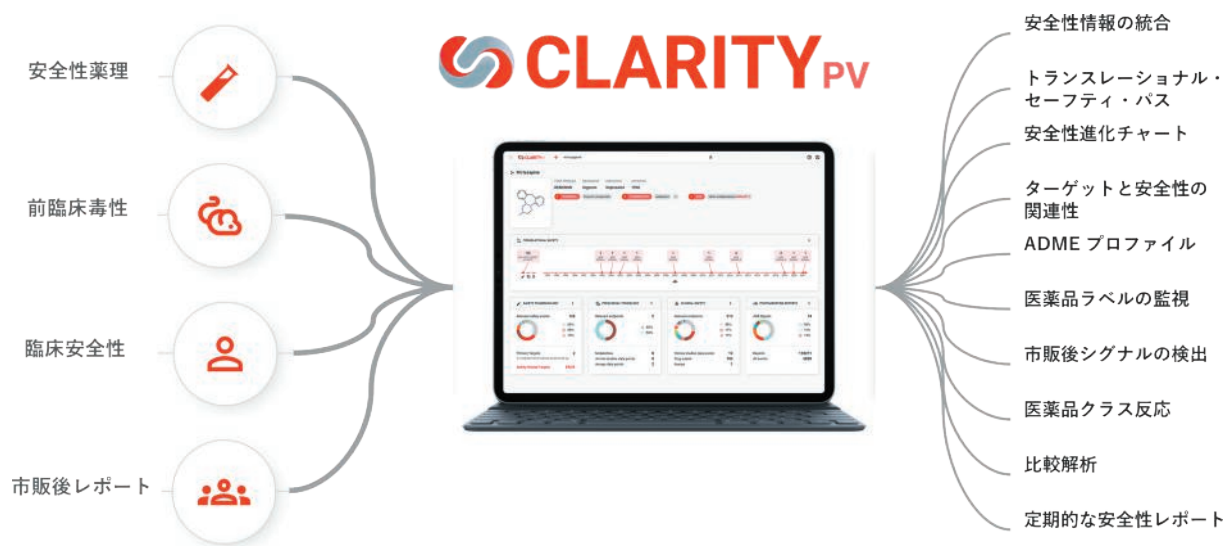


## トランスレーショナル・セーフティ・プラットフォーム CLARITY PV リリース

Chemotargets 社から、新しくトランスレーショナル・セーフティとファーマコビジランスのための研究支援システム CLARITY PV がリリースされました。CLARITY PV は、約 1 万件の医薬品に対して、安全性薬理、前臨床毒性、臨床安全性、市販後レポートの情報をシームレスに接続し、全ての開発フェーズ間で安全性シグナルを追跡可能にします。



### CONTENTS

#### 新製品情報

トランスレーショナル・セーフティ・プラットフォーム CLARITY PV	CLARITY PV リリース	..... 2
遺伝子発現データベース GENEVESTIGATOR	Sample Composition Tools と掲載データのアップデート	..... 4

#### セミナー情報

材料設計支援システム SciMAPS	SciMeeting2021	..... 5
--------------------	----------------	---------

#### 技術情報

統合計算化学システム MOE	REINVENT インターフェースの紹介	..... 6
BioSolveIT 社製創薬支援ツール	Enamine REALSpace からの RAF キナーゼ阻害剤の探索	..... 8
材料設計支援統合システム MedeA	MedeA-VASP による金属表面への分子吸着の評価	..... 10
低分子立体配座発生ツール Corina Classic	CORINA Classic の製品概要	..... 12

トランスレーショナル・セーフティ・プラットフォーム CLARITY PV

# CLARITY PV リリース

CLARITY PV は、医薬品の安全性薬理研究から市販後調査までのすべての毒性・安全性データを統合した、トランスレーショナル・セーフティとファーマコビジランスのための新しいプラットフォームです。前臨床毒性から臨床安全性、更には市販後レポートまで医薬品のライフタイム全体にわたる安全性に関するデータを、高度にキュレーションし、また構造化して収載しています。全ての医薬品の開発フェーズにおけるトレーサビリティと安全性シグナルをシームレスにつなげることができます。

## トランスレーショナル・セーフティの問題

医薬品開発において毒性・安全性データは、in vitro 安全性薬理試験、in vitro スクリーニング、in vivo 安全性評価、臨床実験、市販後調査など、様々なフェーズで得られます。前臨床試験フェーズで得られた毒性・安全性データは、その後の臨床試験、更には市販後調査までトランスレーショナル・セーフティの対象となります。

医薬品開発の各フェーズで得られる毒性・安全性のうち、どれが他のステージに影響し、どれが影響しないのか同定することは、トランスレーショナル・セーフティを研究する上で非常に重要です。

## CLARITY PV のソリューション

CLARITY PV は、トランスレーショナル・セーフティの問題に対して、以下のソリューションを提供します。

### データの統合

医薬品のライフタイム全体にわたる、in vitro で行われる安全性薬理研究、動物実験などの前臨床毒性試験、臨床安全性試験におけるメディカルレビュー、市販後調査における継続的なレポートの全てのデータを統合します。

### レポート

定期的な安全性レポートの収集を支援します。

- ・市販後のデータ更新への迅速なアクセス
- ・臨床安全性イベントの市販後調査
- ・市販後シグナルの層別認識（地域別、年齢別、性別）
- ・有害事象や反応のレポート数の速度と進化

### 予測

市販後の副作用（ADR）シグナルを予測することで、以下の各項目の支援を行います。

- ・重要なマーカーの同定
- ・トランスレーショナル・セーフティ問題の検出
- ・市販後の新薬の安全性傾向の早期の反映
- ・薬と副作用の調査リストの定義

### 解析

以下のユーザーフレンドリーな可視化解析ツールを提供します。

- ・層別分析
- ・トランスレーショナル・セーフティ
- ・比較分析
- ・安全性予測

## トランスレーショナル・チャート

ここで、グリタゾン系に属する Pioglitazone、Rosiglitazone、Trogliatzone の安全性評価に関する比較分析の例を示します。一般的に、これらの分子構造は非常によく似ているため、その安全性を正しく評価することは難しく、多くのプラットフォームでは似た結果を返します。

CLARITY PV による Pioglitazone、Rosiglitazone、Trogliatzone の安全性レポートに関するトランスレーショナル進化チャート（図1）では、これらの医薬品はそれぞれ、泌尿器系、循環器系、消化器系の有害事象レポートが多く出されており、それぞれの有害事象のプロファイルが互いに大きく異なることが分かります。

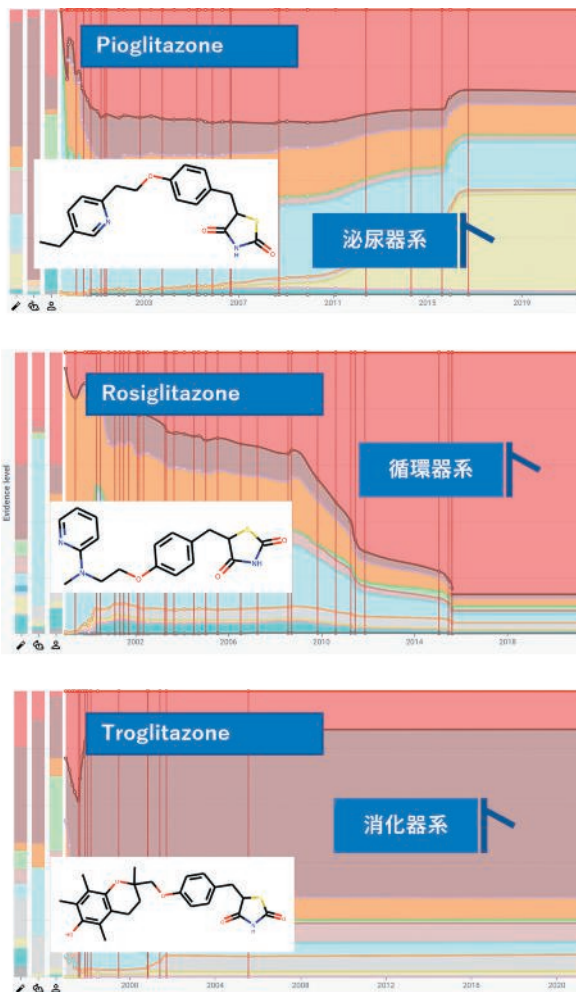


図1. グリタゾン系のトランスレーショナル・チャート。横軸は西暦、縦軸は器官系ごとの安全性レポートの割合。



## ■ Cetuximab の層別分析

CLARITY PV には、低分子医薬品だけでなく抗体医薬品などの生物製剤の情報も含まれています。生物製剤データの検索には、医薬品名やブランド名を使うことができます。

