1 : ITC 測定結果と FMO の計算結果、および、IFIE total と各パラメータの相関係数。

compound	binding constant (M ⁻¹)	ITC			FMO IFIE (kcal/mol)				
		ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	total	ES	EX	CT	DI
1	4.09E+06	-9.02 ± 0.1	-28.3 ± 1.5	19.3 ± 1.3	-	-	-	-	-
2	3.09E+06	-8.85 ± 0.1	-25.8 ± 3.1	16.9 ± 3.1	-105.1	-86.4	73.4	-33.2	-58.8
3	1.14E+07	-9.62 ± 0.1	-27.2 ± 3.2	17.6 ± 3.3	-106.3	-91.5	79.8	-35.2	-59.4
4	1.23E+07	-9.67 ± 0.3	-25.9 ± 1.8	16.3 ± 2.0	-104.9	-89.4	79.9	-34.2	-61.1
5	1.08E+07	-9.59 ± 0.1	-24.7 ± 4.4	15.1 ± 4.4	-102.6	-86.2	79.1	-32.8	-62.7
6	1.40E+07	-9.74 ± 0.2	-23.1 ± 1.7	13.4 ± 2.0	-103.2	-91.7	81.6	-33.6	-59.5
correlation ^{*1}	-	-0.265	0.870	-0.862	1.00	0.30	0.28	0.80	-0.59

*1 IFIE total に対する相関係数

の細胞内競合実験において活性に差がなかったことと同様であり、ER β に対する物理化学的結合定数の差が小さいことが細胞における活性差が小さいことの主な理由の 1 つであると考えられる。塩素原子を含む DPN 誘導体の結合定数をそれぞれ R-/S-DPN と比較したところ、**3**、**4**、**5**、**6** の結合定数はいずれの DPN よりも強い値を示し、塩素原子によって結合定数が増強していることが示された。

結合定数から得られるギブス自由エネルギーの内訳を確認すると、すべての化合物でエンタルピー変化 (ΔH) が負の値であり、エンタルピーが結合に有利な寄与をしていることがわかった。一方、エントロピー変化量 ($-T\Delta S$) は、すべての化合物で正の値を示し、結合に不利な寄与をしていることがわかった。また、 ΔH と $-T\Delta S$ の相関は -0.987 であり、非常に強い逆相関があることがわかった。エンタルピー-エントロピー補償により、 ΔH が変化するとそれを補うように ΔS が変化することがよく知られているが、本研究でも同様の現象を確認した。

各エナンチオペアの熱力学的寄与を比較したところ (エナンチオペアは **1** と **2**、**3** と **4**、**5** と **6** となる)、いずれの組み合わせでも R-体 (**1**、**3**、**5**) のエンタルピー変化が S-体 (**2**、**4**、**6**) よりも強い値であった。また、複合体構造の観察から、R-体のリンカー位置はそれぞれ同じ位置 (pos_LNK_1) であるが、R-体の塩素原子の位置は異なっていることがわかっている。また、塩素原子を含まない DPN においても R-体のエンタルピー寄与が S-体より強かったため、その要因は塩素原子の効果ではなく、結合コンフォメーションの違いによるものと考えられる。

表 2： FMO の計算結果における IFIE の上位 15 フラグメント。（単位 kcal/mol）

順位	フラグメント	ES	EX	CT+mix	DI	Total
1	GLU305	-46.5	29.2	-11.9	-5.2	-34.4
2	HIS475	-28.5	18.1	-6.3	-5.7	-22.4
3	THR299	-4.1	0.6	-1.2	-2.2	-7.0
4	PHE356	-2.3	3.3	-1.5	-5.9	-6.4
5	LEU298	-1.4	2.3	-1.0	-4.1	-4.2
6	LEU476	-0.4	3.3	-1.8	-5.2	-4.0
7	LEU339	-0.6	3.1	-1.5	-4.4	-3.4
8	LEU343	-0.8	0.9	-0.8	-2.2	-2.9
9	MET340	0.9	1.3	-1.3	-3.5	-2.6
10	HOH	-1.9	1.8	-1.1	-1.3	-2.4
11	MET295	-1.5	3.3	-0.9	-3.3	-2.3
12	GLY472	-1.3	0.1	-0.2	-0.4	-1.9
13	ALA302	-0.3	1.1	-0.8	-1.7	-1.7
14	MET336	-0.4	1.9	-0.6	-2.6	-1.6
15	ILE373	-0.1	0.7	-0.4	-1.6	-1.5
16	LEU380	-0.7	0.5	-0.3	-1.0	-1.4
17	ILE376	0.4	0.6	-0.4	-1.8	-1.2
18	MET479	0.2	0.3	-0.4	-1.0	-0.9
19	ARG346	0.3	0.2	-0.3	-0.7	-0.5
20	PHE377	1.1	0.2	-0.3	-1.0	0.0
-	その他	1.4	0.6	-0.4	-4.0	-2.4
総計		-86.4	73.4	-33.2	-58.8	-105.1

2.2.3 FMO による S-DPN と ER β の相互作用解析

化合物 **2-6** の FMO 計算結果を表 1 に、**2** の相互作用が強い上位 15 フラグメントを表 2 に示す。ここで、**4** は 4 つのコンフォマーが観測されたが、X 線構造解析ではその占有率を求めることは不可能であるため、4 つの占有リガンドの平均 IFIE を計算し、それを **4** の標準値としている。それらの IFIE およびそのエネルギー成分分割である PIEDA を含む全ての結果は FMO データベース (<https://drugdesign.riken.jp/FMODB/>) に登録済みである (FMODB codes ; 3QJKL、J3Q49、N1NZQ、829GY、GN581)³⁹)。最も強い相互作用を持つ分子は A 環側の Glu305 で、全相互作用エネルギーの約 33% (-34.4 kcal/mol) を占めており、2 番目に強い相互作用は His475 で約 21% (IFIE は -22.4 kcal/mol) 寄与していた。これらのアミノ酸残基は MOE によって S-DPN と水素結合していることが示唆されている (図 4)。両アミノ酸の IFIE のうち分散相互作用 (DI) はそれぞれ -5.2、-5.7 kcal/mol であり、DI 以外の IFIE 成分の相互作用が大部分を占めていた。その寄与率はそれぞれ Glu305 で 85%、His475 で 75%であった。この場合、S-DPN とフェノールの水酸基との水素結合が相互作用の多くを占めていることを示しており、S-DPN と ER β との相互作用には Glu305、His475 とフェノールの水酸基間の水素結合が重要であることが示された (PIEDA 成分の説明については実験項を参照されたい)。

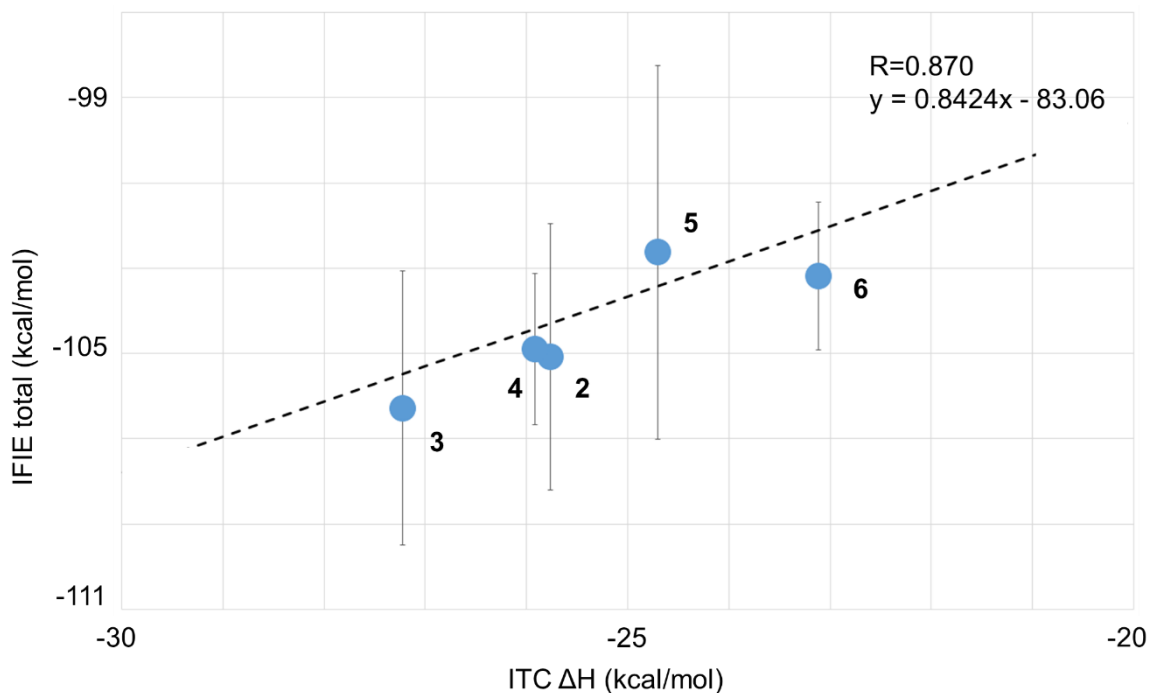


図 7 : ITC 測定結果と FMO の計算結果の散布図。

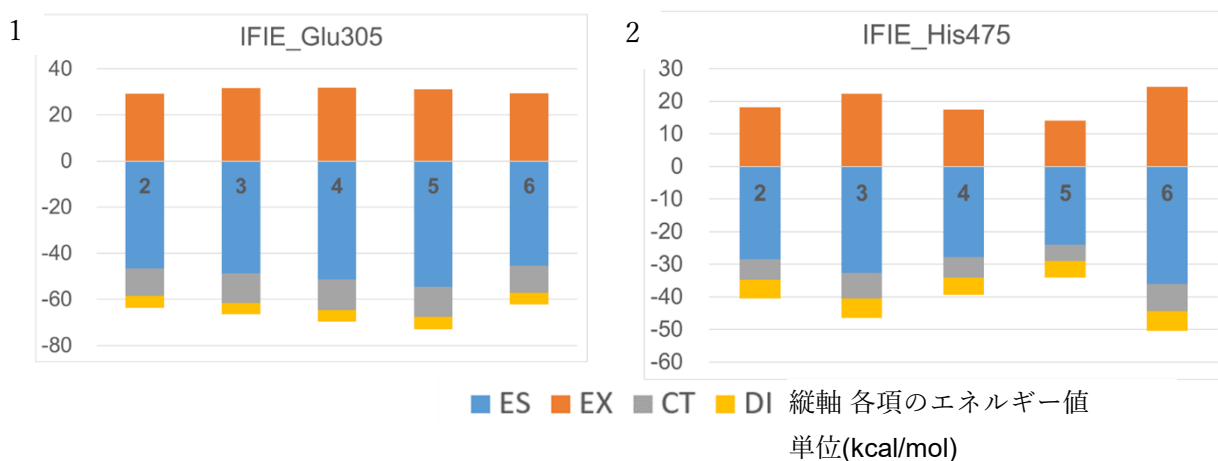


図 8 : 各化合物と Glu305 と His475 間の IFIE。

一方、IFIE DI 項の相互作用が最も強い残基は Phe356 であった (IFIE の DI 項 -5.9 kcal/mol ; IFIE は -6.4 kcal/mol)。したがって、Phe356 は E2 の A 環に相当する DPN のフェノール環と相互作用していることが示された。結晶構造からの予測 (図 4) では、Ile373 も S-DPN と CH/ π 相互作用することが示唆されていた。しかし、Ile373 の S-DPN との相互作用は Phe356 とは対照的に非常に弱かった (IFIE は -1.5 kcal/mol、IFIE の順位は 15 位)。この違いは、構造的特徴のみから相互作用を推測する MOE と

は異なり、量子論に基づく FMO 法によって相互作用が定量的に評価できることを示している。

FMO によって特定された 3 つの重要なアミノ酸 (Glu305、Phe356、His475) は S-DPN の 2 つの水酸基およびフェノール基のフェニル基と相互作用し、これらの相互作用は ER β と S-DPN の相互作用の約 60% を占めていた。Glu305 および Phe356 と主に相互作用をしている E2 の A 環は、男性ホルモンであるテストステロンでは非芳香族であるが、アロマトラーゼ (EC 1.14.14) により芳香環に変換される。すなわち、E2 の A 環は男性ホルモンと女性ホルモンの役割を変える重要な環であり、DPN が女性ホルモンと同様の作用を持つことを支持する結果であった。一方で、DPN の特徴的な部分構造であるシアノ基は選択性に関与していると期待されたが、シアノ基近傍のアミノ酸との IFIE は弱い値であった。近傍には、ER α の Met421 から置き換わった Ile373 も含まれており、シアノ基とサブタイプ選択性の関連は明確にはならなかった。

2.2.4 ITC の親和性と FMO の親和性の相関解析

図 7 に各化合物の FMO 計算の IFIE 合計値と ITC で測定した ΔH との相関係数を示す。また IFIE/PIEDA の各成分を表 1 に示す。結合エネルギーを表す IFIE 合計値は ΔH と強い相関 ($R=0.870$) があり、 ΔG とは相関しない ($R=-0.265$) ことがわかった。 ΔG は ΔH と $-T\Delta S$ の和であり、 ΔS は水分子の脱溶媒など分子の動的な因子である。本研究の FMO は単一座標から計算されており、また温度パラメータ T は 0 である。すなわち、本研究条件の FMO が表現できるのは ΔH のみであるため、 ΔH が IFIE と相関し、 ΔS と ΔG は相関しないことは妥当な結果であるといえる。

FMO で評価可能な化合物の中で、最も安定化が強い ΔH は R-エナンチオマーで D 環側に塩素原子が位置する **3** であり、**3** の IFIE total も最も強い結合安定化を示していた。主要なアミノ酸残基との相互作用では、**3** または **6** と His475 との IFIE が他の化合物より強いことが確認された (図 8-2)。**3** と **6** に共通する点は、D 環側に位置する塩素原

子を含むそのフェノール基が His475 と強く相互作用することであった。一方、Glu305 と 4 または 5 の IFIE は、他の化合物よりも強く、塩素原子が A 環側に位置していた (図 8-1)。PIEDA 解析の結果、すべての塩素原子は IFIE の 静電相互作用 (ES) 項と電荷移動相互作用 (CT) 項を増強し、他の項を増強しないことがわかった。したがって、塩素原子の役割は、フェノール基の水酸基と His475 または Glu305 の間の水素結合を増強していると結論づけられる。

3 と 6 を比較すると、両化合物の ES 項は他の化合物より強いが、3 のトータル IFIE は最も強く、6 は 2 番目に弱いことが判明した。同様に ΔH も 3 が他の化合物より強く、6 が最も弱かった。R-体はそのエナンチオマーペアである S-体よりも ΔH が強く、この傾向は 3 と 6 のトータル IFIE と同様であった。R-エナンチオマーと S-エナンチオマーは複合体構造中において、水素結合している酸素原子の位置はほとんど差がないにも関わらず、FMO の計算ではそれぞれの化合物で異なる IFIE の値を示した。

本研究により、ITC によるエンタルピー変化 (ΔH) は FMO 計算によって求めた IFIE 総計と相関があることが明らかとなった。シミュレーションである FMO 法から実験値のエンタルピー変化寄与を推定することが可能であることが明らかとなった。また、FMO 法の適用範囲には、塩素原子のような力場では化学構造の分極を表現することが困難なハロゲン原子⁴⁰⁾が含まれることも FMO 法の有用性を表している。

2.3 小括

本研究では、S-DPN と ER β の新しい複合体構造を明らかにし、第一原理 FMO 計算により相互作用に重要なアミノ酸を同定した。さらに、ER β と DPN またはその誘導体の結合親和性を ITC で測定または FMO 法で算出した。その相関性を解析した結果、同骨格の化合物群において ITC の ΔH と FMO によるトータル IFIE には非常に強い相関があることを明らかにした。このような、ITC で測定した親和性と FMO で求めた親和性の相関性を解析した報告例は現在までになく、本研究が初めての報告となるとともに、

その相関性が非常に強いという知見も初めての研究成果であると考えられる。

本研究により、創薬研究において **FMO** はエンタルピー寄与の評価に有用なアプリケーションであり、その正確な評価はエンタルピー寄与を高める医薬品候補を設計する上で、重要な指標となることが示唆された。さらに、**FMO** の適用範囲には、医薬品で利用されることがある塩素原子などのハロゲン原子が含まれることも有用であると考えられる。

一方、**FMO** 計算から ΔG を算出することは未解決の課題である。現在、いくつかの方法が開発中であり、例えば、ポアソン-ボルツマン方程式に基づく連続溶媒モデルを使用して、溶媒和の自由エネルギーを考慮することでいくつかの効果を取り入れる試みがある⁴¹⁾。また、古典的な分子動力学と **FMO** を併用することで、**IFIE** と ΔG を比較した最近の研究がある⁴²⁾。これらの水の取り扱いが重要であり、**FMO** と他の計算手法の組み合わせにより、**ITC** の全ての測定値を算出できる発展が期待される。

3 化合物の Ames 予測における Graph Convolutional Network の特徴評価

3.1 研究背景

Ames 試験は、化合物の潜在的な発がん性、催奇形性の早期警戒システムとして、創薬において重要な試験項目の一つである。Ames 予測では LBDD、特に人工知能 (AI) 技術の一つである機械学習を用いた研究が行われており⁴⁴⁾、コンピュータによる Ames 試験予測は、ICH M7「潜在的発がんリスクを低減するための医薬品中 DNA 反応性 (変異原性) 不純物の評価及び管理」ガイドライン⁴⁵⁾において、実験に代わってリスク評価に用いることが可能と言及されており、創薬研究において Ames 試験の予測精度向上、予測精度評価は重要である。

機械学習では、予測モデルが本来の性能を発揮できる場面の研究として、Zhu らは、Applicability Domain (AD) という領域を定義し、その領域内の化合物の予測精度は、領域外の化合物の予測精度より高いことを報告している⁴⁶⁾。Zhu らは、また、予測対象化合物が AD 領域外となる大きな理由として、構造類似度が影響していることを報告しており、予測対象化合物が学習データと類似しているほど、AD 領域内になる可能性が高くなることを示唆している。しかし創薬研究においては、着目する医薬品候補化合物と類似した構造が学習データに含まれないことが多く、医薬品候補化合物が AD 領域外となる傾向が強い。機械学習による Ames 試験の予測において、このような医薬品候補化合物の予測精度は課題である。

一方、化合物の機械学習手法は、情報処理分野における AI 技術の発展とともに、新しい手法が提案されており、その一つがグラフ畳み込みネットワーク (GCN; Graph Convolutional Network) である⁴⁷⁾。機械学習手法の一つであるディープラーニング⁴⁸⁾は、画像処理分野で大きな成果を挙げているが、その成果への貢献が大きい手法の一つが、畳み込みニューラルネットワーク (CNN; Convolutional Neural Network) である⁴⁹⁾。畳み込みとは、関数を平行移動しながら重ねて足し合わせる二項演算で特徴抽出に適し