Cetuximab は、2004 年に承認されたヒト上皮増殖因子受容体 (EGFR) を標的とするマウスとヒト IgG1 のキメラ型モノクローナル抗体です。Cetuximab の皮膚毒性に関する有害事象レポートのチャートでは、2006 年に初めの副作用 (ADR) レポートが報告され、2007 年には新しい ADR Warning として検出されました。その後、数年にわたりレポートの数が増えて、2010 年に新しい ADE シグナルとして検出され、その後継続的にレポートの件数は増加傾向にあります (図 2)。

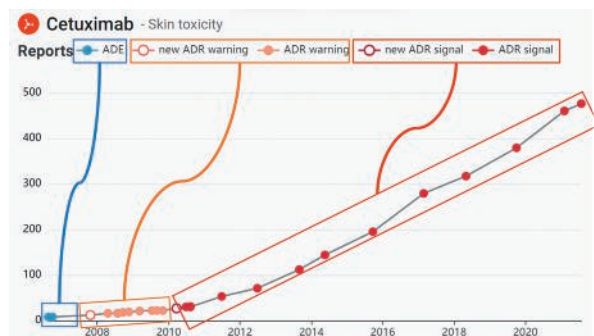


図 2. Cetuximab の皮膚毒性の進化チャート。青：有害事象 (ADE) レポートの報告、橙：ADE warning、赤：ADE シグナル。

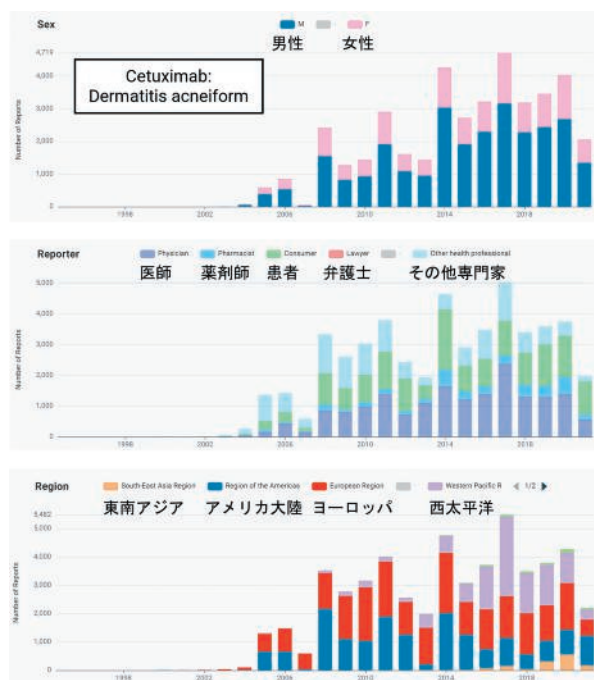


図 3. Cetuximab 層別有害事象レポート数

Cetuximab のざ瘡様皮膚炎 (Dermatitis acneiform) の層別分析では、性別では男性、提出者では医師、また、地域別では日本を含む西太平洋地域から、比較的多くのレポートが提出されていることが分かります (図 3)。

## ■ PARP 阻害剤の比較分析

抗がん剤として知られる PARP (poly ADP-ribose polymerase) 阻害剤のうち、Olaparib、Rucaparib、Niraparib、Talazoparib 起る貧血症の解析事例を示します。

CLARITY PV の解析によると 2014 年承認の Olaparib は 2016 年に、2016 年承認の Rucaparib は 2021 年に、2017 年承認の Niraparib は 2018 年に、2018 年承認の Talazoparib は 2019 年にそれぞれ ADR シグナルが検出されています。

CLARITY PV の比較分析機能 (図 4) を使うと、同一クラスの医薬品に対して、各有害事象レポートの信頼性を評価することができるので、その有害事象がクラス特有のものであるか、医薬品特有のものであるか、また、将来クラス特有の副作用 (ADR) になりそうか判断するための材料を容易に得ることができます。

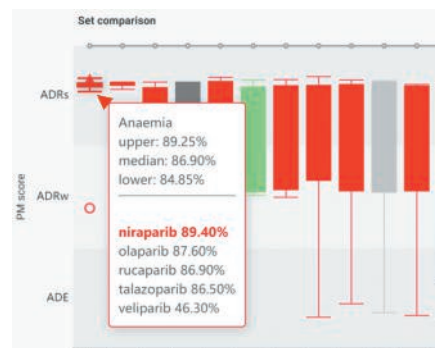


図 4. PARP 阻害剤による貧血症のレポートの比較分析

## ■ まとめ

CLARITY PV は、約 1 万件の医薬品に対する統合された安全性データを提供します。安全性薬理から市販後調査までの全てのデータに関連付けて、可視化する機能を備えており、トランスレーショナル・セーフティやファーマコビジランスを強力に支援します。

遺伝子発現データベース GENEVESTIGATOR

Sample Composition Tools と収載データのアップデート **NEBION**

GENEVESTIGATOR は遺伝子発現データベースのオンライン解析ツールです。公共データベースに登録されたマイクロアレイや次世代シーケンサー (NGS) の膨大な遺伝子発現データをキュレートすることで、様々な研究者により登録された大量の実験結果を統合して解析可能にします。また、GENEVESTIGATOR は使いやすいインターフェースと高速な検索エンジンを搭載しているため、研究者が標的遺伝子の探索などの遺伝子発現解析を行う際に、注目する遺伝子の同定や発現変動遺伝子の優先順位付けなどを簡単かつ正確に行うことができます。本稿では新しい解析ツールである Sample Composition Tools や新たに追加された遺伝子発現データについて紹介します。

■ Sample Composition Tools

GENEVESTIGATOR には多数の実験結果における遺伝子発現を横断的に解析できる COMPENDIUM-WIDE ANALYSIS と、個々の実験における遺伝子発現の解析を行う SINGLE EXPERIMENT ANALYSIS の 2 つの機能が搭載されています。SINGLE EXPERIMENT ANALYSIS では、実験群と対照群の発現変動遺伝子を探索する Differential Expression や多次元データを 2 次元平面上に t 分布型確率的近傍埋め込み法 (t-SNE) で可視化を行う Dimension Reduction などの解析ツールを利用できます。

Sample Composition Tools は SINGLE EXPERIMENT ANALYSIS に新しく追加された解析ツールです。Sample Composition Tools はサンプル中の異なる細胞タイプの組成やその相対量、および疾患や治療、発達などの異なるサンプル特性をグラフに表示することができます。そのため異なるサンプル特性同士の相関関係を解析することが可能です。

ヒトの肺胞マクロファージにおいて正常細胞と新型コロナウイルス (SARS-CoV-2) に感染した細胞を single cell RNA-seq で解析したデータセットを使用しました。図 1 は、新型コロナウイルス感染症 (COVID-19) 患者の重症度ごとに各年齢の割合を表示した Sample Composition Tools の結果です。60-70 歳が赤色、50-60 歳が青色、40-50 歳が黄色、30-40 歳が緑色、20-30 歳がピンク色で色付けされています。COVID-19 の症状が軽症の人のデータや COVID-19 に罹患していない人のデータでは比較的年齢が若い人が多く、COVID-19 の症状が重篤な人のデータでは比較的高齢な人が多いことが分かります。この結果から、COVID-19 の重症度と年齢には相関関係があることが分かります。その他にも細胞に発現しているバイオマーカーや性別、他の疾患情報などの複数の特性との関連性を調べることで、さらに分析を拡張できる可能性があります。

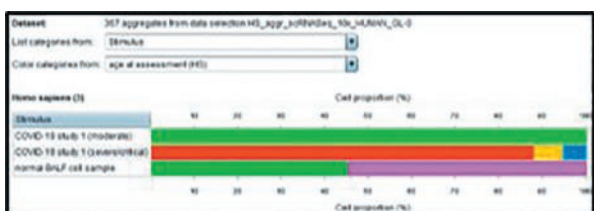


図 1. 年齢と COVID-19 の重症度の Sample Composition Tools

■ 遺伝子発現データの追加収載

GENEVESTIGATOR は公共データベースに登録されたマイクロアレイや RNA-seq などのバルク組織の遺伝子発現データや single cell RNA-seq の遺伝子発現データをキュレートし、統合したデータを利用した解析が可能です。2021 年 10 月現在、バルク組織の遺伝子発現データは、134 研究の合計 11,451 サンプルを利用することが可能です。バルク組織の遺伝子発現データでは、筋萎縮性側索硬化症のデータ、さらに自己免疫疾患や COVID-19、免疫学領域の研究データなどが追加されました (表 1)。特にがん関連の遺伝子発現データベースである The Cancer Genome Atlas (TCGA) の遺伝子発現データでは、がん患者の生存期間だけでなく死因やがん発生中の生存期間、がん再発のイベント、無増悪期間などのパラメーターが追加されています。これらのパラメーターは TCGA の論文<sup>1)</sup>を参考に登録されています。生物種は Homo sapiens (HS) や Mus musculus (MM) や Rattus norvegicus (RN) の 3 種類を利用することができます。Single cell RNA-seq の遺伝子発現データは、19 の研究データが追加されて、合計 138 研究の 570 万細胞のデータが利用可能です。

表 1. 追加されたバルク組織の遺伝子発現データ

Focus area	Organisms	# studies*	# samples*
Cardiovascular System	HS	3	148
Developmental and Regenerative Systems	HS	3	145
Digestive and Endocrine Systems	HS	24	2461
Immune System	HS	43	3283
Nervous System	HS	14	1681
Oncology	HS MM	28	1807
Pharmacology and Toxicology	HS MM RN	3	25
Reproductive and Renal Systems	HS	9	552
Respiratory System	HS	20	2503
Various Systems	HS	7	552

■ ご試用

GENEVESTIGATOR は、無償トライアルにより 30 日間ご評価いただけます。遺伝子発現解析をされる方はぜひ GENEVESTIGATOR をお試しください。トライアルをご希望の方は弊社ウェブサイトよりお問い合わせ下さい。

<https://www.molsis.co.jp/contact/>

1) Liu, J.; et al. *Cell* **2018**, *173* (2), 400-416.

# SciMeeting2021

SciMAPSの開発元である Scienomics 社は去る 10 月 27-29 日の 3 日間、SciMeeting をウェブ上で開催いたしました。SciMeeting は、企業研究でのシミュレーション活用事例やシミュレーション手法の最新の研究等について専門家から講演いただくことで、マテリアルサイエンス分野のシミュレーション技術を紹介することを目的としています。2019 年までは海外で現地開催されておりましたが、昨今の事情を鑑みて本年はウェブ開催となりました。本稿では 4 つの講演の内容について紹介します。

## Refrigerants and Ionic Liquids with Materials Modeling

University of Notre Dame

Prof. Edward Maginn

Classical パッケージに追加されたモンテカルロ法計算プログラム CASSANDRA の開発者である演者は、モンテカルロ法と分子動力学法を組み合わせることで、イオン液体含侵膜中のガス分子の透過性を評価し、混合ガスから二酸化炭素を分離させるのに最適なイオン液体の組み合わせを調べる応用事例を紹介しました。含侵するイオン液体がアミノ酸イオン液体 [P4444] [Gly] の場合には、二酸化炭素とアミノ酸陰イオンが反応することでより多くの二酸化炭素を吸収しますが、生成するカルバミン酸基が水素結合ネットワークを形成し粘性が極度に上昇するため、高い透過性能を発揮できないことを分子動力学計算により示しました。この知見を活かし二酸化炭素を吸収してもカルバミン酸基を生成しない非プロトン性複素環陰イオン (AHA) を導入することで、高い吸収性を持ちかつ反応後も粘性がほとんど変わらないイオン液体を考案しました。他に、臨界点の予測の難しいハイドロフルオロカーボン類について、機械学習を利用した独自の手法による力場パラメータの改良を行い、冷媒の回収、再利用に関するイオン液体の事例を紹介しました。

## Materials simulations at Freudenberg Technology Innovation (FTI)

Freudenberg Technology Innovation

Dr. Daniele Savio

LAMMPS による分子シミュレーションを利用して原子レベルでのポリエーテルエーテルケトン (PEEK) 高分子表面間の摩擦についての研究事例を紹介しました。アスペリティ接触部における摩擦工程や構造変形、局所摩擦に対する物理吸着流体層の影響について調べるために凸凹な表面、平坦な表面、水分子を含まないドライなモデルとウェットなモデルを用意し、さまざまな圧力でシミュレーションを行いました。ドライなモデルでは、大きくサインカーブを描くような凸凹な表面であっても、20 nm 程表面をスライドすると平坦から開始したモデルと構造的には同じようになることが示され、圧力が強いほど平坦化する速度は速くなりました。せん断応力-圧力を調べると、ウェットなモデルでは、表面にせん断方向に整列した水によるナノチャンネルが生成し、高分子界面の距離が保たれ、せん断応力は低減されますが、

プロットの傾きから得られるナノスケールでの摩擦係数はほとんど変化しないことが示されました。ここで得られたナノスケールでの摩擦係数と連続体接触力学シミュレーションを組み合わせたモデル式を考案し、実験で得られる数値と比較してその傾向の再現性と有用性を示しました。

## Parameterizing chemically-specific coarse-grain models of polymer melts: Structure and dynamics

NIST

Dr. Lilian Johnson and Dr. Fred Phelan

ポリマーの粗視化モデルのパラメータ作成についてはボトムアップ的なアプローチを通常行いますが、ポテンシャル面の粗視化に伴うエネルギー表面の平滑化により、粒子の運動が不自然に加速されるという問題があります。この問題を解決するために、化学的な挙動を再現しながら動的に正しいモデルを得る方法として、Langevin サーモスタットの摩擦をあらゆるファクター $\Gamma$ を利用する手法を考案しました。モノマーの拡散、分子鎖の拡散、末端間ベクトルの自己相関、粘性に着目して、全原子モデルの挙動とメソスケールの挙動で、 $\Gamma$ を評価し、どのような $\Gamma$ を選択すべきかを検討した事例を紹介しました。