た手法であり、画像データにおいては威力を発揮する。**GCN**は、グラフ構造、すなわち、ノード（頂点）群とノード間の連結関係を表すエッジ（枝）群で構成される抽象データ型に対して畳み込みを行う手法である。原子をノード、原子の結合をエッジとすれば、化学構造式はグラフ構造であるため、適切に化学構造を表現できると考えられ **GCN** を適用した機械学習の応用が期待されている。

化学構造を表現する代表的手法は、フィンガープリント法であるが、これは分子の局所的な構造を考慮した特徴表現手法であり、近傍の定義を事前に定める必要がある⁵⁰⁾。一方、**GCN**による化学構造表現は、静的距離の近傍の原子だけでなく、畳み込みにより、遠方の原子の情報を動的に反映することができる。また、フィンガープリントでは、単一の原子特性（原子種、もしくは、水素結合性等の特徴）と原子間の結合の種類から表現されるが、**GCN**では原子に複数の特性を持たせる表現が可能である。

GCNを化学構造の実験値予測に応用した例では、Liらの化学的安定性¹³⁾、Ishidaらの合成反応予測に用いた報告⁵¹⁾がある。毒性予測に応用した例では、Wuらは⁵²⁾複数のターゲットの薬理活性、溶解度や脂溶性等の物性、複数の毒性予測について多くの手法の比較結果を報告している。Wuらの報告では、**GCN**の有用性が示唆されているものの、各予測対象の学習データ、テストデータは固定であるため、どのような学習データを用いた場合に予測性能が優れているか、適用範囲が明らかでない。

そこで、本研究では、**GCN**を化学構造に応用した場合、どのような条件下で既存手法より予測性能が向上するかを明らかにするために、**Ames**試験について、フィンガープリント法と代表的な化合物のクラス分けで用いられる既存の3手法（**Support Vector Machine**、**Random Forest**、**Gradient Tree Boosting**）の組み合わせによる予測精度と、**GCN**の予測精度を比較した。特に、**AD**領域外の化合物に対する予測性能について比較するために、テストデータとして学習データと類似している化合物群と類似してない化合物群を用意し、それぞれに対する予測精度を評価し、各手法に与える影響について考察した。

表 3：各種手法の予測精度（ROC-AUC）結果

	静的データセット				ランダムデータセット 1				ランダムデータセット 2			
	GCN	SVM	RF	GB	GCN	SVM	RF	GB	GCN	SVM	RF	GB
テスト類似	0.827	0.790	0.806	0.813	0.895	0.876	0.874	0.886	0.890	0.868	0.878	0.881
テスト非類似	0.764	0.668	0.667	0.670	0.813	0.771	0.780	0.794	0.823	0.774	0.768	0.774
差	0.0631	0.1220	0.1394	0.1437	0.0817	0.1051	0.0942	0.0918	0.0675	0.0946	0.1102	0.1072

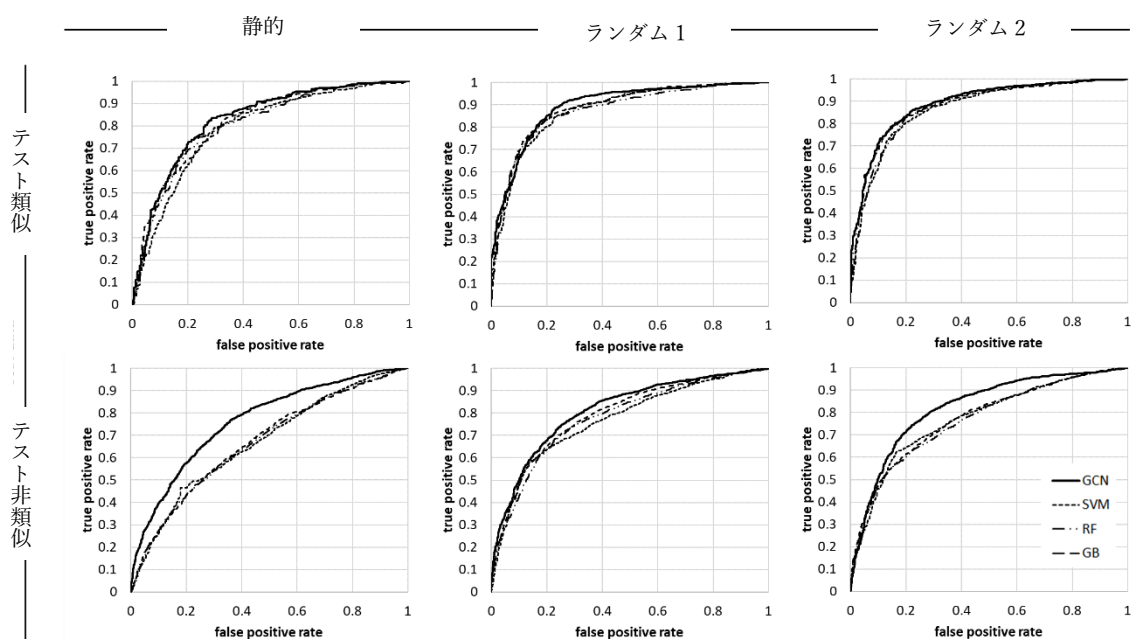


図 9：各種予測手法の ROC 曲線。

3.2 結果と考察

各手法の予測精度（ROC-AUC）を表 3 に、そのグラフを図 9 に示す。また、GCN

表 4 : GCN と各種既存手法の予測精度の差

	静的データセット			ランダムデータセット 1			ランダムデータセット 2		
	SVM	RF	GB	SVM	RF	GB	SVM	RF	GB
テスト類似	0.0374	0.0213	0.0142	0.0184	0.0205	0.0086	0.0223	0.0127	0.0093
テスト非類似	0.0963	0.0976	0.0948	0.0417	0.0329	0.0186	0.0494	0.0554	0.0489

と既存手法との予測精度の差を表 4 に示す。GCN と既存手法との予測精度の差の最小値を確認すると、静的データセットのテスト-類似では、0.0142、テスト-非類似では 0.0948 であった。同様にランダムデータセット 1 においては、それぞれ、0.0086、0.0186、ランダムデータセット 2 では、0.0093、0.0489 であった。このように、GCN の予測精度は全てのケースにおいて、差が正の値となり、既存手法より優れていた。

3.2.1 構造類似度の影響

テストデータの構造類似性に注目すると、どの手法においてもテスト類似に対する予測のほうが、テスト非類似に対する予測より優れていた。Zhu らの AD 領域の報告[2-3]では、学習セットとの構造類似度が高いテスト化合物群は、低いテスト化合物群より AD 領域内となる構造が多い。このことから、テスト類似は AD 領域内になる確率が高い化合物群であり、本研究においても、先行研究と同様に AD 領域内のテストデータのほうが、AD 領域外のテストデータより予測性能が優れる傾向を示した。この傾向は、GCN と既存手法で共通であった。

予測精度が最も悪かったのは、どの手法においても、静的データセットのテスト非類似化合物群に対する予測であった。このデータセットでは、分子量 250 以下の化合物の割合が、学習データでは約 70%であるのに対して、テストデータでは 55%と分子量分布が異なっている(図 10)。さらに、学習データと構造類似度も低いことから、学習デ

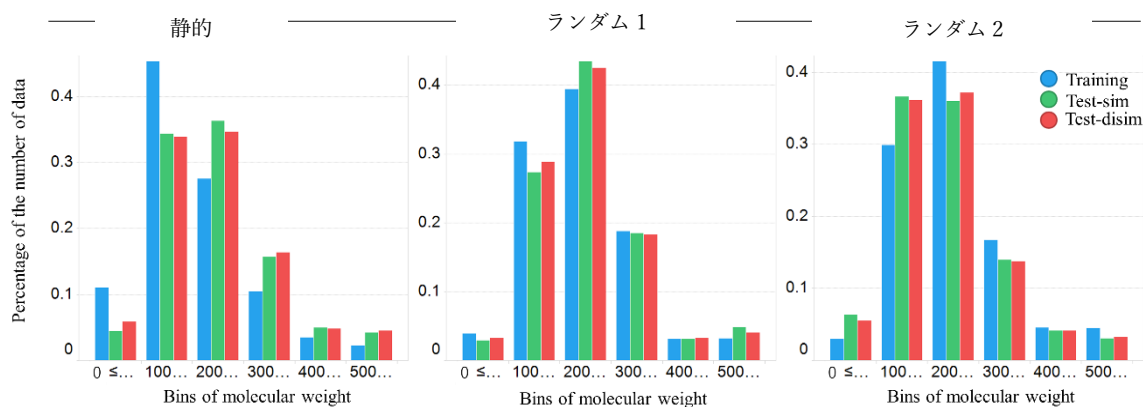


図 10：各データセットの分子量ヒストグラム

表 5：各条件におけるメジャーフィンガープリントビットの数

条件①	条件②	静的データセ ット	ランダムデー タセ ット 1	ランダムデー タセ ット 2
学習データに存在し ていないビット	テスト類似のみ	3	2	1
	テスト非類似のみ	92	22	23
	両方にあり	120	10	13
学習データに存在し ているビット	テスト類似のみ	1	16	2
	テスト非類似のみ	67	67	76
	両方にあり	738	895	900
その他		3	12	9

ータに含まれてない部分構造がテストデータに存在していると考えられる。一方、ランダムデータセット 1・2 では、全体の化合物群から無作為に学習データとして選択しているため、テストデータを分割した構造類似度は 0.5 で静的データセットと変わらないにも関わらず、分子量分布は学習データと類似している。

表 5 は、学習データに含まれてない部分構造を確認するために、各データセットの化合物のフィンガープリントビットのうち、出現回数が 1%以上のフィンガープリントビット（メジャーフィンガープリントビットと呼ぶ）を集計した表である。静的データセ

ットでは、学習データにはなく、テストデータのみにあるメジャーフィンガープリントビットは 120 個であったが、ランダム 1・2 のデータセットでは、それぞれ 10、18 個と少なかった。また、学習データとテストデータすべてに存在していたメジャーフィンガープリントビットは、静的、ランダム 1・2 の各データセットにおいて、それぞれ、738、895、900 個であり静的データが最も少なく、静的データとランダム 1・2 で異なる傾向であった。

予測精度は、全ての手法において、静的データセットよりランダム 1・2 のデータセットのほうが優れており、静的データセットでは、学習データに含まれてないメジャーフィンガープリントビットがテストデータに多く含まれていたことが、他のデータセットより予測精度が低くなった原因と考える。各データセットの類似/非類似の分割に用いた類似度閾値は同じであるため、予測性能に与える影響は、分子全体の類似度より、テストセットに含まれていないメジャーフィンガープリントビット（≒部分構造）の影響が大きいといえる。

3.2.2 GCN と既存手法との比較

各手法における、テスト-類似、非類似の予測精度の差を確認すると、既存手法では、最小 0.095、最大 0.144 であったのに対して、GCN は最小 0.063、最大 0.082 と、いずれも既存手法より差が小さかった。このことから、GCN は、学習データと構造類似度が低く AD 領域外になる確率が高い化合物群の予測性能が、既存手法より、より優れていた。

特に、静的データセットのテスト-非類似の予測精度は、GCN と既存手法の差が最も大きかった。各テストデータに対する既存手法と GCN の ROC 散布図（図 11）を見ると、プロット点は ROC AUC が小さいほど直線から離れている傾向であり、既存手法での予測精度が悪いほど、GCN の予測精度がより優れている傾向であった。

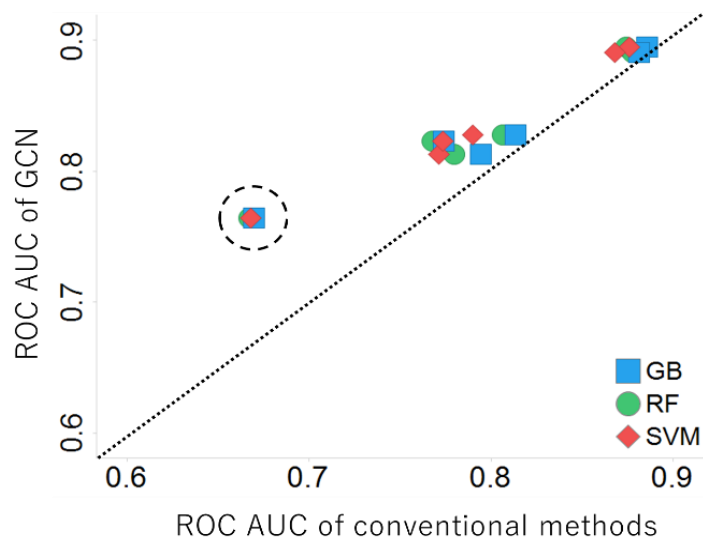


図 11： 既存手法の予測精度（X軸）とGCNの予測精度（Y軸）の散布図。予測精度は ROC-AUC で算出。

すなわち、GCN は、学習データと構造類似度が低い化合物群の予測性能が既存手法より優れており、特に、学習データには少ない部分構造を持つ化合物群において、より、既存手法より予測性能が優れている特徴を持つ。

一方、テスト-類似に対する予測精度は、ランダム 1・2 のデータセットにおいて、既存手法と GCN と予測精度の差が小さく、学習データと類似している構造群に対する予測性能は、GCN のほうが優れていたものの、既存手法と大きな違いはないと考える。

3.2.3 既存手法間比較

GCN 以外の SVM、RF、GB、各手法間の予測結果を比較すると、ROC の差は最大で 0.023、ほとんどの場合 0.015 未満であり、GCN との差に比べると小さく、既存手法でディスクリプタをフィンガープリントとした場合、既存手法間の差は小さかった。

3.2.4 GCN のハイパーパラメータと予測精度の関連

表 6 : GCN の各ハイパーパラメータ組み合わせに対する予測精度 (ROC-AUC)

		ドロップアウト		
		0.2	0.5	0.9
畳 み 込 み 層	64*2	0.764	0.750	0.721
	64*3	0.711	0.762	収束せず
	64*4	0.720	0.757	収束せず

予測精度に影響を与えるハイパーパラメータの効果を確認するために、既存手法との差が大きかった静的データセットのテスト-非類似に対する予測精度を、全ハイパーパラメータの組み合わせについて表 6 にまとめた。ドロップアウトは頑健な予測モデル作成に影響するハイパーパラメータであるが、ドロップアウト 0.9 の場合、64*3、64*4 階層において学習が収束しなかった。これは情報が欠損しすぎて学習できなかったと推察する。ドロップアウト 0.5 の 64*3 階層の時には、最も良い予測精度に近い 0.762 の値をとり、同じドロップアウト 0.5 の 64*4、64*2 階層も 0.75 を超える良い予測精度であり、ドロップアウト 0.5 は総じて良い予測精度であった。しかし、最も良い予測精度であったのは、ドロップアウトが 0.2 ; 畳み込み層が 64*2 階層の時であり、ROC-AUC は 0.764 であった。AD 領域外となる可能性が高い化合物群の予測精度が既存手法より優れていた理由がドロップアウトの効果であるか否か、また、畳み込み層の構造が影響しているかは、不明瞭な結果であった。

3.2.5 GCN と既存手法の予測精度の差の理由考察

Steven らは、グラフ構造による化合物表現では、ディープラーニングのネットワーク構造によりフィンガープリントに比べて化合物構造表現の柔軟性が異なることを報告している⁵³⁾。また、グラフ構造による化合物表現は予測モデルと一体化しており、フィ

ンガープリントと予測モデル作成手法の関係のように明確な区別がない統合的なアプローチであることを述べている。本研究においては、ディープラーニング系の特徴であるハイパーパラメータ（ドロップアウト、畳み込み層の構造）の影響は明らかではなく、学習が収束したどの組み合わせでも 0.7 以上の予測精度であった。このことから、GCN の予測精度が優れていた理由として、ディープラーニングによる予測モデル作成の影響よりも、化合物をグラフ構造で表現し予測モデルと統合することで化合物構造表現が柔軟になる GCN の特徴であると考えられる。

3.3 小括

本研究では、化学構造から Ames の実験結果を予測する手法について、複数の予測手法を比較した。ディープラーニング系の機械学習アルゴリズムである GCN は、本研究で作成した全てのテストデータセットに対して、既存の機械学習手法より最も優れた予測精度であった。また、GCN は、AD 領域外となる可能性が高い化合物群に対して、より優れた予測性能を示した。特に、学習データセットに少数しか含まれていない部分構造を含むような化合物群の予測において、既存手法より優れていた。創薬研究においては、医薬品候補化合物を特許・権利化するために、化学構造の新規性が求められる。その結果、学習データセットには含まれない部分構造を持つ化合物の Ames 予測が求められることは多い。本研究で示した GCN の特性は、そのような場合において、既存手法より有効であった。すなわち、GCN が既存手法と比べて創薬研究において優れた性能の機械学習アルゴリズムであり、LBDD のアプローチにおいて有効な手法であると結論付ける。

4 結語

本研究では、SBDD、LBDD 両アプローチの計算化学について、創薬研究における有用性について研究した。SBDD アプローチにおいては、FMO 計算により求めた相互作用エネルギーが ITC で測定した ΔH と強い相関があることを示した。この結果は、FMO 計算による相互作用エネルギーが強くなるように化合物設計を行うことで、化合物のエンタルピー寄与を強くできる事を示している。製薬企業は創薬研究段階において、特異的結合であるエンタルピー駆動型の医薬品候補化合物を望んでおり、本研究成果は FMO 法の有用性の新たな知見を加えるとともに、より良い医薬品の創製に貢献する成果であると考えられる。

また、LBDD においては、GCN が既存手法より優れた機械学習予測アルゴリズムであることを示した。特に、AD 領域外となる化合物群に対して、既存手法より予測精度の低下が緩やかであることを明らかにした。近年の創薬研究においては、多くの機械学習予測モデルにより合成前化合物のプロファイリングが行われており、その結果は合成順位の優先順位付けに利用されている。本研究では GCN を利用することで、より正確な優先順位付けが可能となり、有望な医薬品候補化合物を優先的に合成することで、製薬企業の課題の一つである、創薬期間の短縮に貢献出来ると考える。

以上、本研究にて明らかにした SBDD、LBDD の両知見は、計算化学により創薬研究を効率化できる場面を新たに提供する知見であり、計算化学がより一層創薬研究の効率化に貢献できることを示す成果だと考える。

謝辞

本論文の執筆に際し、終始ご懇切なるご指導を賜りました星薬科大学 米持 悦生 教授、大阪大学 大学院 薬学研究科 福澤 薫 教授に深甚なる謝意を表します。有益なご助言を頂きました星薬科大学 古石 誉之 講師に謹んで感謝の意を表すとともに、実験に協力してくださった山崎 優輝 氏に厚くお礼申し上げます。また、2章の研究において、FMO drug design consortium の協力をいただきました。立教大学 理学部 化学科の望月 祐志 教授には、ABINIT-MP プログラムを提供していただきました。

本研究に際し、多大なるご支援と発表の機会を与えてくださいました、キッセイ薬品工業株式会社代表取締役 神澤 陸雄 会長並びに代表取締役 竹花 泰雄 社長に深甚なる感謝の意を表します。

本研究に際し、多大なるご支援とご助言を頂きました、キッセイ薬品工業 研究本部長 菊池 伸次 博士、基盤技術研究所長 小沢 知永 博士に深く感謝いたします。

本研究の遂行にあたり、化学合成においては、米窪 滋 氏、タンパク質精製、ITC 実験においては、百瀬 隆喜 博士、X線結晶構造解析においては、古屋 憲孝 博士に多大なるご助言を頂きました。また、キッセイ薬品工業株式会社中央研究所の皆様には、様々な場面でご協力とご助言を承りましたことを感謝いたします。

最後に、本研究の遂行を陰ながら支えてくれた家族に、心より感謝します。

引用文献

- 1) C. M. Song, S. J. Lim, and J. C. Tong, Recent advances in computer-aided drug design, *Briefings in Bioinformatics* 10 (2009) 579-591 【1章】
- 2) J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, Innovation in the pharmaceutical industry: New estimates of R&D costs, *Journal of Health Economics* 47 (2016) 20-33 【1章】
- 3) D. Vohora and G. Singh, *Pharmaceutical Medicine and Translational Clinical Research*, Academic Press 2018. 【1章】
- 4) F. Zhong, J. Xing, X. Li, Artificial intelligence in drug design, *Science China Life Sciences* 61 (2018) 1191-1204 【1章】
- 5) W. Yu and A. D. Mac Kerell, *Computer-aided drug design methods, in Antibiotics*, Humana Press (2017) 【1章】
- 6) 理研 スーパーコンピュータ「富岳」, <https://www.r-ccs.riken.jp/fugaku/> 【1章】
- 7) S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi, Role of computer-aided drug design in modern drug discovery, *Archives of Pharmacal Research* 38 (2015) 1686-1701 【1章】
- 8) W. Duch, K. Swaminathan, and J. Meller, Artificial intelligence approaches for rational drug design and discovery, *Current Pharmaceutical Design* 13 (2007) 1497-1508 【1章】
- 9) M. Batool, B. Ahmad, and S. Choi, A structure-based drug discovery paradigm, *International Journal of Molecular Sciences*, 20 (2019) 2783 【1章】
- 10) H.-J. Huang, H. W. Yu, C.-Y. Chen et al., Current developments of computer-aided drug design, *Journal of the Taiwan Institute of Chemical Engineers* 41 (2010) 623-635 【1章】
- 11) M. Hassan Baig, K. Ahmad, S. Roy, Computer aided drug design: success and limitations, *Current Pharmaceutical Design* 22 (2016) 572-581 【1章】
- 12) A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, K. M. Al-Anazi, An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19,

BioMed Research Internationa 2021 (2021) 【1 章】

- 13) Li X, Yan X, Gu Q, Zhou H, Wu D, Xu J., DeepChemStable: Chemical Stability Prediction with an Attention-Based Graph Convolution Network, *J Chem Inf Model.* 59 (2019) 1044-1049 【1、3 章】
- 14) K. Pettersson、 J. A. Gustafsson, Role of estrogen receptor beta in estrogen action, *Annu. Rev. Physiol.* 63 (2001) 165-192 【2 章】
- 15) G. G. J. M. Kuiper, B. Carlsson, K. Grandien, E. Enmark, J. Häggblad, S. Nilsson, J. Gustafsson, Comparison of the Ligand Binding Specificity and Transcript Tissue Distribution of Estrogen Receptors α and β , *Endocrinology* 138 (1997) 863– 870 【2 章】
- 16) H. A. Harris, Preclinical characterization of selective estrogen receptor beta agonists: new insights into their therapeutic potential, *Ernst. Schering. Found Symp. Proc.* (2006) 149–161 【2 章】
- 17) G. I. Paterni, C. Granchi, J. A. Katzenellenbogen, F. Minutolo, Estrogen Receptors Alpha ($ER\alpha$) and Beta ($ER\beta$): Subtype-Selective Ligands and Clinical Potential, *Steroids* 90 (2014) 13–29 【2 章】
- 18) M. S. Malamas, E. S. Manas, R. E. McDevitt, I. Gunawan, Z. B. Xu, M. D. Collini, C. P. Miller, T. Dinh, R. A. Henderson, J. C. Keith Jr., H. A. Harris, Design and Synthesis of Aryl Diphenolic Azoles as Potent and Selective Estrogen Receptor- β Ligands, *J. Med. Chem.* 47 (2004) 5021-5040 【2 章】
- 19) E. S. Manas, R. J. Unwalla, Z. B. Xu, M. S. Malamas, C. P. Miller, H. A. Harris, C. Hsiao, T. Akopian, W-T Hum, K. Malakian, S. Wolfrom, A. Bapat, R. A. Bhat, M. L. Stahl, W. S. Somers, Structure-based design of estrogen receptor-beta selective ligands, *J. Am. Chem. Soc.* 126 (2004) 15106-15119 【2 章】
- 20) E. S. Manas, Z. B. Xu, R. J. Unwalla, W. S. Somers, J. C. Understanding the Selectivity of Genistein for Human Estrogen Receptor- β Using X-Ray Crystallography and Computational Methods, *Structure* 12 (2004) 2197-2207 【2 章】
- 21) H. A. Harris, L. M. Albert, Y. L. Leathurby, M.S. Malamas, R. E. Mewshaw, C. P. Miller, Y. P. Kharode, J. Marzolf, B. S. Komm, R. C. Winneker, D. E. Frail, R. A. Henderson , Y. Zhu, J. C. Keith, Evaluation of an estrogen receptor- β

- agonist in animal models of human disease. *Endocrinology* 144 (2003) 4241–4249 【2 章】
- 22) T. I. Richardson, J. A. Dodge, G. L. Durst, L. A. Pfeifer, J. Shah, Y. Wang, J. D. Durbin, V. Krishnan, B. H. Norman, Benzopyrans as selective estrogen receptor β agonists (SERBAs). Part 3: Synthesis of cyclopentanone and cyclohexanone intermediates for C-ring modification. *Bioorg. Med. Chem. Lett.*, 17 (2007) 4824–4828 【2 章】
- 23) M. J. Meyers, J. Sun, K. E. Carlson, G. A. Marriner, B. S. Katzenellenbogen, J. A. Katzenellenbogen, Estrogen receptor beta potency-selective ligands: structure–activity relationship studies of diarylpropionitriles and their acetylene and polar analogues. *J. Med. Chem.* 44 (2001) 4230–4251 【2 章】
- 24) J. Sun, M. J. Meyers, B. E. Fink, R. Rajendran J. A. Katzenellenbogen, B. S. Katzenellenbogen, Novel ligands that function as selective estrogens or antiestrogens for estrogen receptor alpha or estrogen receptor-beta. *Endocrinology*. 140(1999), 800–804 【2 章】
- 25) J. Sun, J. Baudry, J. Katzenellenbogen, B. S. Katzenellenbogen, Molecular Basis for the Subtype Discrimination of the Estrogen Receptor-Beta-Selective Ligand, Diarylpropionitrile. *Molecular Endocrinology* 17 (2003) 247–258 【2 章】
- 26) V. M. Carrola, M. Jeyakumara, K. E. Carlson, J. A. Katzenellenbogen, Diarylpropionitrile (DPN) Enantiomers: Synthesis and Evaluation of Estrogen Receptor Beta-Selective Ligands. *J. Med. Chem.* 2012 (55) 528–537 【2 章】
- 27) J. B. Chaires, Calorimetry and thermodynamics in drug design. *Ann. Rev. Biophys.* 37 (2008) 135–151 【2 章】
- 28) E. Freire, Do enthalpy and entropy distinguish first in class from best in class? *Drug Discovery Today* 13 (2008) 869-874 【2 章】
- 29) A. Biela, F. Sielaff, F. Terwesten, A. Heine, T. Steinmetzer, G. Klebe, Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. *J. Med. Chem.* 55 (2012) 6094-6110 【2 章】
- 30) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, Fragment molecular orbital method: an approximate computational method for large molecules,

- Chem. Phys. Lett. 313 (1999) 701-706 【2 章】
- 31) T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, K. Kitaura, Fragment molecular orbital method: use of approximate electrostatic potential, Chem. Phys. Lett. 351 (2002) 475-480 【2 章】
- 32) Y. Mochizuki, S. Tanaka, K. Fukuzawa, Recent Advances of the Fragment Molecular Orbital Method, <https://doi.org/10.1007/978-981-15-9235-5> 【2 章】
- 33) K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma, T. Nakano, Ab initio quantum mechanical study of the binding energies of human estrogen receptor α with its ligands: an application of fragment molecular orbital method. J. Comput. Chem. 26 (2005) 1–10 【2 章】
- 34) K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura, T. Nakano, Molecular interactions between estrogen receptor and its ligand studied by the ab initio fragment molecular orbital method, J. Phys. Chem. B 17 (2006) 16102-16110 【2 章】
- 35) M. Kawasaki, A. Kambe, Y. Yamamoto, S. Arulmozhiraja, S. Ito, Y. Nakagawa, H. Tokiwa, S. Nakano, H. Shimano, Elucidation of molecular mechanism of a selective PPAR α modulator, pemafibrate, through combinational approaches of X-ray crystallography, thermodynamic analysis, and first-principle calculations, Int. J. Mol. Sci., 21 (2020), p. 361 【2 章】
- 36) B. Drobot, M. Schmidt, Y. Mochizuki, T. Abe, K. Okuwaki, F. Brulfert, S. Falke, S.A. Samsonov, Y. Komeiji, C. Betzel, T. Stumpf, J. Raff, S. Tsushima, Cm 3+/Eu 3+ induced structural, mechanistic and functional implications for calmodulin, Phys. Chem. Chem. Phys., 21 (2019), pp. 21213-21222 【2 章】
- 37) E. J. Ariëns, Stereochemistry: A source of problems in medicinal chemistry, Medicinal Research Reviews., 6 (1986) 451–466 【2 章】
- 38) E. J. Ariëns, E. W. Wuis, E. J. Veringa, Stereoselectivity of bioactive xenobiotics, Biochemical Pharmacology. 37 (1988) 9–18 【2 章】
- 39) D. Takaya, C. Watanabe, S. Nagase, K. Kamisaka, Y. Okiyama, H. Moriwaki, H. Yuki, T. Sato, N. Kurita, Y. Yagi, T. Takagi, N. Kawashita, K. Takaba, T. Ozawa, M. Takimoto-Kamimura, S. Tanaka, K. Fukuzawa, T. Honma, FMODB: The World's First Database of Quantum Mechanical Calculations for

- Biomacromolecules Based on the Fragment Molecular Orbital Method, J. Chem. Inf. Model, 61 (2021) 777-794 【2章】
- 40) K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, AD. Mackerell, CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. J. Comput. Chem. 31 (2010) 671–690 【2章】
- 41) Y. Okiyama, C. Watanabe, K. Fukuzawa, Y. Mochizuki, T. Nakano, S. Tanka, Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson–Boltzmann Equation: II. Protein and Its Ligand-Binding System Studies, J. Phys. Chem. B, 123 (2019), pp. 957-973 【2章】
- 42) K. Takaba, C. Watanabe, A. Tokuhisa, Y. Akinaga, B. Mao, R. Kanada, M. Araki, Y. Okuno, Y. Kawashima, H. Moriwaki, N. Kawashita, T. Honma, K. Fukuzawa, S. Tanaka, Protein–ligand binding affinity prediction of cyclin-dependent kinase-2 inhibitors by dynamically averaged fragment molecular orbital-based interaction energy, J. Comput. Chem, 43 (2022), pp. 1362-1371 【2章】
- 43) PCT Souza, LC Textor, DC Melo, AS Nascimento, MS Skaf, I. Polikarpov. An alternative conformation of ER β bound to estradiol reveals H12 in a stable antagonist position. Sci. Rep. 7 (2017) 3509 【2章】
- 44) Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N, Müller KR., "Benchmark Data Set for in Silico Prediction of Ames Mutagenicity", J Chem Inf Model., 49 (2009) 2077-2081 【3章】
- 45) ICH-M7 潜在的発がんリスクを低減するための医薬品中 DNA 反応性（変異原性）不純物の評価及び管理， <https://www.pmda.go.jp/int-activities/int-harmony/ich/0036.html> 【3章】
- 46) Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, Oberg T, Dao P, Cherkasov A, Tetko IV., Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*, J Chem Inf Model. 48(2008) 766-84 【3章】
- 47) Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg,

- Ivan Titov, Max Welling., Modeling Relational Data with Graph Convolutional Networks, arXiv:1703.06103 (2017) 【3 章】
- 48) J Weston, F Ratle, H Mobahi, R Collobert., Neural Networks, Tricks of the Trade (2012) 639-655 【3 章】
- 49) A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Proc. Adv. Neural Inf. Process. Syst. (2012) 1097-1105 【3 章】
- 50) H. L. Morgan, The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service., J. Chem. Doc 5, pp (1965) 107-113 【3 章】
- 51) Ishida S, Terayama K, Kojima R, Takasu K, Okuno Y., Prediction and Interpretable Visualization of Retrosynthetic Reactions Using Graph Convolutional Networks, J Chem Inf Model. 59 (2019) 5026-5033 【3 章】
- 52) Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V., MoleculeNet: a benchmark for molecular machine learning, arXiv:1703.00564v2 (2017) 【3 章】
- 53) Steven K., Kevin M.C., Marc B., Vijay P., Patrick R, "Molecular graph convolutions: moving beyond fingerprints", Journal of Computer-Aided Molecular Design 30 (2016) 595-608 【3 章】

実験項

2章に関する実験項

ER β リガンド結合ドメインのタンパク発現

ヒト ER β LBD の 3 変異体（残基 261-500、C334S/C369S/C481S）をコードする cDNA を pGEX6P-1（GE Healthcare）にクローニングし、C 末端にさらに 2 つの Asp 残基を付加したものを得た。得られたプラスミドとシャペロンプラスミド pGro7（タカラバイオ社製）で大腸菌 BL21（DE3）細胞を共発現させた。ER β LBD を GroEL および GroES と 20℃で一晩共発現させた後、細胞を採取し、超音波処理により破碎した。上清を遠心分離により回収し、ER β LBD を Glutathione Sepharose 4B（GE Healthcare）に捕捉し、PreScission Protease（GE Healthcare）で 4℃、一晩ゲル処理し、GST を除去した。結晶化のために、ER β LBD を Mono Q 5/50 GL（GE Healthcare）と Superdex75 10/300 GL カラムでさらに精製した。ITC 解析を目的としたタンパク質の最終精製は、Mono Q 5/50 GL（GE Healthcare）カラムで行った。

化学構造の合成

1-6 は、キッセイ薬品工業株式会社の化学合成チームによりキラル合成を行い、純度は、HPLC にて 99%以上であることを確認した。合成方法の詳細は、巻末の実験データ集に記載する。

等温滴定型カロリメトリー（ITC）の測定方法

ITC 測定は、MicroCal iTC200 熱量計を用いて、50 mM NaPi pH 7.4, 150 mM NaCl

および 1 mM tris(2-carboxyethyl)phosphine 中、25℃で行った。滴定は、各試験化合物 (70 μM) を、サンプルセル内の ERβ LBD (10 μM) に行った。各試験は、化合物 0.2 μL の予備注入した後、120 秒間隔で 2 μL ずつ注入した。ITC の測定データは、MicroCal iTC200 Origin ソフトウェアパッケージにて解析した。

X 線結晶構造の解析

結晶化へ向けてタンパク質の純度を高めるために、50 mM Tris-HCl pH8.0, 150 mM NaCl, 5% glycerol, 1 mM tris(2-carboxyethyl)phosphine, 1 mM エチレンジアミン四酢酸の superdex 75 column (GE healthcare) でさらに蛋白質を精製した。精製したタンパク質を 5~7 mg/mL に濃縮し、モル比 10 倍の化合物、モル比 4 倍のコアクチベーターと混合した。20℃でシッティングドロップ法により結晶を大きくし、0.2μL のリザーバー溶液と 0.2μL のタンパク質溶液を混合して、X 線構造解析用の結晶を得た。リザーバー溶液は、16-20% PEG3350、100 mM BisTris pH6.5、200 mM Magnesium Acetate で調整した。回折データの収集は、KEK ビームライン BL-17A、SPring-8 ビームライン BL41XU で行った。得られた回折データは XDS^{e1)}によって統合し、初期位相は Molrep^{e2)}を用いた分子置換によって決定した。CCP4 プログラム^{e3)}の Refmac5^{e4)}と Coot^{e5)}を用いて、モデル構造の作成と修正を繰り返した。複合体の最終座標は Protein Data Bank に登録した (コード: 7XVY、7XVZ、7XWP、7XWQ、7XWR)。

フラグメント分子軌道法による計算

FMO 法では、タンパク質構造を小さな断片 (フラグメント) に分割し、フラグメントおよびそのペアの電子状態を計算して過不足なく集計することで、タンパク質全体電子状態を構築することができる。また各フラグメント単位でタンパク質とリガンド間の相互作用エネルギーを評価することができる^{30)e6)}。本研究では、すべての計算におい

て、ABINIT-MP プログラム^{32)e9)e10)}を用いた。計算レベルは 2 次の Møller-Plesset 摂動理論 (MP2)^{e7)e8)}を用い、6-31G*基底関数を用いた。

FMO 法を用いたタンパク質-リガンド相互作用エネルギーの解析には、2 つの段階がある。まず、各アミノ酸残基とリガンド間の相互作用エネルギーである IFIE^{e10)}を算出する。次に、pair interaction energy decomposition analysis (PIEDA)により、さらに IFIE 値を静電相互作用 (ES)、交換反発 (EX)、電荷移動 (CT-mix)、分散相互作用 (DI) の 4 種類のエネルギー成分に分解し、分子間相互作用の由来を把握できる^{e11)}。

IFIE および PIEDA は次の式であらわされる。

$$\Delta\tilde{E}_{IJ} = \Delta\tilde{E}_{IJ}^{\text{ES}} + \Delta\tilde{E}_{IJ}^{\text{EX}} + \Delta\tilde{E}_{IJ}^{\text{CT+mix}} + \Delta\tilde{E}_{IJ}^{\text{DI}}$$

これらのエネルギー項は、例えば、塩橋や極性相互作用を ES で、水素結合を ES および CT+mix で、立体反発を EX で、CH/ π などの疎水性タイプの相互作用を DI でそれぞれ表すことができる。(なお、本研究では、フラグメント "J" は常に化合物、フラグメント "I" はタンパク質のアミノ酸または水分子となる)。