## Molecular engineering for environment friendly industrial processes and materials design

Texas A&M University at Qatar

Prof. Ioannis G. Economou

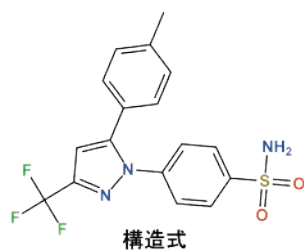
分子シミュレーションによる細孔内の分子の挙動についての研究を中心に環境にやさしいプロセスと材料デザインに関する研究を紹介しました。ゼオライト様イミダゾレート骨格材料として知られる ZIF-8 については、細孔径の大きさと拡散速度に基づく分離および吸着に基づく分離の 2 つについての事例を紹介しました。前者については、ZIF-8 のヒンジ部分の金属原子イオンと有機リンカー部分に置換基を導入することで、分離するガス種や拡散速度を調整し、商用的に期待される今までにない分離能を持つ骨格構造の候補を示しました。後者については、イオン液体を細孔内に導入することで二酸化炭素の選択的な分離に最適なイオン液体の ZIF-8 に対する質量比をイオン種の組み合わせと共に示し、分子シミュレーションの有用性を示しました。

# REINVENT インターフェースの紹介

低分子の生成モデルである REINVENT<sup>1)</sup> は、近年注目されているディープ・ニューラル・ネットワークの技術を創薬研究に取り入れた応用事例として注目されています。MOE の REINVENT インターフェース (I/F) では、MOE 上から REINVENT を実行するための使いやすい GUI と、強化学習の評価関数として使用できる MOE のさまざまな機能を提供します。REINVENT I/F は、MOE のアドオンプログラム共有サイトである SVL Exchange から入手できます。本稿では、REINVENT の概要と REINVENT I/F の機能や使用例について紹介します。

## REINVENT の概要

REINVENT は、プログラミング言語 Python で実装された Recurrent Neural Network (RNN) に基づく SMILES 生成モデルです。SMILES は、分子の化学構造の表記法の一つで、図 1 のように文字列で記述します。REINVENT では化合物ライブラリーを用いて教師あり学習されたモデルをフィンガープリント類似度や活性予測モデルなどの評価関数を取り入れた強化学習によってさらに調整することができます。



O=S(=O)(c3ccc(n1nc(cc1c2ccc(cc2)C)C(F)(F)F)cc3)N

SMILES

図 1. 構造式 (上) とその SMILES (下) の例

## RNN の教師あり学習

RNN とは、時系列データを扱えるように拡張したニューラルネットワークです。主に、自然言語処理で用いられています。RNN では、一つ前の隠れ層の出力が次の入力の一部となり、次の出力に影響を与えます。図 2 は、RNN の教師あり学習を示したものです。例えば、 $X_0$  から  $X_4$  で構成された配列を教師あり学習するとします。このとき、 $X_0$  は開始の記号で  $X_4$  は終了の記号です。 $X_0$  を最初の入力として与えると、次の要素の確率分布が得られます。図 2 では、正解の  $X_1$  の確率を出力として示しています。ここで得られる隠れ層の出力 (図 2 の右向きの矢印) が次の  $X_1$  とともに入力として用いられます。この計算ステップを最後の  $X_4$  の確率分布が得られるまで繰り返されます。教師あり学習では、正解の要素の確率の対数和にマイナスをつけたものを損失関数として定義し、これを最小化するように RNN の重み ( $W$ ) やバイアス ( $b$ ) といったパラメーターを最適化します。REINVENT においては、ここで示した配列が SMILES に対応します。SMILES の生成過程では、各計算ステップで得られる確率分布に基づいて要素 (SMILES を構成する文字) がサンプリングされます。

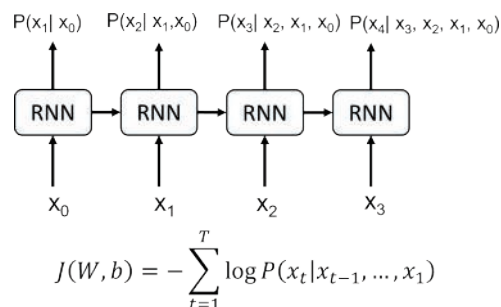


図 2. RNN の教師あり学習の概念図 (上) と損失関数 (下)

## REINVENT の強化学習

REINVENT では、まず、医薬品候補化合物のアッセイデータベースの ChEMBL から SMILES を抽出して、これを基に上記の教師あり学習が行われ、妥当な SMILES を生成するモデル (Prior Network) を得ます (図 3)。次に強化学習では、Prior Network が Agent Network (強化学習によって調整されたモデル) の初期値として用いられます。強化学習の各計算ステップでは、Agent Network によって SMILES 群を生成し、それらの評価値を計算します。既存機能では、いくつかのフィンガープリントによる類似度や活性予測モデルを評価関数に使用できますが、REINVENT I/F では、さらに MOE の機能が使用できます。これらの評価値と Prior Network から計算した SMILES 群の生成確率から、理想的な生成確率を求めます。さらにこれと Agent Network による SMILES 生成確率から損失関数を定義し、これを最小化するように Agent Network のパラメーターが更新されます。このような計算ステップを繰り返し行うことで、Agent Network は高い評価値を出すような SMILES を生成できるようになります。

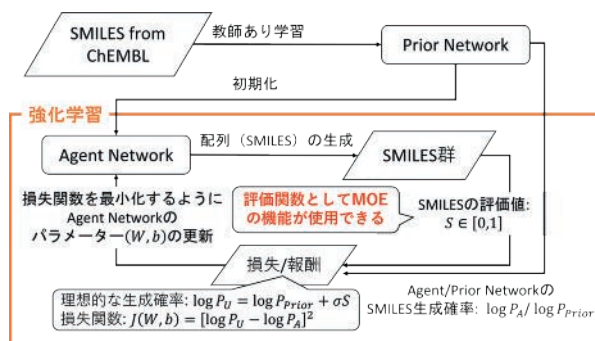


図 3. REINVENT の強化学習の概念図

## REINVENT I/F

MOE の REINVENT I/F では、使いやすい GUI とともに以下の機能を提供します。

- MOE の評価関数：強化学習における評価関数として、MOE の記述子、QSAR モデル、フィンガープリントモデル、ファーマコフォアモデル、コンセンサスモデル、ドッキングスコアが使用できます。
- 強化学習の実行と監視：GUI からさまざまなパラメーターを設定し、強化学習を実行できます。実行している間、生成された SMILES の構造とそのスコアを監視できます (図 4)。
- サンプリングした化合物の評価：生成モデルでサンプリングした SMILES の三次元化や評価関数の適用が行えます。
- Prior Network の作成と教師あり学習の GUI：SMILES ファイルだけでなく、MDB/SD ファイルを入力データとして Prior Network の作成と教師あり学習が GUI から行えます。

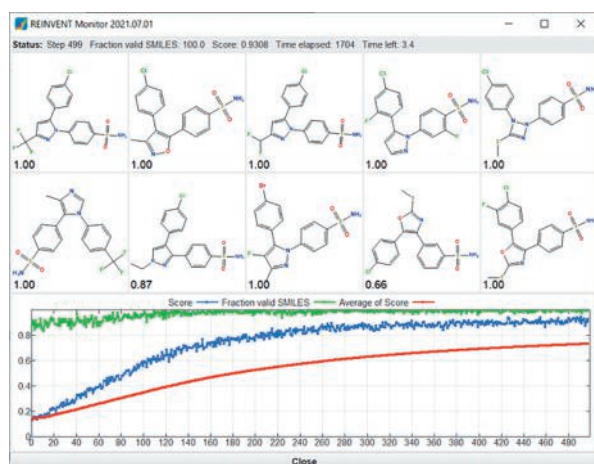


図 4. REINVENT の強化学習における REINVENT I/F の監視機能。上部には、各計算ステップで生成された化合物と評価値が表示され、下部には横軸を計算ステップとした妥当な SMILES の割合 (緑)、評価値の平均 (青色) と累積の平均 (赤) がプロットされている。

## 使用例

一つ目は、ファーマコフォアモデルを評価関数に使用した例です。図 5 の左のように、上皮成長因子受容体 (EGFR) と阻害剤のエルロチニブの共結晶構造を基にして、水素結合アクセプターと芳香環中心の特性球を持つファーマコフォアモデルを構築しました。それを評価関数として 500 ステップの強化学習を行いました。その結果得られた化合物が図 5 の右になります。エルロチニブと同様の縮合環を持つ構造が得られているのが分かります。

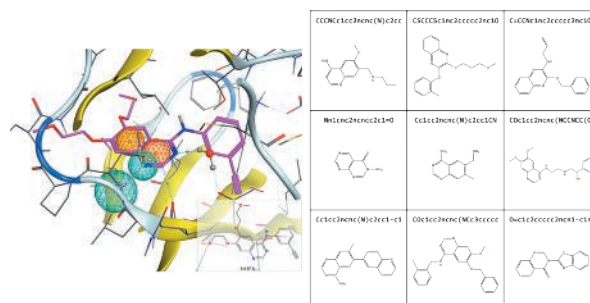


図 5. 強化学習の評価関数に使用したファーマコフォアモデル (左) と生成された化合物群 (右)。ファーマコフォアモデルは水素結合アクセプター (水色) と芳香環中心 (橙色) の特性球を含む。

二つ目は、コンセンサスモデルを評価関数に用いた例です。コンセンサスモデルは QSAR モデル、フィンガープリントモデル、ファーマコフォアモデルなどを組み合わせたモデルです。ここでは、ファーマコフォアベースのフィンガープリント GpiDAPH3 で EGFR 阻害活性の高い化合物群と比較したときの類似度と、血液脳関門の透過性を予測する線形モデルを組み合わせたコンセンサスモデルを使用しました。このモデルの評価値は、血液脳関門の透過性が高く、既知の EGFR 阻害剤と似ている程、高くなります。得られた結果では、既知の EGFR 阻害剤と同様な縮合ヘテロ環を部分構造にもつ化合物が得られているのが分かりました (図 6)。

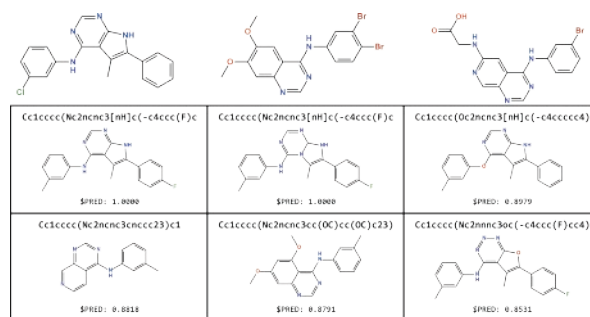


図 6. 既知の EGFR 阻害剤 (上) と強化学習にコンセンサスモデルを用いて得られた化合物群 (下)。コンセンサスモデルでは、EGFR 阻害剤のフィンガープリントモデルと血液脳関門の透過性を予測する線形モデルを組み合わせている。

- 1) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* **2017**, *9* (1), 1–14.

## BioSolveIT 社製創薬支援ツール

## Enamine REALSpace からの RAF キナーゼ阻害剤の探索

ほぼ無限のサイズの化合物空間（ケミカルスペース）の中から目的分子を探索するのは容易ではありません。BioSolveIT 社の infiniSee は、数百億個を超える合成可能あるいは購入可能な分子群の中より、クエリー構造に類似した構造を短時間で発見することができます。ここでは 200 億個以上の分子に相当するケミカルスペースである Enamine REALSpace の中から、infiniSee を用いた RAF キナーゼ阻害剤の探索事例を紹介します。

## ■ infiniSee 概要

infiniSee は、閲覧が不可能なほど巨大なサイズのケミカルスペースを高速に探索し、クエリーと類似した、合成可能あるいは購入可能な化合物を出力します。図 1 に示すように、まずはクエリーを化学的特徴とその繋がりを定義した 2 次元の Feature Tree に変換します。次に、類似の Feature Tree を構築するために、化学反応ルールに基づき分子フラグメントを組み合わせます。構築した Feature Tree を分子構造に変換することで、クエリーとの類似構造をオンデマンドで構築します。分子の 2 次元構造情報のみを使用するため、立体配座に影響しない検索が可能です。化学反応ルールには、フラグメント間の結合ルールが定義されており、合成の容易性を加味した構造が得られます。設定として、類似構造の Target Similarity、出力分子数、多様性、拘束する官能基を指定できます。Target Similarity はクエリーと類似構造の類似度を意味し、指定した類似度付近の分子構造を出力します。

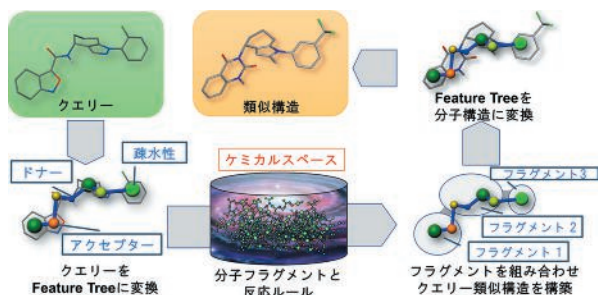


図 1. infiniSee によるケミカルスペース探索

## ■ Enamine REALSpace

Enamine 社は、世界最大のスクリーニング化合物やビルディングブロックの在庫保有数を誇る、ウクライナのサプライヤーです<sup>1)</sup>。REALSpace は、Enamine 社が取り扱う 200 億分子以上に相当するケミカルスペースです。104,492 個の在庫試薬と 171,912 個のビルディングブロック、156 個の自社の化学反応ルールから構成されています（図 2）。このケミカルスペース中の構造は 80%

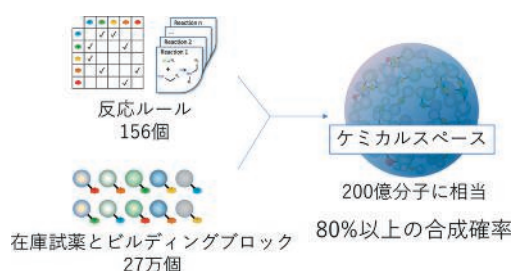


図 2. Enamine REALSpace

以上の合成確率を保っています。infiniSee では、REALSpace の他に、CHEMriya（OTAVA Chemicals 社、110 億分子に相当）、Galaxi（Wuxi LabNetwork 社、20 億分子に相当）、KnowledgeSpace（バーチャル化合物、 $10^{14}$  分子に相当）等の様々なケミカルスペース<sup>2)</sup>を検索対象に用いることができます。

## ■ RAF キナーゼ阻害剤の類似構造探索

infiniSee と REALSpace を用いた、RAF キナーゼ阻害剤の探索事例を紹介します。Sorafenib（図 3）は RAF キナーゼ阻害剤の一つで、その他のさまざまなキナーゼを阻害することが知られています<sup>3)</sup>。