複合体構造座標の入力ファイルは全て MOE 2018.0101^{e12)}で作成した。リガンド結合部位から離れたアミノ酸残基がいくつか欠損していたため、全複合体構造に共通しているアミノ酸配列として、ER β の Leu263-Ser283、Glu291-Ser409、Asp421-Ala497、co-factor の His604-Thr613 をそれぞれ解析対象とした。配列の末端アミノ酸のプロトン化状態は中性と仮定し、Amber10:EHT 力場を用いて、すべての複合体座標のエネルギー最小化を段階的に行った。最初に主鎖の原子を固定し、側鎖の原子には緩い拘束をかけて構造を最小化した。次に、すべての原子に対して緩い拘束をかけて最小化した。拘束時の強さはパラメータ tether=1.0 とした。

3章に関する実験項

データセット

手法評価には、Hansen らの論文⁴⁴⁾の Ames の実験データセットを用いた。このデータセットは、CCRIS^{e13)}、複数の論文 (Helma ら^{e14)}、Kazius ら^{e15)}、Feng ら^{e16)}、VITIC[15])、及び GeneTox データセット^{e18)}から得ており、機械学習用のデータとして整備済みである。その主な処理内容は次の通りである。

- ・ 重複構造を削除
- ・ 異常および無機分子構造を削除
- ・ DEREK^{e19)}、または、MULTICASE^{e20)}の予測モデル作成に用いられたデータに対して矛盾する実験結果である化合物を削除

最終的なデータセットの分子の数は、陽性 3,503 個、陰性 3,009 個、計 6,512 個である。これらのデータについて、KNIME^{e21)}の RDKit^{e22)}ノードを用いて、本研究の評価用に再処理を行った。その際、6 化合物は、RDKit においてエラーが発生したため、本研究で解析に用いた化合物数は 6,506 個であり、陽性化合物 3,497 個、陰性化合物 3,009 個の内訳であった。

このデータを、学習用データとテスト用データに分割するにあたり、2 つの方法で行った。一つ目の方法では、Hansen らが「静的」と定義したデータセットを学習データとして、残りのデータをテストデータとした。静的とは、DEREK、MULTICASE の双方において、信頼性が高いと判断できた既知化合物群である。二つ目の分割方法では、全化合物からランダムに、学習データとテストデータに分割した。その際、学習データの総数、陽性、陰性化合物数がそれぞれ上記の静的と同じ数となるように分割した。さらに、ランダムに分割したデータセットは、乱数の設定を変えた 2 種類のデータを作成し、計 3 つのデータセット (それぞれ、静的、ランダム 1、ランダム 2 とする) を作成

した。

化合物の構造類似性による予測精度の違いを観察するために、それぞれのデータセットのテストデータを、学習データに類似している化合物とそれ以外の化合物に分割した。化合物の類似度は、RDKit Morgan Fingerprint(1024bit; radius=3)を用いたTanimoto 係数により算出し、0.5 を閾値として、学習データに対する最近傍類似度 (closest similarity) が 0.5 より大きい値を類似している化合物セット (テスト-類似, Test-類似) とし、それ以外の化合物を類似していない化合物 (テスト-非類似, Test-非類似) とした。

予測精度評価

予測精度の評価は、各予測モデルの Probability による ROC 曲線 (Receiver Operating Characteristic curve) の AUC (Area Under the Curve) で行った。ROC 曲線は、予測性能を 2 次元のグラフで表したものであり、縦軸に真陽性率、横軸に偽陽性率をプロットして線で結ぶ。このグラフの下の部分の面積が AUC の値で 0 から 1 までの値をとり、値が 1 に近いほど予測精度が高いことを示す。

Graph Convolutional Network

GCN のネットワーク構成は、入力層、畳み込み層、プーリング層、集約層からなり、各層の概要は次のとおりである。入力層では、指定された化学構造のグラフ情報と、各原子のプロパティのベクトルにより入力値が作成される。畳み込み層では、任意の原子に対して、隣接している原子の特徴ベクトルを畳み込み関数により算出する。この処理を全原子に対して繰り返し行う事で、隣接している原子の特徴が反映されたベクトルが作成される。プーリング層では、畳み込み層で算出された特徴ベクトルの各要素について、最小値等の代表値を算出する。この処理では、代表値を残し他のデータは破棄するため、

情報量を下げ、グラフ構造の特徴を際立たせる処理となる。集約層では、畳み込み層、プーリング層で算出された各原子の特徴ベクトルを、要素ごとに加算して化合物の特徴ベクトルを算出する層である。

GCN は、ディープラーニングの一手法であり、各層において、ディープラーニングの大きな特徴の一つである、ドロップアウトアルゴリズムが適用可能である。ドロップアウトとは、各層のニューロンを定めた確率により働かなくすることで、意図的に情報を欠損させた状態で学習させる。そのような状態でも予測性能を維持するモデルを作成することで、機械学習の大きな課題である過学習を防ぎ、予測モデルの頑健性を高める手法である。

本研究では、DeepChem^{e23)}を用いて GCN の予測モデルを作成しており、各原子の特性は {原子種、原子が結合している水素以外の原子数、データ上で明示されていない結合水素数、原子のフォーマルチャージ、ラジカル電子数、軌道の種類、芳香環に含まれるか否か、総結合水素数 (データ上の明示、非明示双方) } の 8 種類である。予測モデルを作成するには、多くのハイパーパラメータ設定が必要となるが、本研究では、畳み込み層を {64 要素*2 階層、64 要素*3 階層、64 要素*4 階層} (それぞれ DeepChem の表記で、[64, 64], [64, 64, 64], [64,64,64,64]) の 3 パターン、ドロップアウト確率を {0.2, 0.5, 0.9} の 3 パターン、学習回数であるエポック数は 200 に固定し、計 9 パターンの組み合わせで予測モデルを作成し、各テストセットに対する ROC の最大値を GCN の予測値として採用した。その他のハイパーパラメータはデフォルト設定を用いた。

Support Vector Machine (SVM)

SVM アルゴリズムは、Ivanciu らによって開発された^{e24)}。SVM アルゴリズムの大きな特徴は、マージン最大化とカーネルトリックの導入である。SVM では、各サンプルはベクトルで表現されるが、サポートベクトルと呼ばれるクラス境界近傍に位置するベクトルをクラス分けの基準とする。この際、その距離が最も大きくなるような位置に識別

境界を設定する。マージン最大化とは、マージンと呼ばれるこの距離を最大化するように学習を行う手法である。カーネルトリックとは、SVMの予測性能を高めたアルゴリズムであり、線形空間では分離が難しいデータを非線形空間で行う手法である。本研究では、SVMをscikit-learn^{e25)}で実装し、カーネルはデフォルトのRBFカーネルを用いた。RBFカーネルを用いた学習時には、CostとGammaのハイパーパラメータチューニングを行う。Costは誤分類をどの程度許容するかのパラメータであり、Costが大きいほど誤分類を許容せず、学習データへフィッティングしやすくなる。Gammaは境界の複雑さであり、大きくなるほど境界は複雑となる。本研究では、各データセットに対して、Costを{1, 10, 100, 1000}、Gammaを{1.0, 0.1, 0.01, 0.001}の総当たりの組み合わせによる16個の予測モデルを作成し、テストセットそれぞれに対して最も予測精度が高い結果を予測性能として採用した。

ランダム Forest (RF)

RFは、複数の決定木予測モデルを組み合わせるより強力な予測モデルを作るアンサンブル学習手法の一つである。決定木とは、条件に基づいてデータを分割していく学習方法で、「ある変数について、ある閾値で予測結果が分岐する」ということを繰り返して予測モデルを構築する。

RFのアンサンブル学習は、学習データセットから重複を許してサンプリングし、そのサンプル集合に対して決定木の作成を繰り返し複数の決定木を作成する。これら複数の決定木はそれぞれ独立しており、予測結果は、作成された決定木の予測結果の平均となる。

本研究では、決定木の階層数を{5, 7, 10, 15}で試行し、アンサンブルの数は100に固定、scikit-learnにより各データセットに対し4個の予測モデルの作成を行った。決定木の分岐に用いられる変数の選択方法である重要度指標は、デフォルト設定であるGini indexを用い、各テストセットの最も予測精度が高い結果を、予測性能として採用した。

Gradient Tree Boosting(GB)

GB も RF と同じく決定木ベースのアンサンブル学習の一つである。GB と RF の大きな違いは、GB ではブースティングという手法により、複数の決定木を作成する。ブースティングでは、まず、基本となる決定木予測モデルを作成してベースラインとする。その後は、この基本モデルの間違った予測に焦点を当てて、間違った予測を修正するように「重み」を加味して次の予測モデルを作成することで、予測結果を改善していく。これを繰り返し行い、最終的に全ての決定木の結果から予測を行う。

予測木を作成する時のチューニングパラメータとして、決定木の階層数を{5, 7, 10, 15}、学習レートを{0.1、0.05、0.01}の組み合わせを全て試行した。予測モデルの作成は、scikit-learn により行い、その他のパラメータはデフォルト値を用いた。各テストデータあたり 12 個の予測モデルを作成し、その中で最も予測精度が高い結果を採用した。

ディスクリプタ

SVM、RF、GB においては、予測を行うために化合物を表現するディスクリプタが必要となる。本研究では、データセット作成の類似度計算に用いたフィンガープリントをディスクリプタとした。フィンガープリントは 1024 ビットで作成しているため、入力ディスクリプタの数は 1024 個である。

実験項引用文献

- e1) W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125-132.
- e2) A. Vagin, A. Teplyakov, Molecular replacement with MOLREP, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 22-25.
- e3) M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G.

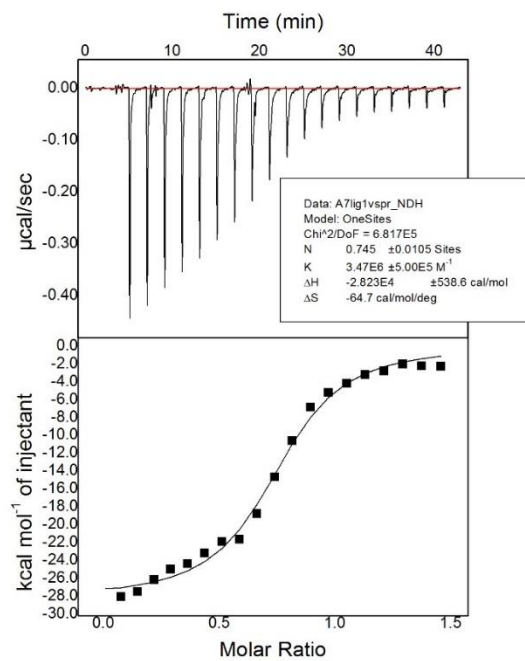
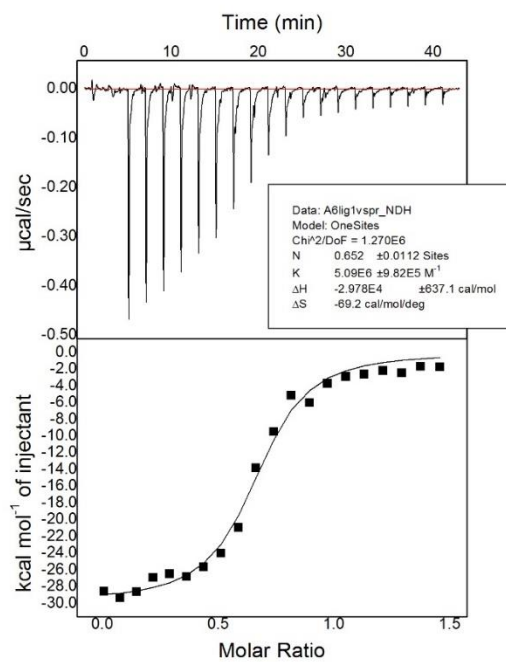
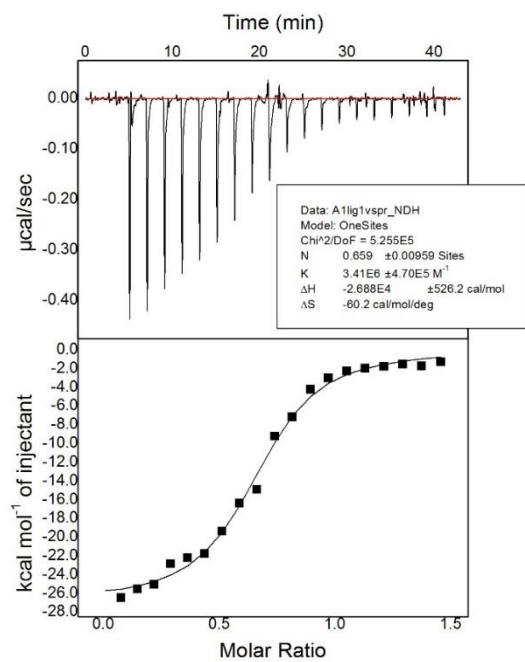
- N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. D Biol. Crystallogr.* 67 (2011) 235-242.
- e4) A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, G. N. Murshudov, REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use, *Acta Crystallogr. D Biol. Crystallogr.* 60 (2004) 2184-2195.
- e5) P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot, *Acta. Crystallogr. D Biol. Crystallogr.* 66 (2010) 486-501.
- e6) D.G. Fedorov, K. Kitaura, The Fragment Molecular Orbital Method: PRACTICAL APPLICATION TO LARGE MOLECULAR SYSTEM, CRC Press, 2009.
- e7) Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima, K. Kitaura, A parallelized integral-direct second-order Møller-Plesset perturbation theory method with a fragment molecular orbital scheme, *Theor. Chem. Acc.* 112 (2004) 442–452.
- e8) Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari, K. Kitaura, Large scale MP2 calculations with fragment molecular orbital scheme, *Chem. Phys. Lett.* 396 (2004) 473–479.
- e9) Y. Mochizuki, K. Yamashita, T. Nakano, Y. Okiyama, K. Fukuzawa, N. Taguchi, S. Tanaka, Higher-order correlated calculations based on fragment molecular orbital scheme, *Theor. Chem. Acc.* 130 (2011) 515–530.
- e10) S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, K. Fukuzawa, Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems, *Phys. Chem. Chem. Phys.* 16 (2014) 10310-10344
- e11) D.G. Fedorov, K. Kitaura, Pair interaction energy decomposition analysis, *J. Comput. Chem.* 28 (2007) 222-237.
- e12) Molecular Operating Environment (MOE), Chemical Computing Group Inc, 1010 Sherbrooke St. West, Suite #910, Canada, Montreal, QC, 2016. H3A 2R7.
- e13) NCRI Informatics Initiative Homepage., "Chemical Carcinogenesis Research Information System", <http://www.cancerinformatics.org.uk/matrix/CCRIS.htm>, (accessed July 17, 2009)
- e14) Helma C, Cramer T, Kramer S, De Raedt L., "Data mining and machine learning techniques for the identification of mutagenicity inducing

- substructures and structure activity relationships of noncongeneric compounds.", J Chem Inf Model., 44, pp. 1402-1411 (2004)
- e15) Kazius J, McGuire R, Bursi R., "Derivation and Validation of Toxicophores for Mutagenicity Prediction", J Chem Inf Model., 45, pp. 312-320 (2005)
- e16) Feng J, Lurati L, Ouyang H, Robinson T, Wang Y, Yuan S, Young S., "Predictive toxicology: benchmarking molecular descriptors and statistical methods.", J Chem Inf Model., 43, pp.1463-1470 (2003)
- e17) Judson PN, Cooke PA, Doerrer NG, Greene N, Hanzlik RP, Hardy C, Hartmann A, Hinchliffe D, Holder J, Miller L, Steger-Hartmann T, Rothfuss A, Smith M, Thomas K, Vessey JD, Zeiger E., "Towards the creation of an international toxicology information centre", Toxicology, 213, pp. 117-128 (2005)
- e18) "Genetic Toxicity, Reproductive and Developmental Toxicity, and Carcinogenicity Database", <http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092217.htm>, (accessed July 17, 2009)
- e19) Sanderson D, Earnshaw C., "Computer prediction of possible toxic action from chemical structure, the DEREK system.", Hum. Exp. Toxicol., 10, pp. 261-273 (1991)
- e20) Klopman G., "MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program.", Quant. Struct.-Act. Relat., pp. 176-184 (1992)
- e21) "KNIME Version 4.1", available at <http://knime.com/>
- e22) "RDKit: Open-Source Cheminformatics Software", available at <https://www.rdkit.org/>
- e23) "DeepChem 2.1", available at <https://www.deepchem.io/>
- e24) Ivanciuc, O., "In Applications of Support Vector Machines in Chemistry", Lipkowitz, K. B., Cundari, T. R., Eds., Chapter 6, pp 291-400 (2007)
- e25) "scikit-learn 0.22", available at <https://scikit-learn.org/>

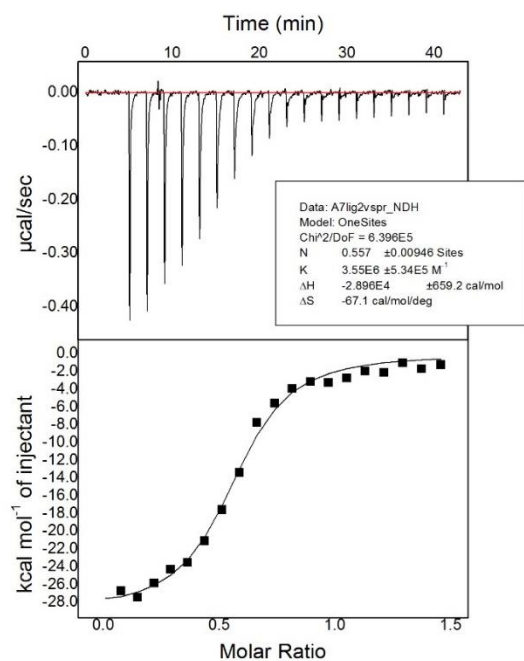
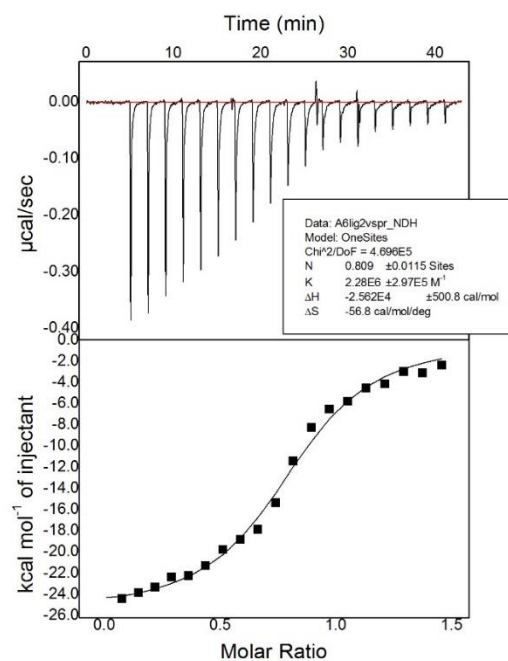
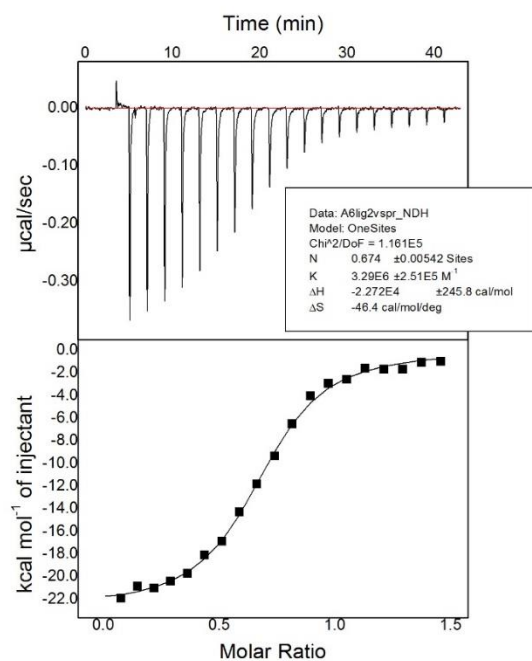
実験データ集