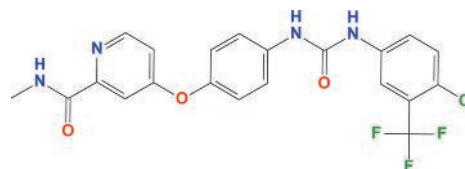


図 3. Sorafenib

Sorafenib をクエリーとして、infiniSee を用いて類似構造探索を行います。ここでは、ケミカルスペースを REALSpace、Target Similarity を 1.00 あるいは 0.85、出力分子数を各 1000 分子に設定します。infiniSee は、2 次元の Feature Tree が類似した構造を高速に見つけることができます（図 4）。infiniSee の計算時間は一般的なノート PC で 10 分間程度です。

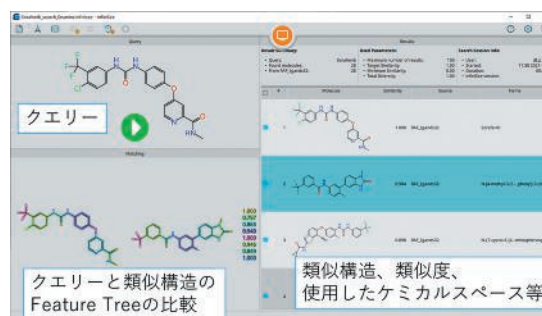


図 4. infiniSee インターフェース。クエリーとの類似構造を出力し、クエリーと類似構造との Feature Tree による類似性を可視化。

infiniSee で得られた類似構造に対して、ドッキングシミュレーション等の分子の立体構造に基づく計算手法を適用することで、より効率的なバーチャルスクリーニングを行えます。ドッキングシミュレーションの準備と実行には、統合計算化学システム MOE を用いました。MOE に搭載された Kinase Database から、RAF キナーゼの複合体構造を検索し、32 個の複合体を重ね合わせました。Sorafenib とその他の RAF キナーゼ複合体を含む計 26 個の複合体に共通する 3 個のファーマコフォアフィー



チャーを抽出しました (図5左)。これらのファーマコフォアフィーチャーをガイドとして Induced Fit ドッキングを実施しました。Sorafenib のリドッキングでは X 線構造を精度良く再現できました (図5右、X 線との RMSD 値は 0.18 Å)。

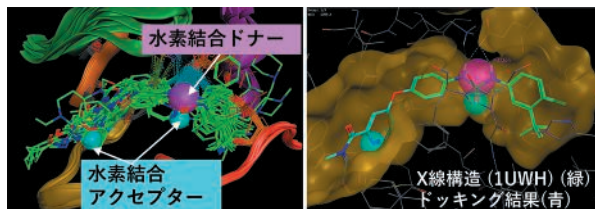


図5. RAF キナーゼリガンド複合体の重ね合わせ構造とファーマコフォア (左)、Sorafenib のリドッキングの結果 (右)

ここでは、Sorafenib をクエリーとして、Target Similarity を 1.00 に設定し infiniSee を実行しました。1000 個のヒット構造に対してドッキングシミュレーションを行いました。

図6に、ドッキングスコアが上位のヒット構造を示しています。図6左の構造は Sorafenib との類似度は 0.97、図6右の構造の類似度は 0.96 です。いずれも Sorafenib と非常に類似した構造を持ち、ドッキングポーズも類似していました。infiniSee では Target Similarity を大きく設定することで、クエリーの置換基や元素を一部入れ替えた構造を見つけることができます。

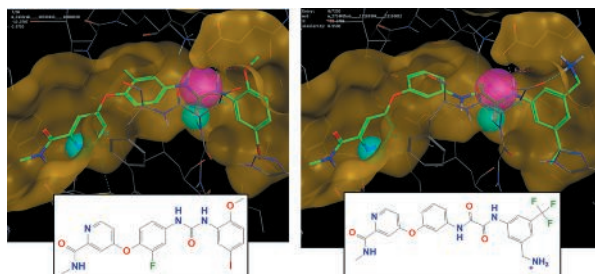


図6. Target Similarity 1.00 で探索しヒットした類似構造とそのドッキング結果

## ■ RAF キナーゼ阻害剤の Scaffold hopping

Sorafenib をクエリーとして、Target Similarity を 0.85 に設定し infiniSee を実行しました。ヒット構造の中に、Sorafenib と別の母核を有する Ponatinib を見出しました。図7左は、Ponatinib の 2次元構造と Sorafenib との類似度を示し、図7右は Ponatinib のドッキング構造を示しています。X 線構造との RMSD 値は 1.10 Å でした。infiniSee では Target Similarity を小さく設定することで、

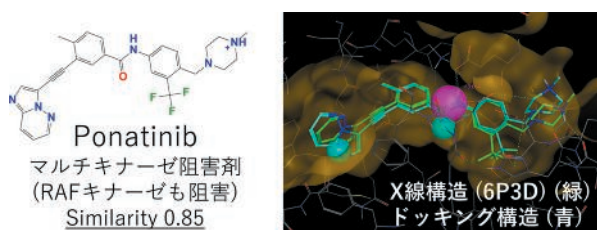


図7. Target Similarity 0.85 で探索しヒットした Ponatinib とそのドッキング結果

Scaffold Hopping やケミカルクラスをジャンプした構造をヒットさせることができます。

## ■ 他のフィンガープリントとの比較

Sorafenib と Ponatinib 間の類似度を、Feature Tree と各フィンガープリントを用いて比較しました。図8は、両分子の Feature Tree による比較図です。同色は類似の官能基を意味し、数値は官能基ごとの類似度を指します。表1は、他のフィンガープリントを用いて2化合物間の類似度を評価したものです。Feature Tree による比較が最も高い類似度を示しており、部分構造のフィンガープリント (MACCS、ECFP4、Morgan) では2化合物の類似性を判定することが難しいことがわかります。この例では、ファーマコフォアフィンガープリント TGT が Feature Tree と同じ類似度を示しました。ファーマコフォアフィンガープリントは Feature Tree と似た傾向を示す時がありますが、Feature Tree は化学的な特徴に加えて体積や環構造の個数等の特徴も含む他、各特徴間の繋がりも木構造として考慮しており、より精密な類似度を求めることができます。

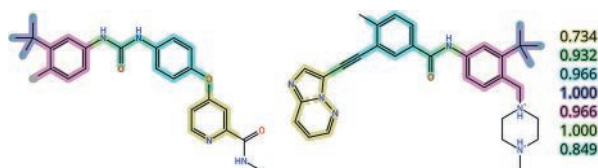


図8. Sorafenib と Ponatinib の Feature Tree の比較

表1. Feature Tree と他のフィンガープリントとの類似度の比較

フィンガープリント	類似度
infiniSee Feature tree	0.85
MOE MACCS (166 substructures)	0.41
MOE TGD (2 points pharmacophores)	0.73
MOE TGT (3 points pharmacophores)	0.85
KNIME ECFP4 (circular substructures)	0.22
KNIME Morgan (circular substructures)	0.26

## ■ まとめ

infiniSee を用いて、膨大なケミカルスペースの中より、クエリーと化学的特徴が類似した構造を高速に探索できます。infiniSee の GUI を用いて簡単な操作で検索が可能で、ヒット構造との類似性を可視化できます。複数のクエリーを入力とした検索も、コマンドラインや KNIME を用いることで可能です。Target Similarity を設定することで、置換基の一部を入れ替えた類似構造や母核置換した類似構造を探索できます。infiniSee による 2次元構造を用いた探索の後に、さらにドッキングシミュレーションなどの分子の立体構造を用いた手法と組み合わせることで、より効率的なバーチャルスクリーニングが可能になります。

- 1) <https://www.namiki-s.co.jp/supply/common.php?id=enamine>
- 2) <https://www.biosolveit.de/infiniSee>
- 3) 竹内伸司. 肺癌. 2015, 55 (6), 941-947.

# MedeA-VASP による金属表面への分子吸着の評価

MedeA は、第一原理計算や力場計算を基に材料のさまざまな物性を算出することができます。MedeA-VASP には、van der Waals (vdW) 相互作用を評価できる手法が搭載されています。本稿では、MedeA-VASP を用いた金属表面への分子吸着を事例として、その効果について解説します。

## ■ vdW 密度汎関数法

一般的な密度汎関数法 (Density Functional Theory, DFT) では vdW 力が十分に考慮されません。そこで、非局所相関項を取り入れた vdW 密度汎関数 (vdW-DF) が開発されています。vdW-DF がその特長を発揮する対象の一つとして、金属表面と分子の間の相互作用評価があげられます。特に、ベンゼンなどの芳香族分子が金属表面に吸着する系は、物理的な吸着系でありその典型と言えます。

本稿では、モデルとして Cu(111) 面へのベンゼンの吸着を取り扱い、金属表面と分子の間の相互作用について評価します。

## ■ ポテンシャルエネルギー曲線と吸着エネルギー

MedeA のフローチャート機能を用い、Cu(111) 表面 - ベンゼン間距離に対するエネルギー変化をプロットするための計算を実行しました。vdW-DF と通常の DFT で比較を行うため、通常の DFT には GGA-PBE<sup>1)</sup> を、vdW-DF には rev-vdW-DF2<sup>2)</sup> を交換相関汎関数として採用しました。計算に用いたフローチャートを図 1 に示します。

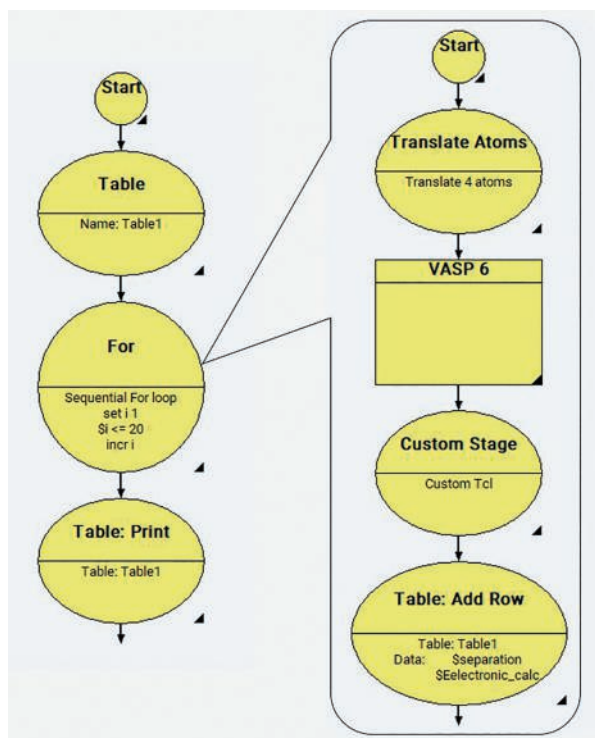


図 1. ポテンシャルエネルギー曲線作成のためのフローチャート

図 1 右側の囲み内のフローで具体的な操作・計算を行います。囲み部分は For ループで繰り返されます。「Translate Atoms」ステージでベンゼンを 0.2 Å ずつ移動します。C 原子は計算時には固定します。Cu(111) 表面モデルは 5 層とし、最下部 2 層を固定しています。

「VASP 6」ステージで MedeA-VASP による構造最適化計算を実行します。計算条件は以下のとおりです。

- ・平面波カットオフ値: 500 eV
- ・k-spacing: 0.3 Å<sup>-1</sup> shift to  $\Gamma$
- ・エネルギー収束条件: 1.0E-6 eV
- ・構造最適化収束条件: 1.0E-2 eV/Å

For ループで指定した回数操作が完了すると、結果は図 2 のようにテーブル形式にまとめて出力されます。テーブルへの出力は桁数や出力する値の追加などフレキシブルに設定することができます。また、図 2 のようなテキスト出力の他、csv 形式等のファイルに出力することも可能です。

Stage 3: Print Table Table1

Separation A	Energy eV
3.53	-114.157906
3.73	-114.111793
3.93	-114.003343
4.13	-113.816569
4.33	-113.541977

図 2. 表面 - 分子間距離とエネルギーの出力 (例)

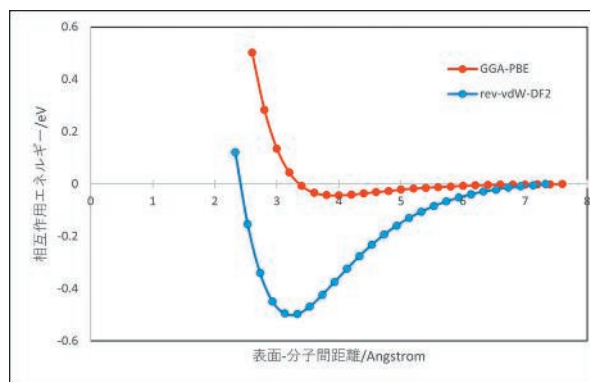
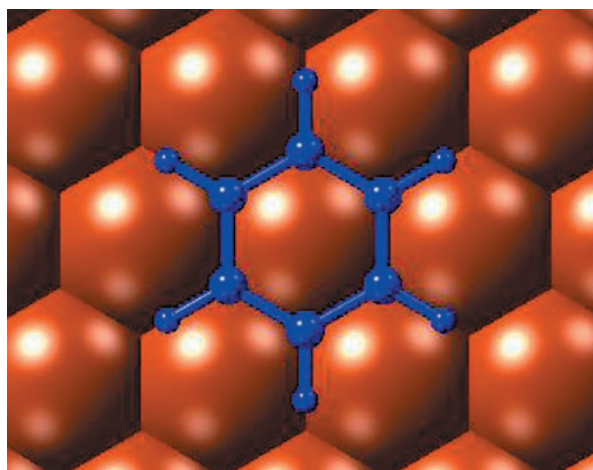


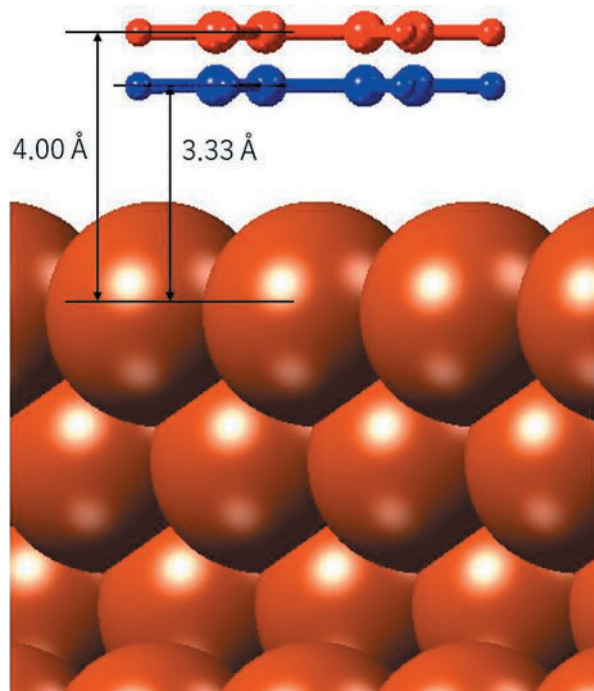
図 3. Cu(111) の表面 - ベンゼン分子間距離に対するポテンシャルエネルギー曲線。赤: GGA-PBE、青: rev-vdW-DF2。