ITC 測定チャート

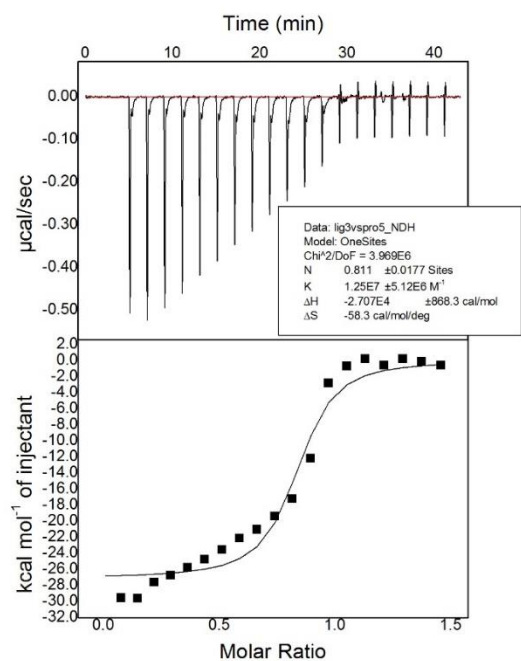
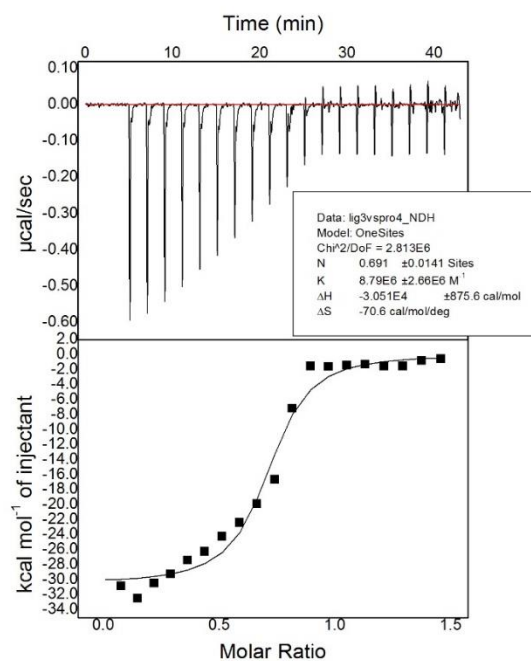
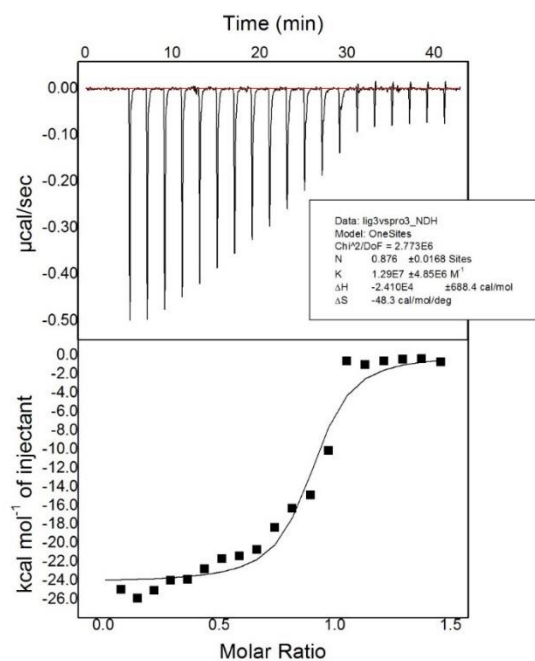
化合物 1(R-DPN)



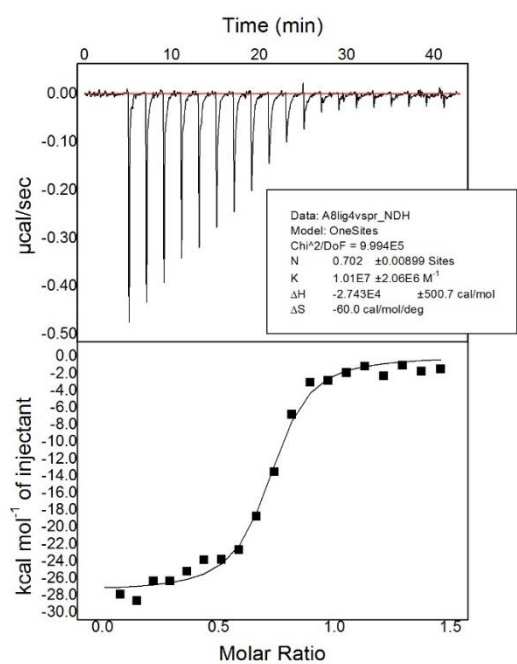
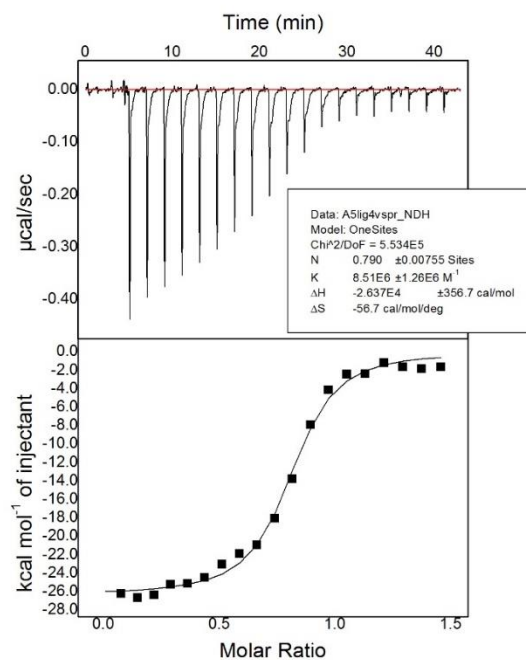
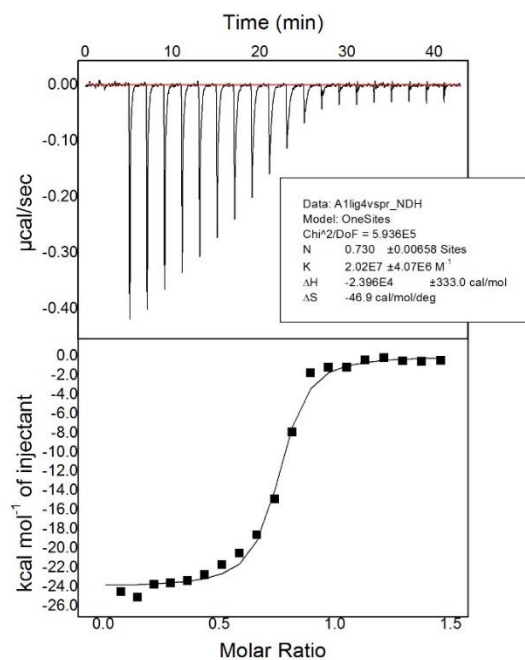
化合物 2(S-DPN)



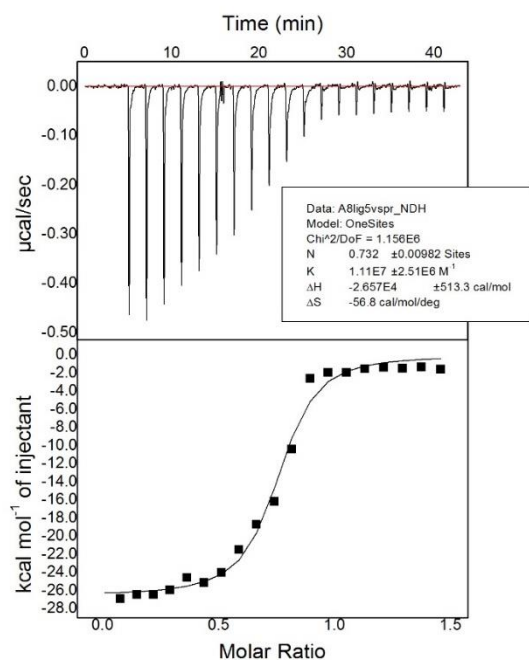
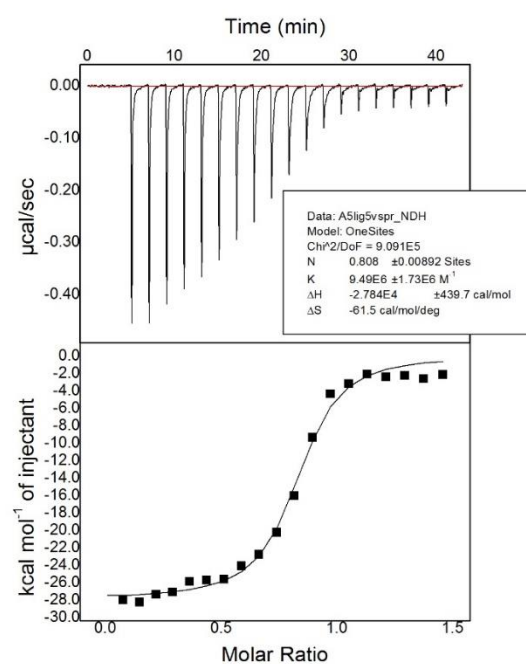
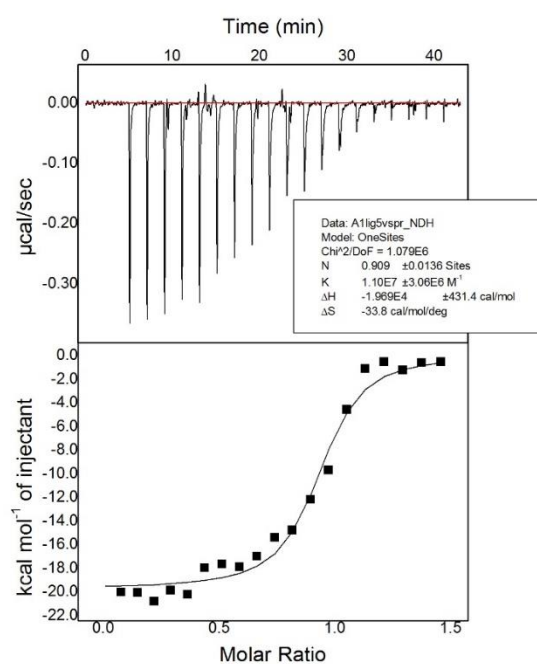
化合物 3



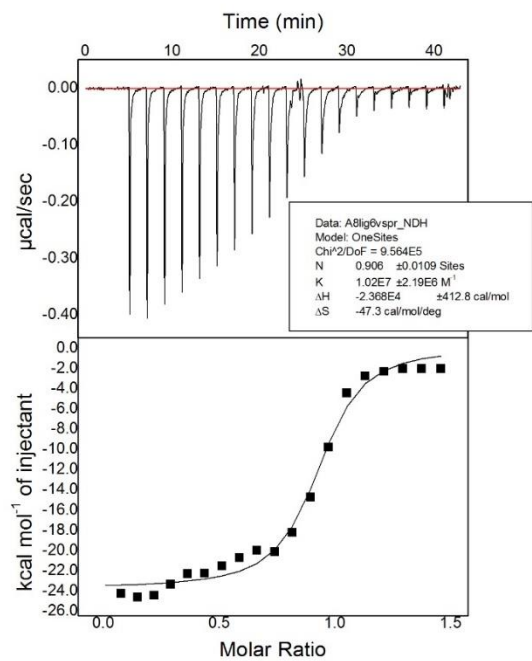
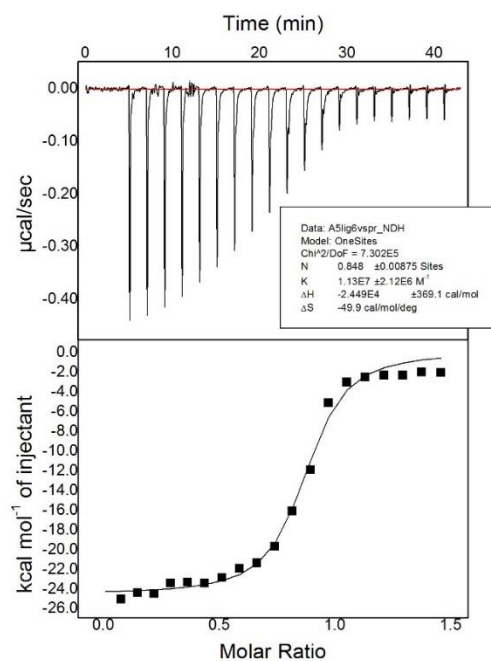
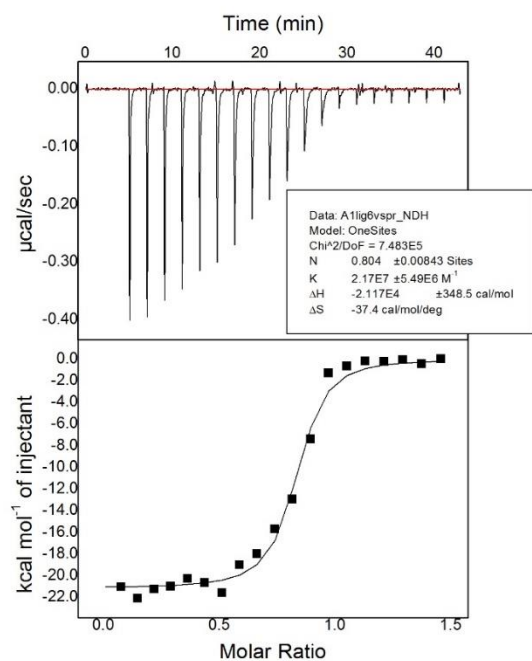
化合物 4



化合物 5



化合物 6



1. Chemistry

All reagents and solvents were purchased from commercial sources and used without further purification. All moisture and air sensitive reactions were carried out under an argon atmosphere. Reactions were monitored by thin layer chromatography (TLC) using Merck Kieselgel 60 F₂₅₄ plates or Fuji Silysia Chemical Ltd. Chromatorex NH-TLC plates. Flash column chromatography was performed on Biotage prepacked columns or Yamazen Hi-Flash columns using an automated flash chromatography system (Biotage, Isolera One). Melting points were determined on a Yanaco micro melting point apparatus (MP-J3). ¹H NMR spectra were recorded on a Bruker AV400M spectrometer at 400.1 MHz. Chemical shifts (δ) are reported in parts per million (ppm) units using tetramethylsilane as an internal standard. Data are presented as follows; chemical shift, multiplicity (d, doublet; m, multiplet), coupling constant, and integration. ¹³C NMR spectra were recorded on a Bruker AV400M spectrometer at 100.6 MHz with complete proton decoupling. Chemical shifts (δ) are reported in parts per million (ppm) with the solvent as the internal reference (δ 49.00 in CD₃OD).^{S1} High resolution mass spectra (HRMS) were recorded on an Agilent Technologies 6520 Accurate-Mass Q-TOF instrument using electrospray ionization (ESI, positive or negative ion mode). The chemical purity and optical purity of tested compounds were determined by HPLC–UV analysis, which was performed on a Shimadzu LC-VP series instrument.

2. Preparation of materials

(*R*)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#1)^{S2}

Step 1: To a solution of lithium diisopropylamide (LDA) (1.08 M in THF-hexane, 14.6 mL, 15.8 mmol) in tetrahydrofuran (THF) (50 mL) was added dropwise a solution of methyl 2-(4-methoxyphenyl)acetate (2.70 g, 15.0 mmol) in THF (10 mL) at $-78\text{ }^{\circ}\text{C}$, and the mixture was stirred at the same temperature for 0.5 h. To the reaction mixture was added dropwise a solution of 4-methoxybenzyl bromide (3.32 g, 16.5 mmol) in THF (5 mL) at $-78\text{ }^{\circ}\text{C}$. After being stirred at the same temperature for 0.5 h, the reaction mixture was allowed to warm to room temperature and stirred for 0.5 h. The reaction mixture was poured into a saturated aqueous NH_4Cl solution and the resulting mixture was extracted with Et_2O . The extract was washed with water and brine, dried over anhydrous Mg_2SO_4 , and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 5–20% EtOAc in hexane) to give methyl 2,3-bis(4-methoxyphenyl)propanoate (4.35 g).

Step2: To a solution of methyl 2,3-bis(4-methoxyphenyl)propanoate (4.35 g, 14.5 mmol) in a mixed solvent of THF (50 mL) and MeOH (25 mL) was added 2.0 M aqueous LiOH solution (22.5 mL, 45 mmol), and the reaction mixture was stirred at room temperature for 4 h. After addition of 1 M hydrochloric acid (50 mL), the mixture was poured into water and the resulting mixture was extracted with EtOAc . The extract was washed with water and brine, dried over anhydrous Mg_2SO_4 , and concentrated under reduced pressure. The residue was suspended in EtOAc /hexane (1/5, 30 mL). The insoluble material was collected by filtration, washed with EtOAc /hexane (1/5, 20 mL) and dried under reduced pressure to give 2,3-bis(4-methoxyphenyl)propanoic acid (3.75 g, 87%) as a white solid.

Step 3: To a solution of 2,3-bis(4-methoxyphenyl)propanoic acid (1.43 g, 5.0 mmol) and *N,N*-dimethylformamide (DMF) (10 μL) in dichloromethane (DCM) (20 mL) was added dropwise oxalyl chloride (0.76 g, 6.0 mmol), and the mixture was stirred at room temperature for 1 h. The reaction mixture was concentrated under reduced pressure and the residue was dissolved in THF (10 mL) to give an intermediate solution-A. To a solution of (*R*)-4-phenyloxazolidin-2-one (1.06 g, 6.5 mmol) in THF (40 mL) was added dropwise LDA (1.08 M in THF-hexane, 5.5 mL, 5.9 mmol) at $-78\text{ }^{\circ}\text{C}$, and the mixture was stirred at the same temperature for 0.25 h. The above intermediate solution-A was added dropwise to the reaction mixture at $-78\text{ }^{\circ}\text{C}$. After being stirred at the same temperature for 0.25 h, the reaction mixture was allowed to

warm to room temperature and stirred for 1 h. To the reaction mixture was added a saturated aqueous NH_4Cl solution, and the resulting mixture was extracted with EtOAc. The extract was washed with brine, dried over anhydrous Mg_2SO_4 , and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 20–35% EtOAc in hexane). The obtained compound was dissolved in EtOAc (10 mL) and hexane (10 mL) at 70 °C. Hexane (30 mL) was added dropwise to the mixture at 70 °C, and the mixture was stirred at the same temperature for 1 h and at room temperature overnight. The precipitate was collected by filtration, washed with EtOAc/hexane (1/5), and dried under reduced pressure to give (*R*)-3-((*R*)-2,3-bis(4-methoxyphenyl)propanoyl)-4-phenyloxazolidin-2-one (0.76 g, 35%) as a white solid.

Step 4: To a solution of 4.0 M aqueous LiOH solution (0.45 mL, 1.8 mmol) in water (11.5 mL) was added 30% aqueous H_2O_2 solution (0.85 g) under ice-cooling, and the mixture was stirred at the same temperature for 10 min. The mixture was added to a suspension of (*R*)-3-((*R*)-2,3-bis(4-methoxyphenyl)propanoyl)-4-phenyloxazolidin-2-one (0.65 g, 1.5 mmol) in THF (60 mL) under ice-cooling, and the reaction mixture was stirred at the same temperature for 3 h. To the mixture was added a solution of Na_2SO_3 (1.89 g, 15.0 mmol) in water (7.5 mL), and the mixture was stirred at room temperature for 0.5 h. The reaction mixture was poured into 1 M hydrochloric acid (23 mL), and the resulting mixture was extracted EtOAc twice. The combined extracts were washed with brine, dried over anhydrous Mg_2SO_4 , and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 20–35% EtOAc in hexane). The obtained compound (0.37 g) was suspended in EtOAc/hexane (1/3, 4 mL) and the mixture was stirred at room temperature for 1 h. Hexane (2 mL) was added to the mixture, and the insoluble material was collected by filtration, washed with EtOAc/hexane (1/5), and dried under reduced pressure to give (*R*)-2,3-bis(4-methoxyphenyl)propanoic acid (0.34 g, 79%) as a white solid.

Step 5: To a solution of (*R*)-2,3-bis(4-methoxyphenyl)propanoic acid (0.15 g, 0.52 mmol) in THF (5.2 mL) were added *N*-methylmorpholine (0.064 g, 0.63 mmol) and isobutyl chloroformate (0.089 g, 0.66 mmol) at –20 °C, and the mixture was stirred at the same temperature for 0.25 h. To the mixture was added ammonia (2 M in 2-propanol, 2.6 mL, 5.2 mmol) at –20 °C, and the mixture was stirred at the same temperature for 0.5 h. After removal of the insoluble materials by filtration, the filtrate

was concentrated under reduced pressure, and the residue was dissolved in THF (5.2 mL). To the solution were added pyridine (0.17 g, 2.2 mmol) and trifluoroacetic anhydride (0.22 g, 1.0 mmol) under ice-cooling, and the mixture was stirred at the same temperature for 0.5 h. The reaction mixture was poured into 0.2 M hydrochloric acid (7.5 mL), and the resulting mixture was extracted with EtOAc. The extract was washed with water, a saturated aqueous NaHCO₃ solution and brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was dissolved in DCM (5.2 mL) and the solution was cooled to -78 °C. BBr₃ (1.0 M in DCM, 4.5 mL, 4.5 mmol) was added dropwise to the solution at the same temperature, and the mixture was allowed to warm to 0 °C over a period of 4.5 h, and stirred at 0 °C for 1 h. After addition of MeOH (1.0 mL), the mixture was stirred at 0 °C for 10 min. The mixture was poured into water, and the resulting mixture was extracted with EtOAc. The extract was washed with brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was suspended in EtOAc/hexane (1/1, 3 mL), and the suspension was stirred at 70 °C for 0.5 h, and cooled to room temperature. To the mixture was added hexane (1.0 mL), and the mixture was stirred at room temperature for 1 h. The insoluble material was collected by filtration, washed with EtOAc/hexane (1/4), and dried under reduced pressure to give the title compound (0.082 g, 65%) as an off-white solid (mp 189–192 °C). ¹H NMR (CD₃OD) δ 2.95–3.07 (m, 2H), 4.04–4.11 (m, 1H), 6.65–6.71 (m, 2H), 6.72–6.78 (m, 2H), 6.93–7.00 (m, 2H), 7.05–7.12 (m, 2H); ¹³C NMR (CD₃OD) δ 39.9, 42.3, 116.1, 116.6, 122.4, 127.9, 129.1, 129.8, 131.5, 157.6, 158.4; HRMS (ESI) calcd for C₁₅H₁₇N₂O₂ (M + NH₄)⁺ 257.1285, found 257.1286; calcd for C₁₅H₁₂NO₂ (M - H)⁻ 238.0874, found 238.0875; HPLC purity: 99.7% (Retention time (t_R) 13.7 min); Optical purity: 99.0% ee.

(S)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#2)^{S2}

The title compound was prepared in a manner similar to that described for the synthesis of #1, using (S)-4-phenyloxazolidin-2-one instead of (R)-4-phenyloxazolidin-2-one, and obtained as a white solid (mp 190–192 °C). ¹H NMR (CD₃OD) δ 2.95–3.07 (m, 2H), 4.03–4.12 (m, 1H), 6.65–6.71 (m, 2H), 6.72–6.78 (m, 2H), 6.93–7.00 (m, 2H), 7.05–7.12 (m, 2H); ¹³C NMR (CD₃OD) δ 39.9, 42.3, 116.1, 116.6, 122.4, 127.9, 129.1, 129.8, 131.5, 157.6, 158.4; HRMS (ESI) calcd for C₁₅H₁₇N₂O₂ (M + NH₄)⁺ 257.1285, found 257.1286; calcd for C₁₅H₁₂NO₂ (M - H)⁻

238.0874, found 238.0874; HPLC purity: 99.5% (t_R 13.7 min); Optical purity: 98.5% ee.

(*R*)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#3)

Step 1: To a mixture of 2-(4-methoxyphenyl)acetic acid (1.00 g, 6.0 mmol), (*R*)-benzyloxazolidin-2-one (0.53 g, 3.0 mmol) and triethylamine (1.21 g, 12.0 mmol) in toluene (8 mL) was added dropwise a solution of pivaloyl chloride (0.72 g, 6.0 mmol) in toluene (2 mL) at 80 °C. The mixture was refluxed for 16 h, and then cooled to room temperature. To the mixture was added 1 M hydrochloric acid (8 mL), and the resulting mixture was extracted with EtOAc. The extract was washed with a saturated aqueous NaHCO₃ solution twice and brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 15–33% EtOAc in hexane). The obtained compound (0.59 g) was dissolved in toluene (4.8 mL) at 70 °C. Hexane (7.2 mL) was added dropwise to the solution at the same temperature, and the mixture was stirred at the same temperature for 1 h and at room temperature overnight. The precipitate was collected by filtration, washed with toluene/hexane (1/4), and dried under reduced pressure to give (*R*)-4-benzyl-3-(2-(4-methoxyphenyl)acetyl)oxazolidin-2-one (0.44 g, 45%) as a white solid.

Step 2: To a solution of (*R*)-4-benzyl-3-(2-(4-methoxyphenyl)acetyl)oxazolidin-2-one (0.44 g, 1.4 mmol) in THF (4.3 mL) was added dropwise sodium hexamethyldisilazide (1.0 M in THF, 1.4 mL, 1.4 mmol) at –78 °C, and the mixture was stirred at the same temperature for 1 h. To the mixture was added dropwise a solution of 2-chloro-4-methoxybenzyl bromide (0.64 g, 2.7 mmol) in THF (1.4 mL) at the same temperature. The mixture was allowed to warm to room temperature and stirred for 5 h. To the reaction mixture was added water, and the resulting mixture was extracted with EtOAc. The extract was washed with brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 15–30% EtOAc in hexane) to give (*R*)-4-benzyl-3-((*R*)-3-(2-chloro-4-methoxyphenyl)-2-(4-methoxyphenyl)propanoyl)oxazolidin-2-one (0.55 g, 85%) as a white amorphous solid.

Step 3: To a solution of 4.0 M aqueous LiOH solution (0.35 mL, 1.4 mmol) in water (8.9 mL) was added 30% aqueous H₂O₂ solution (0.65 g) under ice-cooling, and the

mixture was stirred at the same temperature for 10 min. The mixture was added to a solution of (*R*)-4-benzyl-3-((*R*)-3-(2-chloro-4-methoxyphenyl)-2-(4-methoxyphenyl)propanoyl)oxazolidin-2-one (0.55 g, 1.2 mmol) in THF (46 ml) under ice-cooling, and the reaction mixture was stirred at the same temperature for 3 h. To the mixture was added a solution of Na₂SO₃ (1.46 g, 11.5 mmol) in water (5.8 mL), and the mixture was stirred at room temperature for 0.5 h. The reaction mixture was poured into 1 M hydrochloric acid (18 mL), and the resulting mixture was extracted EtOAc twice. The combined extracts were washed with brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (gradient: 20–35% EtOAc in hexane). The obtained compound was suspended in EtOAc/hexane (1/3, 4 mL) and the mixture was stirred at room temperature for 1 h. Hexane (2 mL) was added to the mixture, and the insoluble material was collected by filtration, washed with EtOAc/Hexane (1/5), and dried under reduced pressure to give (*R*)-3-(2-chloro-4-methoxyphenyl)-2-(4-methoxyphenyl)propanoic acid (0.27 g, 74%) as a white solid.

Step 4: To a solution of (*R*)-3-(2-chloro-4-methoxyphenyl)-2-(4-methoxyphenyl)propanoic acid (0.17 g, 0.52 mmol) in THF (5.2 mL) were added *N*-methylmorpholine (0.064 g, 0.63 mmol) and isobutyl chloroformate (0.089 g, 0.66 mmol) at –20 °C, and the mixture was stirred at the same temperature for 0.25 h. To the mixture was added ammonia (2 M in 2-propanol, 2.6 mL, 5.2 mmol) at –20 °C, and the mixture was stirred at the same temperature for 0.5 h. After removal of the insoluble materials by filtration, the filtrate was concentrated under reduced pressure, and the residue was dissolved in THF (5.2 mL). To the solution were added pyridine (0.17 g, 2.2 mmol) and trifluoroacetic anhydride (0.22 g, 1.0 mmol) under ice-cooling, and the mixture was stirred at the same temperature for 0.5 h. The reaction mixture was poured into 0.2 M hydrochloric acid (7.5 mL), and the resulting mixture was extracted with EtOAc. The extract was washed with water, a saturated aqueous NaHCO₃ solution and brine, dried over anhydrous Mg₂SO₄, and concentrated under reduced pressure. The residue was dissolved in DCM (5.2 mL) and the solution was cooled to –78 °C. BBr₃ (1.0 M in DCM, 4.5 mL, 4.5 mmol) was added dropwise to the solution at the same temperature, and the mixture was allowed to warm to 0 °C over a period of 3 h, and stirred at 0 °C for 1.5 h. After addition of MeOH (1.0 mL), the mixture was stirred at 0 °C for 10 min. The mixture was poured into water, and the resulting mixture was extracted with EtOAc. The extract was washed with brine, dried

over anhydrous Mg_2SO_4 , and concentrated under reduced pressure. The residue was dissolved in EtOAc (1.5 mL) at 70 °C. To the solution was added hexane (1.5 mL), and the mixture was stirred at 70 °C for 0.25 h, and cooled to room temperature. To the mixture was added hexane (1.0 mL), and the mixture was stirred at room temperature for 1 h. The precipitate was collected by filtration, washed with EtOAc/hexane (1/4), and dried under reduced pressure to give the title compound (0.101 g, 71%) as an off-white solid (mp 209–212 °C). ^1H NMR (CD_3OD) δ 3.07–3.21 (m, 2H), 4.13 (dd, $J = 6.9, 8.7$ Hz, 1H), 6.64 (dd, $J = 2.5, 8.3$ Hz, 1H), 6.74–6.81 (m, 2H), 6.82 (d, $J = 2.5$ Hz, 1H), 7.03 (d, $J = 8.3$ Hz, 1H), 7.10–7.17 (m, 2H); ^{13}C NMR (CD_3OD) δ 38.2, 40.2, 115.3, 116.7, 117.2, 122.0, 126.3, 127.7, 129.7, 133.5, 135.5, 158.6, 159.0; HRMS (ESI) calcd for $\text{C}_{15}\text{H}_{16}\text{ClN}_2\text{O}_2$ ($\text{M} + \text{NH}_4$) $^+$ 291.0895, found 291.0895; calcd for $\text{C}_{15}\text{H}_{11}\text{ClNO}_2$ ($\text{M} - \text{H}$) $^-$ 272.0484, found 272.0486; HPLC purity: 99.0% (t_{R} 16.0 min); Optical purity: 98.1% ee.

(S)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#4)

The title compound was prepared in a manner similar to that described for the synthesis of **#3**, using (S)-benzyloxazolidin-2-one instead of (R)-benzyloxazolidin-2-one, and obtained as an off-white solid (mp 209–212 °C). ^1H NMR (CD_3OD) δ 3.07–3.21 (m, 2H), 4.13 (dd, $J = 6.9, 8.7$ Hz, 1H), 6.64 (dd, $J = 2.5, 8.4$ Hz, 1H), 6.74–6.80 (m, 2H), 6.82 (d, $J = 2.5$ Hz, 1H), 7.03 (d, $J = 8.4$ Hz, 1H), 7.10–7.17 (m, 2H); ^{13}C NMR (CD_3OD) δ 38.2, 40.2, 115.3, 116.7, 117.2, 122.0, 126.3, 127.7, 129.7, 133.5, 135.5, 158.6, 159.0; HRMS (ESI) calcd for $\text{C}_{15}\text{H}_{16}\text{ClN}_2\text{O}_2$ ($\text{M} + \text{NH}_4$) $^+$ 291.0895, found 291.0894; calcd for $\text{C}_{15}\text{H}_{11}\text{ClNO}_2$ ($\text{M} - \text{H}$) $^-$ 272.0484, found 272.0485; HPLC purity: 99.4% (t_{R} 16.1 min); Optical purity: > 99% ee.

(R)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#5)

The title compound was prepared in a manner similar to that described for the synthesis of **#3**, using 2-(2-chloro-4-methoxyphenyl)acetic acid and 4-methoxybenzyl bromide instead of 2-(4-methoxyphenyl)acetic acid and 2-chloro-4-methoxybenzyl bromide, respectively, and obtained as an off-white solid (mp 210–214 °C). ^1H NMR (CD_3OD) δ 2.99 (dd, $J = 8.5, 13.7$ Hz, 1H), 3.07 (dd, $J = 5.8, 13.7$ Hz, 1H), 4.45 (dd, $J = 5.8, 8.5$ Hz, 1H), 6.66–6.72 (m, 2H), 6.73 (dd, $J = 2.5, 8.5$ Hz, 1H), 6.86 (d, $J = 2.5$ Hz, 1H), 6.96–7.02 (m, 2H), 7.20 (d, $J = 8.5$ Hz, 1H); ^{13}C NMR (CD_3OD) δ 37.6, 40.2, 115.8, 116.2, 117.6, 121.5, 124.8, 128.7, 131.2, 131.5, 134.2, 157.8, 159.7; HRMS

(ESI) calcd for $C_{15}H_{16}ClN_2O_2$ ($M + NH_4$)⁺ 291.0895, found 291.0890; calcd for $C_{15}H_{11}ClNO_2$ ($M - H$)⁻ 272.0484, found 272.0485; HPLC purity: 96.7% (t_R 16.1 min); Optical purity: > 99% ee.

(S)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#6)

The title compound was prepared in a manner similar to that described for the synthesis of **#3**, using 2-(2-chloro-4-methoxyphenyl)acetic acid, (*S*)-benzyloxazolidin-2-one and 4-methoxybenzyl bromide instead of 2-(4-methoxyphenyl)acetic acid, (*R*)-benzyloxazolidin-2-one and 2-chloro-4-methoxybenzyl bromide, respectively, and obtained as an off-white solid (mp 211–214 °C). ¹H NMR (CD₃OD) δ 2.99 (dd, *J* = 8.5, 13.6 Hz, 1H), 3.07 (dd, *J* = 5.7, 13.6 Hz, 1H), 4.45 (dd, *J* = 5.7, 8.5 Hz, 1H), 6.66–6.72 (m, 2H), 6.73 (dd, *J* = 2.5, 8.5 Hz, 1H), 6.86 (d, *J* = 2.5 Hz, 1H), 6.96–7.03 (m, 2H), 7.20 (d, *J* = 8.5 Hz, 1H); ¹³C NMR (CD₃OD) δ 37.6, 40.2, 115.8, 116.2, 117.5, 121.5, 124.8, 128.7, 131.2, 131.5, 134.2, 157.8, 159.6; HRMS (ESI) calcd for $C_{15}H_{16}ClN_2O_2$ ($M + NH_4$)⁺ 291.0895, found 291.0890; calcd for $C_{15}H_{11}ClNO_2$ ($M - H$)⁻ 272.0484, found 272.0485; HPLC purity: 98.4% (t_R 16.1 min); Optical purity: > 99% ee.

2,3-Bis(4-hydroxy-phenyl)propionitrile (racemate of #1(#2))

The title compound was purchased from SIGMA (SIGMA H5915, purity ≥ 98%).

3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (racemate of #3(#4))

The title compound was prepared in a manner similar to that described for the synthesis of **#1** (step 1, 2 and 5), using 2-chloro-4-methoxybenzyl bromide instead of 4-methoxybenzyl bromide, and obtained as an off-white solid (mp 199–202 °C). ¹H NMR (CD₃OD) δ 3.07–3.21 (m, 2H), 4.13 (dd, *J* = 7.0, 8.8 Hz, 1H), 6.64 (dd, *J* = 2.5, 8.5 Hz, 1H), 6.74–6.80 (m, 2H), 6.82 (d, *J* = 2.5 Hz, 1H), 7.03 (d, *J* = 8.5 Hz, 1H), 7.10–7.17 (m, 2H); ¹³C NMR (CD₃OD) δ 38.2, 40.2, 115.3, 116.7, 117.2, 122.0, 126.3, 127.7, 129.7, 133.5, 135.5, 158.6, 159.0; HRMS (ESI) calcd for $C_{15}H_{16}ClN_2O_2$ ($M + NH_4$)⁺ 291.0895, found 291.0893; calcd for $C_{15}H_{11}ClNO_2$ ($M - H$)⁻ 272.0484, found 272.0487; HPLC purity: 99.4% (t_R 16.1 min).