図 2 で得られたデータから作成したポテンシャルエネルギー曲線を図 3 に示します。赤の GGA-PBE と比較して、青の rev-vdW-DF2 は極小点が深く、表面 - 分子間距離が短いことが分かります。ベンゼンが Cu(111) 表面

から最も離れた時のエネルギーを基準とすると、表面モデルへのベンゼンの吸着エネルギーは GGA-PBE で  $-0.043$  eV、rev-vdW-DF2 で  $-0.50$  eV となりました。実験値は  $-0.55$  eV との報告<sup>3)</sup> があり、rev-vdW-DF2 の結果とよく一致します。一方 GGA-PBE では、吸着エネルギーは一桁過小評価していることとなります。



a)



b)

図 4. Cu(111) 面へのベンゼンの吸着安定構造。a) Cu(111) 表面の法線方向からの図 b) Cu(111) 面を横から見た図。赤：GGA-PBE、青：rev-vdW-DF2。a) では rev-vdW-DF2 のみ表示。

## ■安定吸着構造の比較

図 4 に Cu(111) 面上のベンゼンの吸着安定構造を示します。ベンゼンは Cu 表面に対して、芳香環が最密充填の六角形にはまり込むように重なります。交換相関関数の違いによって、重なったときの表面との距離が異なることが図からわかります。やはり rev-vdW-DF2 の方が GGA-PBE より距離が  $0.7$  Å ほど短く、相互作用の強さが窺われます。

Cu(111) 面へのベンゼンの吸着については数多くの先行研究があります。例えば Toyoda らの報告<sup>4)</sup> では他の vdW-DF の手法<sup>5)</sup> を使用しており、吸着エネルギーは本稿に近い値を得ている一方、表面 - 分子間距離は GGA-PBE とほぼ変わらない結果となっています。同じ vdW-DF であってもさまざまであり、使用の際には自身のモデルで検証を行う必要があることを示しています。本稿では vdW-DF として rev-vdW-DF2 を選択しましたが、MedeA-VASP では他に以下の vdW-DF や分散項を追加した DFT+D を使うことができます。

- vdW-DFT: optB86b-vdW, optPBE-vdW, BEEF-vdW, rev-vdW-DF2, revPBE-vdW, SCAN+rVW10 など
- DFT+D: Grimme D2, Grimme D3, Tkatchenko-Scheffler, DFT-dDsC など

## ■まとめ

芳香族分子の金属表面への吸着状態を評価する例として、Cu(111) 面モデルとベンゼンを用い、MedeA-VASP による計算を行いました。vdW 相互作用の評価に長けた vdW-DF と一般的な DFT による結果を比較し、vdW-DF により吸着構造、吸着エネルギーが改善されることを示しました。

vdW 密度汎関数法は、分子結晶の構造・安定性評価に対しても有用です。詳細は、弊社ニュースレター 18 号に掲載の「MedeA-VASP による分子結晶の安定性評価」をご覧ください。

- 1) Perdew, J.P., Burke, K., Ernzerhof, M., *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- 2) Hamada, I., *Phys. Rev. B*, **2014**, *89*, 121103.
- 3) Caputo, R., Prascher, B.P., Staemmler, V., Bagus, P.S., Wöll, C., *J. Phys. Chem. A*, **2007**, *111*, 12778–12784.
- 4) Toyoda, K., *Panasonic Technical Journal*, **2010**, *55*, 27–32.
- 5) Dion, M., Rydberg, H., Schröder, E., Langreth, D.C., Lundqvist, B.I., *Phys. Rev. Lett.* **2004**, *92*, 246401.

## CORINA Classic の製品概要



CORINA Classic は、高精度かつ高速に低分子量から中分子量までの有機分子の三次元分子構造を生成するソフトウェアです。高品質な三次元分子構造を生成する標準のソフトウェアとして世界的に認められています。CORINA Classic は、これまで数多くのバージョンアップを重ねてきており、2021年9月にバージョン4.4.0をリリースしました。本稿では、CORINA Classic の概要とバージョン4.4.0の主な新機能について紹介します。

## CORINA Classic の特徴

## ■ 高速な処理

CORINA Classic は、X線構造に近い立体構造をルールベースで生成します。そのため、力場ベースでの処理と比較し格段に高速に立体構造を生成できます。一般的なx86 Linux ワークステーション上の中小規模の有機分子を3ミリ秒未満で処理します。米国国立がん研究所(NCI)のオープンデータベースの265,242の構造に対して優れた変換率(99.6%)を達成しています。

## ■ バッチモード実行

コマンドラインベースで動作するので、連続的なバッチ処理による大量のデータ処理に向いています。また、他のソフトウェアと連携も可能です。例えば、統合計算化学システム MOE は CORINA のインターフェイスを備えており、MOE の化合物データベースに対し、CORINA による化合物の三次元化ができます。

## ■ 複雑な構造の三次元化

CORINA Classic は、立体異性体(四面体キラル中心およびE/Z二重結合)の生成や比較的複雑な環構造をもつ分子の三次元化を行うことができます。例えば、シクロファン架橋ポルフィリン分子やC60フラレンといった構造の三次元構造を生成することができます。

また、CORINA は、最大6つの隣接原子を持つ化合物を処理できます。したがって、八面体中心までの金属錯体を扱うことができます。金属-配位子結合の長さは、それらの非共有特性を考慮した特定の要因によって修正されます。結果として得られる構造は、実験的に決定された形状に非常によく対応しています。

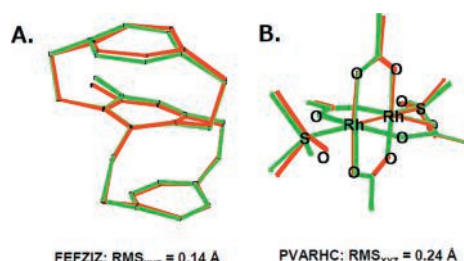


図1. 複雑な環構造(A)と金属錯体(B)の三次元化  
緑: CORINA で生成した配座、赤: X線配座

## ■ バージョン4.4.0の主な新機能

- 立体異性体生成モジュールが拡張され、定義されたE/Z二重結合を個別に処理できるようになりました。つまり、立体異性体の列挙中に、定義されたE/Z二重結合の構成を保持して立体構造を生成できます。
- 新しいオプションの oesterplane は、エステル基の5つの原子「CC(=O)OC」を同一平面上に配置します。これにより、REFMACによる最適化中もエステル基の安定した平面性が保証されます。
- 入力構造に3D座標がある場合、CORINAはプロキラル中心を認識するようになりました。これにより、プロキラル炭素原子の水素原子が、入力ファイルの3D構造と同じ相対3D位置を維持することが保証されます。

## ■ 参考文献

Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.

- 弊社ウェブサイトには、取り扱い製品、最新のニュース、お客様向けサポート情報、他のサイトへのリンクなどを掲載しております。
- 電子媒体でのニュースレターおよびバックナンバー(RSIニュースレターを含む)は、弊社ウェブサイトをご覧ください。
- 本ニュースレターは、弊社取り扱い製品をご購入いただきましたサイトのご担当者様にお送りしています。その他に発送のご希望がございましたら弊社までご連絡ください。
- 記載されている会社名及び商品名は、各社の商標または登録商標です。

MOLSIS ニュースレター

No. 20, 2022.01

2022年1月1日発行

発行所 株式会社モルシス

Copyright © 2022 MOLSIS Inc.



株式会社モルシス

URL: <https://www.molsis.co.jp/>

TEL: 03-3553-8030 FAX: 03-3553-8031 E-mail: [sales@molsis.co.jp](mailto:sales@molsis.co.jp)