2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (racemate of

#5(#6))

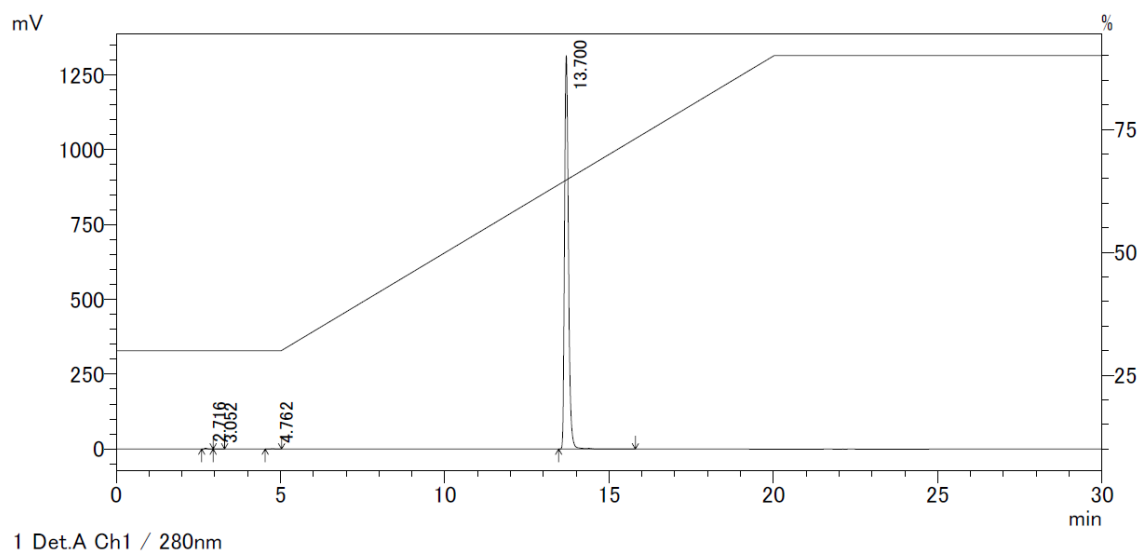
The title compound was prepared in a manner similar to that described for the synthesis of **#1** (step 1, 2 and 5), using methyl 2-(2-chloro-4-methoxyphenyl)acetate instead of methyl 2-(4-methoxyphenyl)acetate, and obtained as an off-white solid (mp 192–195 °C) in 68% yield. ¹H NMR (CD₃OD) δ 2.99 (dd, *J* = 8.5, 13.7 Hz, 1H), 3.07 (dd, *J* = 5.8, 13.7 Hz, 1H), 4.45 (dd, *J* = 5.8, 8.5 Hz, 1H), 6.66–6.72 (m, 2H), 6.73 (dd, *J* = 2.5, 8.5 Hz, 1H), 6.86 (d, *J* = 2.5 Hz, 1H), 6.96–7.02 (m, 2H), 7.20 (d, *J* = 8.5 Hz, 1H); ¹³C NMR (CD₃OD) δ 37.6, 40.2, 115.8, 116.2, 117.6, 121.5, 124.8, 128.7, 131.2, 131.5, 134.2, 157.8, 159.7; HRMS (ESI) calcd for C₁₅H₁₆ClN₂O₂ (M + NH₄)⁺ 291.0895, found 291.0894; calcd for C₁₅H₁₁ClNO₂ (M – H)⁻ 272.0484, found 272.0484; HPLC purity: 99.8% (*t*_R 16.1 min).

3. HPLC chromatograms for chemical purity of tested compounds

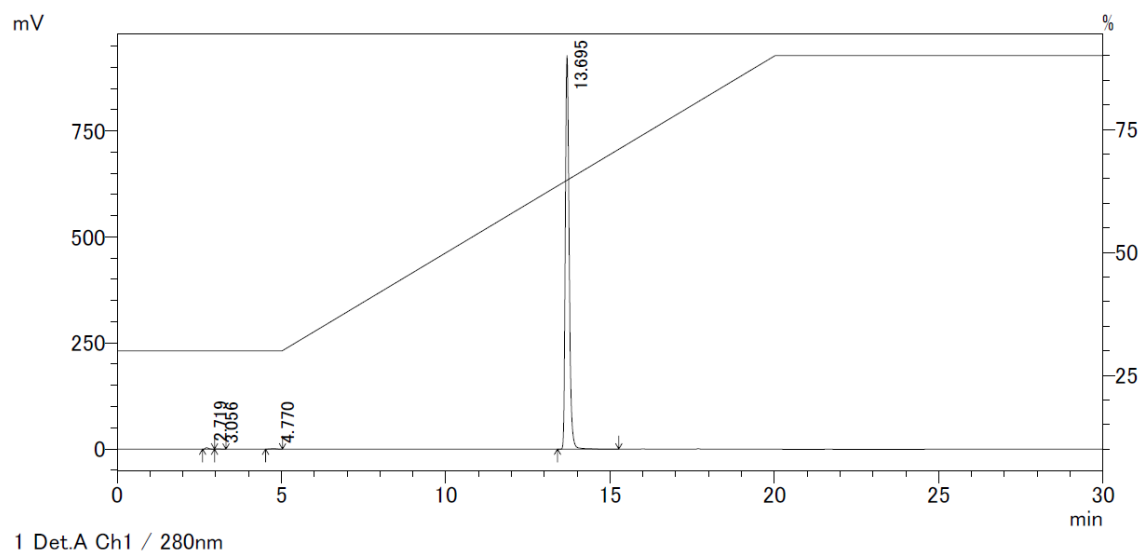
conditions

analytical column	InertSustain C18, 4.6 mm × 250 mm, 5 μm (GL Sciences Inc.)
mobile phase	MeCN in 10 mM aq AcONH ₄
gradient	30% (0–5 min), linear gradient of 30–90% (5–20 min), 90% (20–30 min)
flow rate	1.0 mL/min
column temperature	35 °C
detection wavelength	280 nm

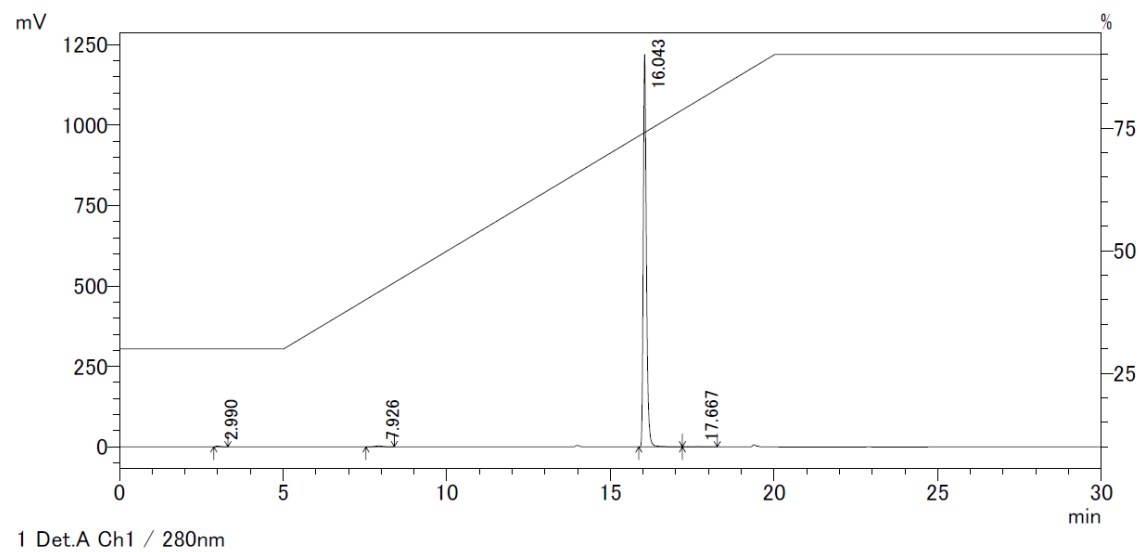
(R)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#1): 99.735% (Retention time: t_R 13.700 min)



(S)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#2): 99.505% (t_R 13.695 min)

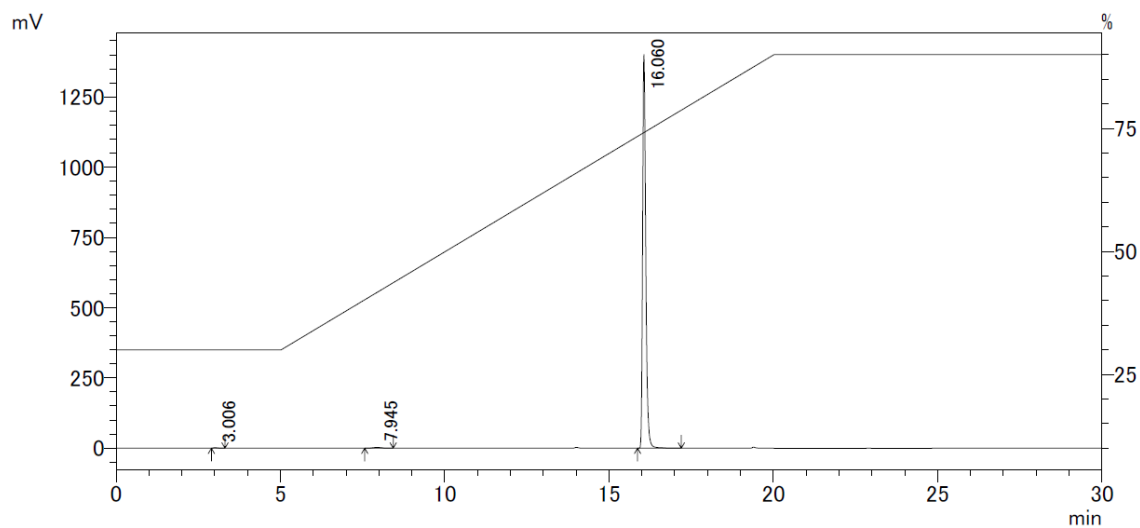


(R)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#3): 99.021% (t_R 16.043 min)



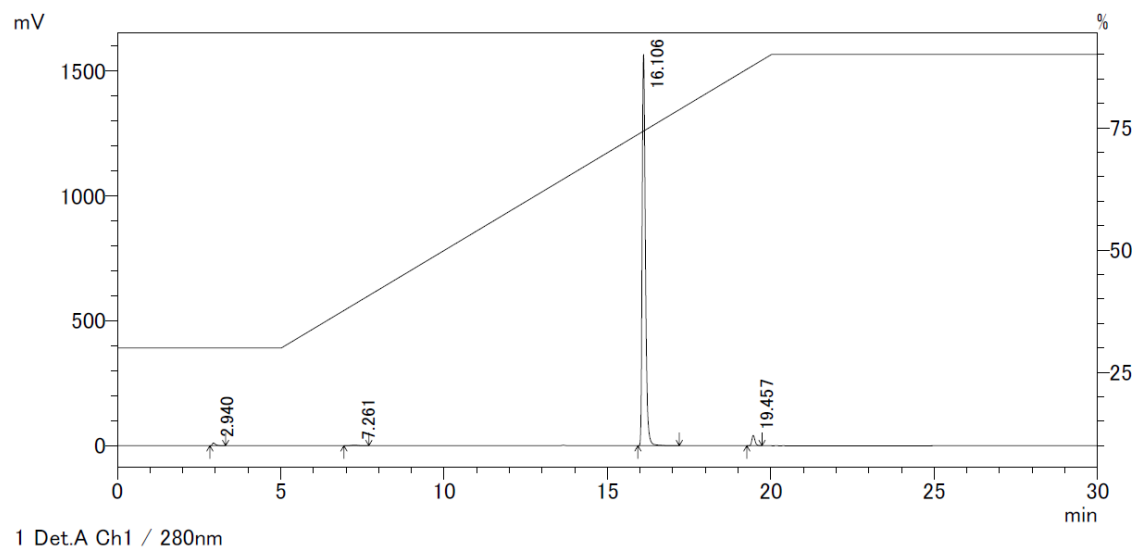
(S)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#4):

99.389% (t_R 16.060 min)



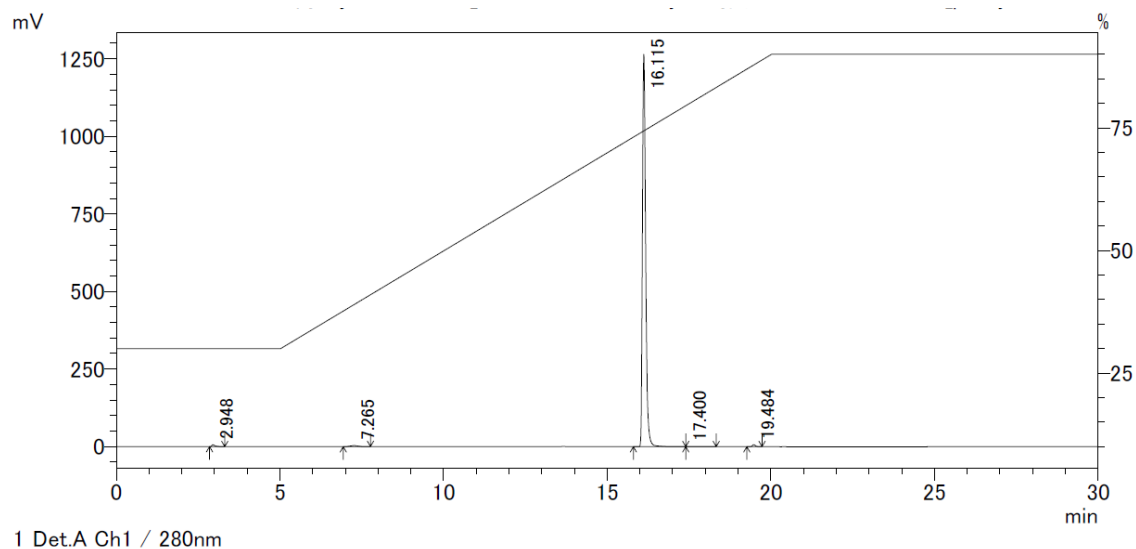
(R)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#5):

96.673% (t_R 16.106 min)



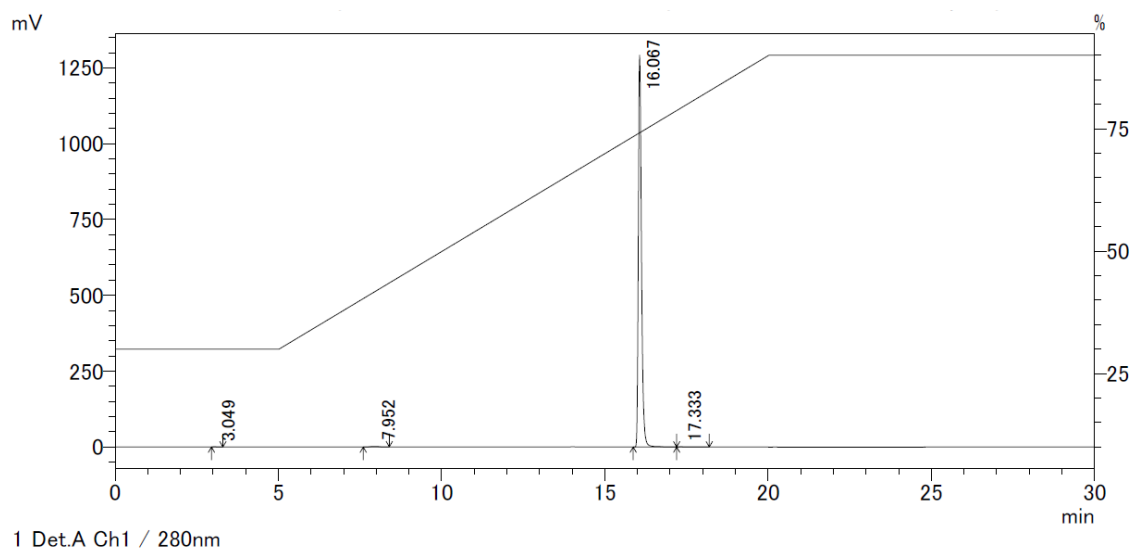
(S)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#6):

98.399% (t_R 16.115 min)

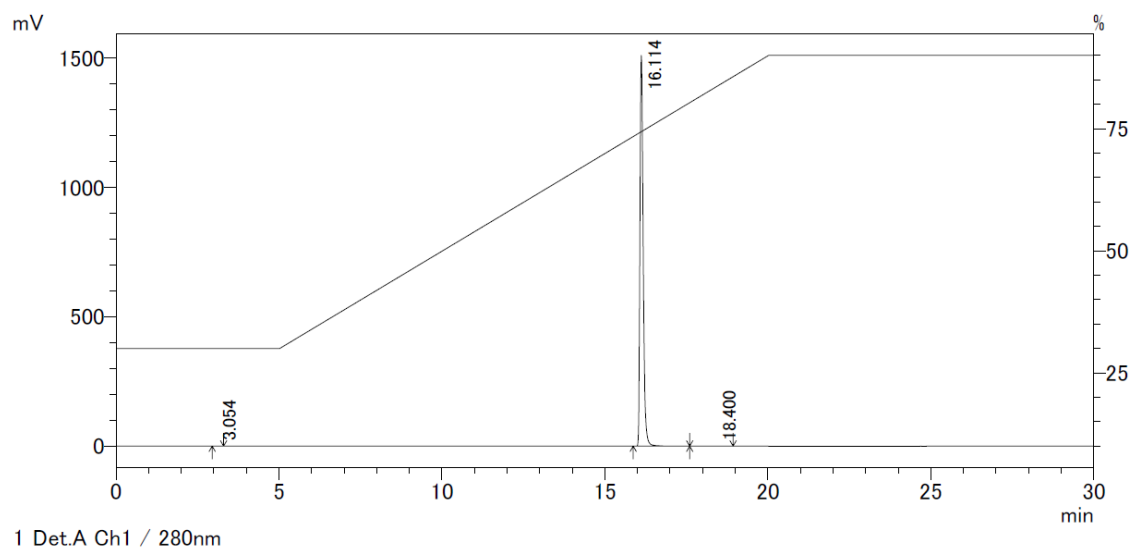


3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (racemate of

#3(#4): 99.437% (t_R 16.067 min)



2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (racemate of #5(#6)): 99.844% (t_R 16.114 min)

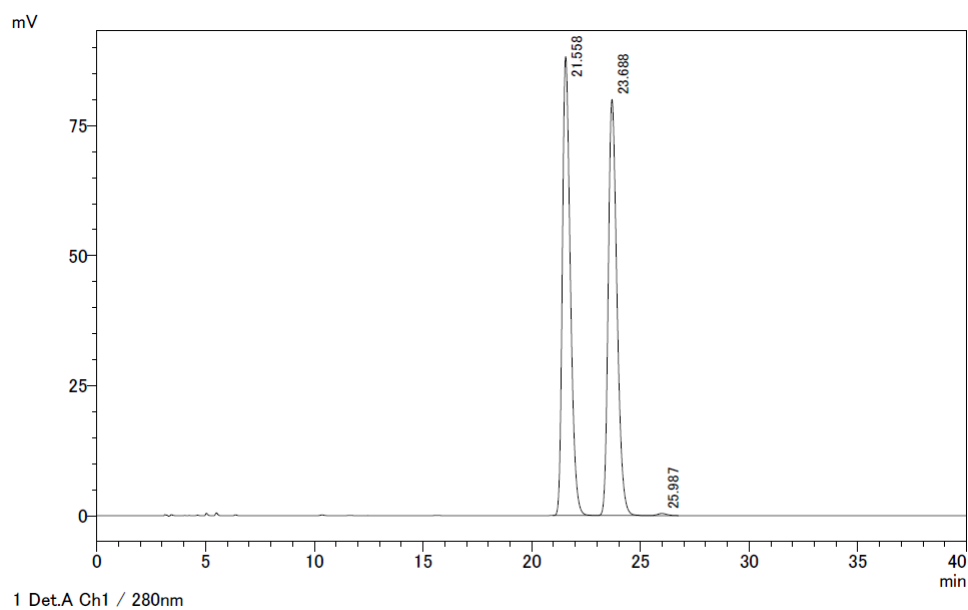


4. HPLC chromatograms for optical purity of tested compounds

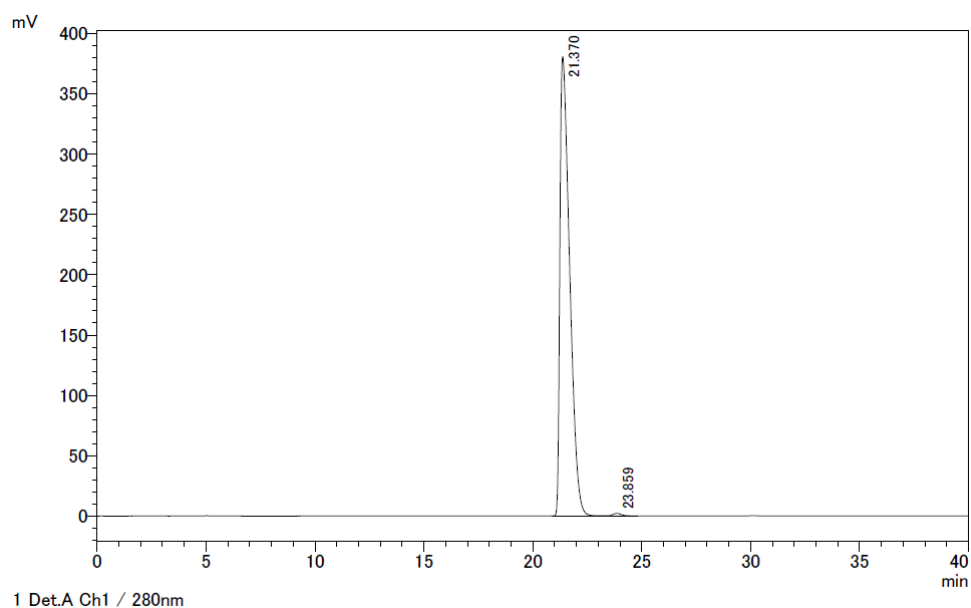
conditions

analytical column	CHIRALPAK OJ-3, 4.6 mm × 250 mm, 3 μm (DAICEL CORPORATION)
mobile phase	20% EtOH in hexane
flow rate	1.0 mL/min
column temperature	30 °C
detection wavelength	280 nm

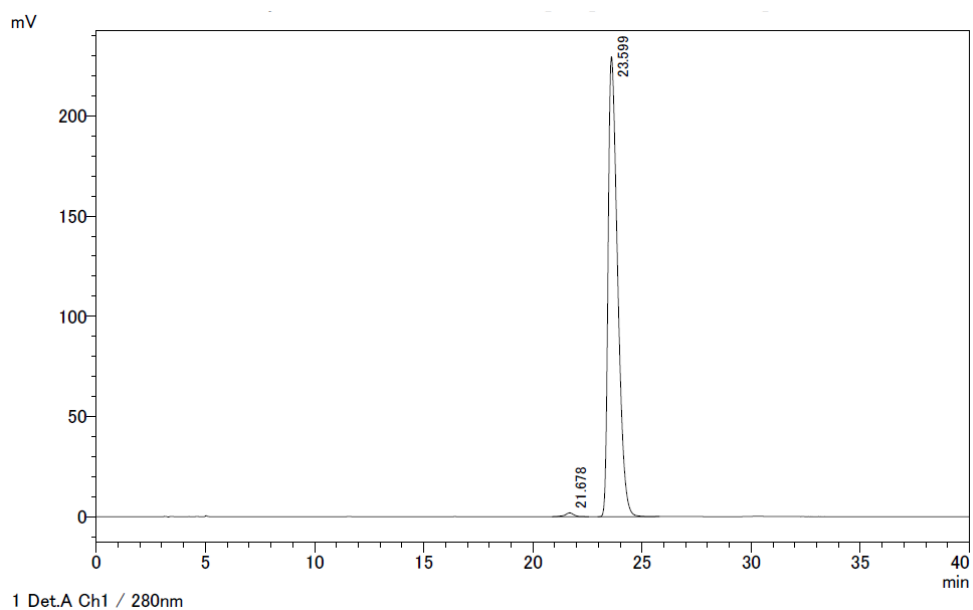
2,3-Bis(4-hydroxy-phenyl)propionitrile (racemate of **#1(#2)**): 49.861% (t_R 21.558 min), 49.880% (t_R 23.688 min)



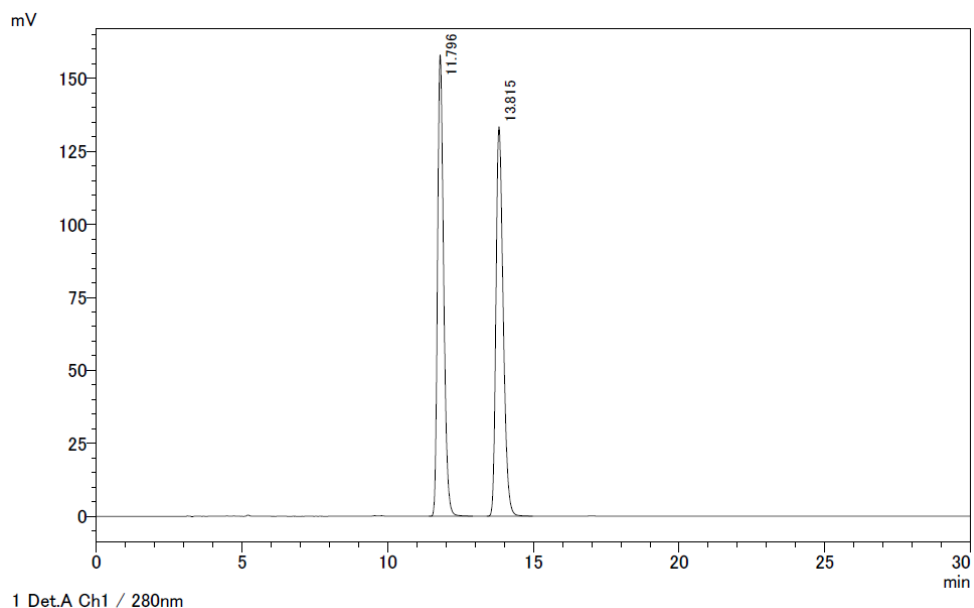
(R)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#1): 99.480% (t_R 21.370 min), 0.520% (t_R 23.859 min)



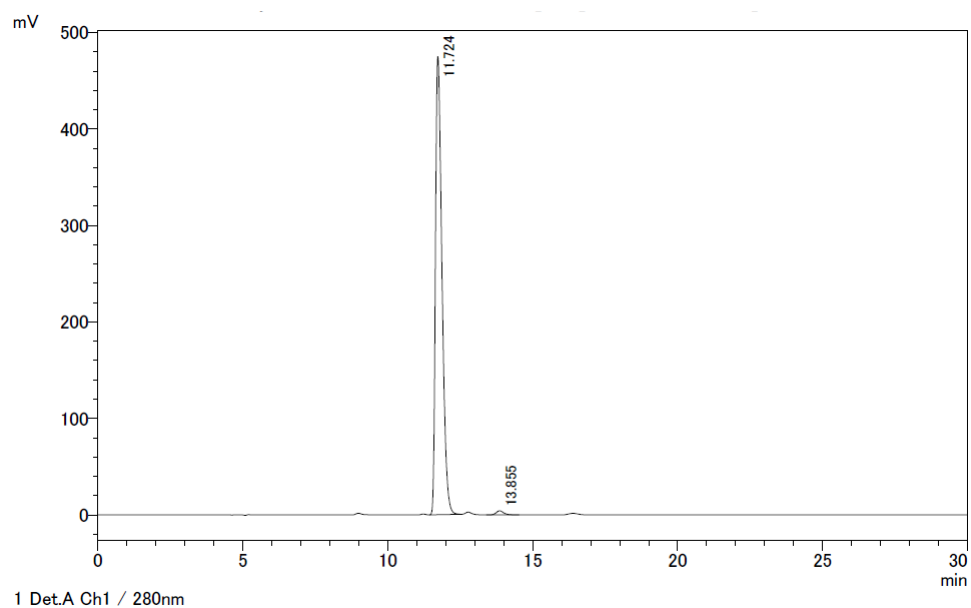
(S)-2,3-Bis(4-hydroxyphenyl)propanenitrile (#2): 0.772% (t_R 21.678 min), 99.228% (t_R 23.599 min)



3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (racemate of #3(#4)): 50.008% (t_R 11.796 min), 49.992% (t_R 13.815 min)

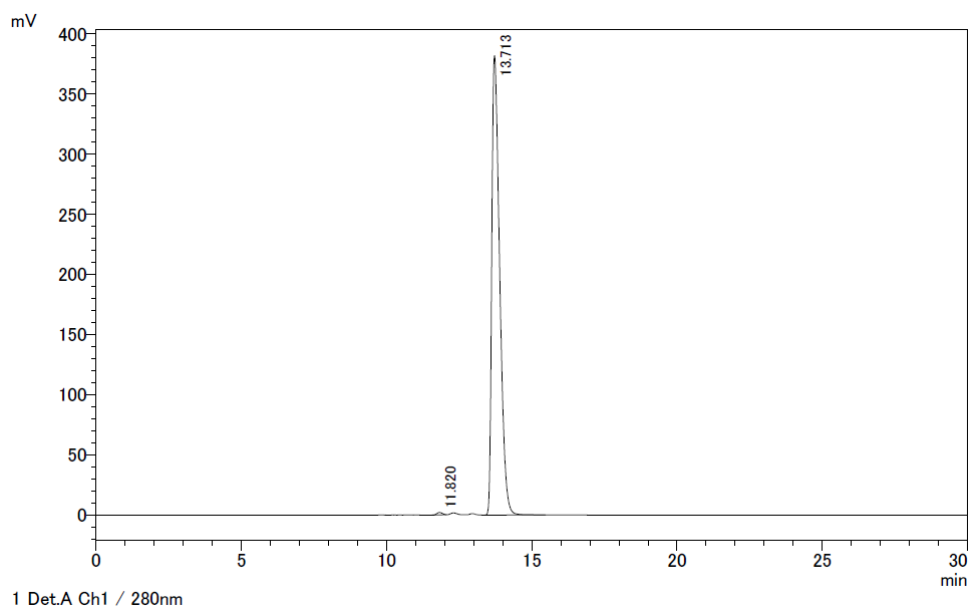


(R)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#3): 99.028% (t_R 11.724 min), 0.972% (t_R 13.855 min)



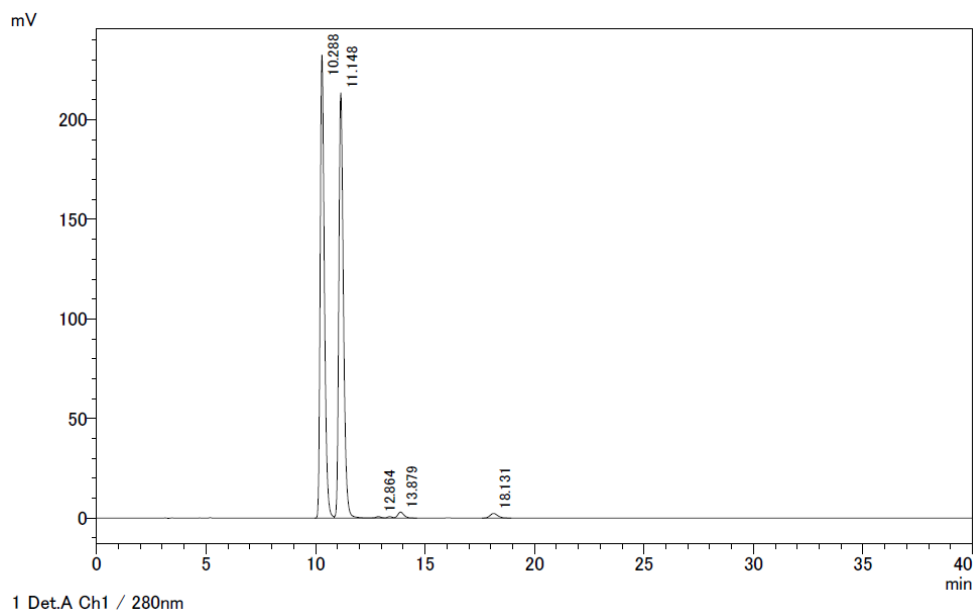
(S)-3-(2-Chloro-4-hydroxyphenyl)-2-(4-hydroxyphenyl)propanenitrile (#4):

0.355% (t_R 11.820 min), 99.645% (t_R 13.713 min)

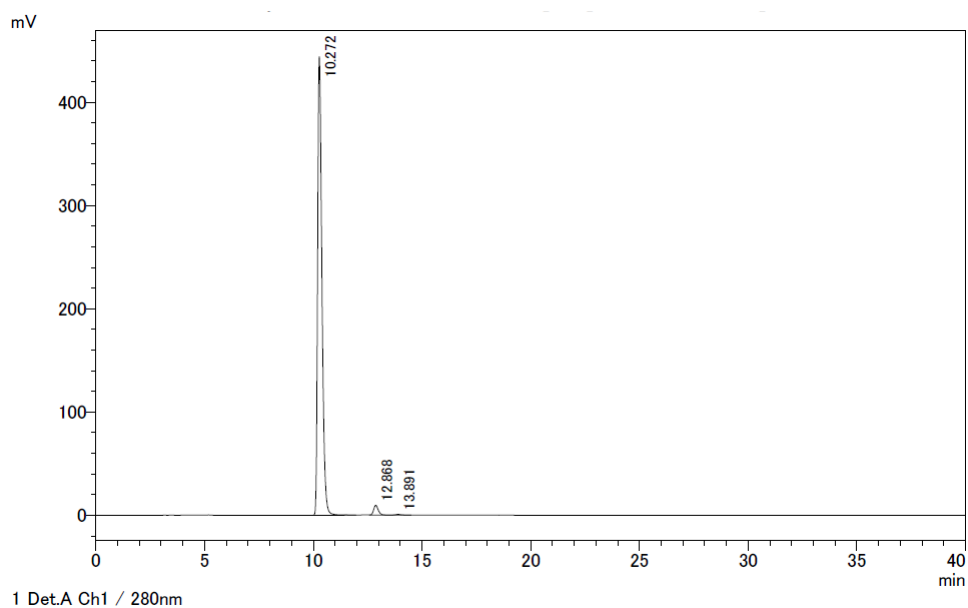


2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (racemate of

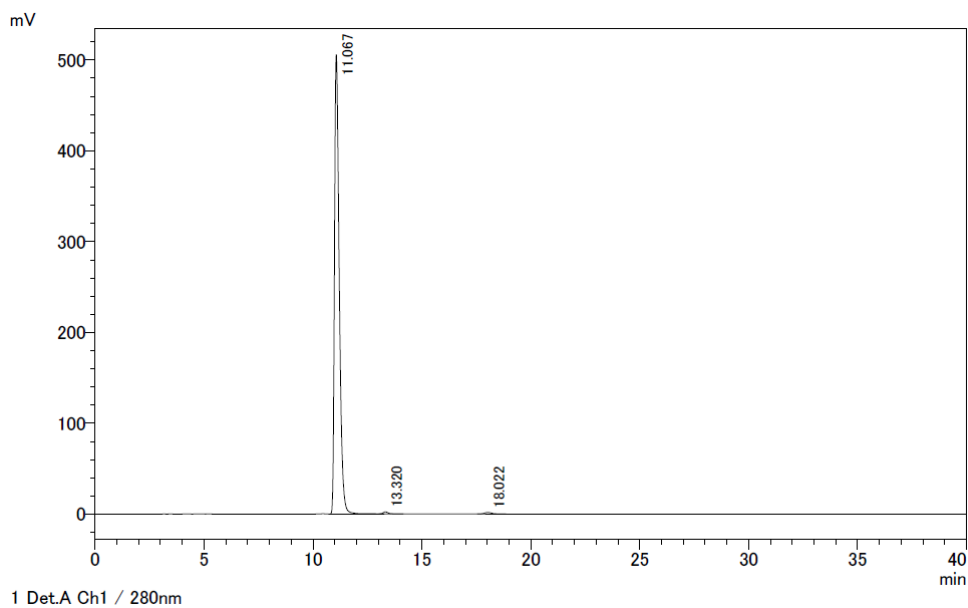
#5(#6)): 48.711% (t_R 10.288 min), 49.300% (t_R 11.148 min)



(R)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#5): single enantiomer



(S)-2-(2-Chloro-4-hydroxyphenyl)-3-(4-hydroxyphenyl)propanenitrile (#6): single enantiomer



References (Chemistry)

- S1. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. *Organometallics* **2010**, *29*, 2176.
- S2. Carroll, V. M.; Jeyakumar, M.; Carlson, K. E.; Katzenellenbogen, J. A. *J. Med. Chem.* **2012**, *55*, 528.