

統合計算化学システム

MOE 2019.01 リリース

2019年3月に統合計算化学システムMOEの最新バージョン MOE 2019.01をリリースしました。MOE 2019.01では、リガンド-受容体間の結合自由エネルギー計算、タンパク質物性推算、SAR解析、実験情報に基づく立体配座予測の機能の向上に加えて、タンパク質立体構造データベースシステムPSILOとの連携が強化されます。これらの機能向上により低分子からバイオ医薬品までの多様なモダリティ研究を強力に支援します。またHTMLに対応した柔軟な原子ラベル設定や、原子選択、バブルヘルプなどインターフェース面も改良され、より使いやすくなります。詳細については2、3ページの記事をご覧ください。

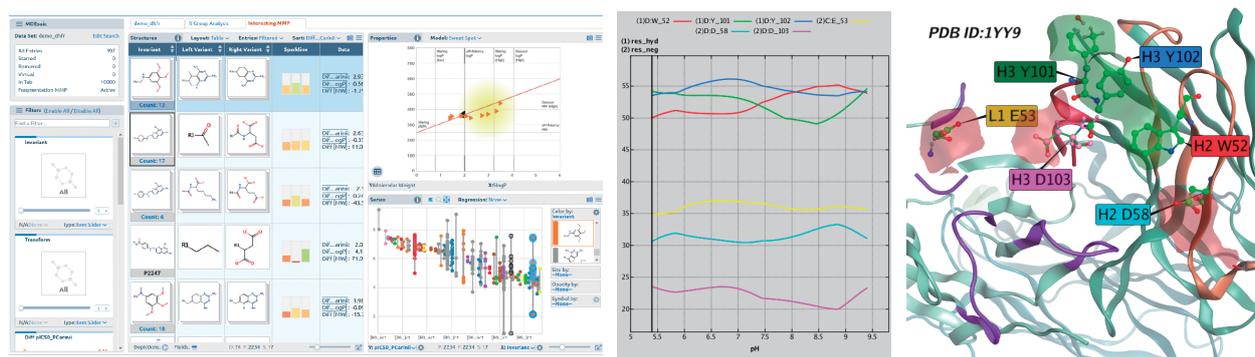


図1. MOEsaic(左)、タンパク質物性推算(中)、改良されたラベル表示(右)

CONTENTS

新製品情報

統合計算化学システム	MOE 2019.01 リリース	2
創薬研究情報共有クラウドシステム	CDD Vault新機能紹介	4
配座発生/プロファイリング/データセット管理	CORINA Symphony 1.1 機能紹介	5

技術情報

創薬支援ツール	KNIMEによるドッキング計算解析事例	6
医薬品安全性情報サービス	OFF-X: 副次的薬理スクリーニングの支援	7
次世代シーケンサーデータ解析ソフトウェア	Flow: Single Cell RNA-seqデータ解析	8
材料設計支援統合システム	ハイブリッド汎関数による高精度なバンド構造評価	10

セミナー情報

学会・展示会出展およびセミナー開催予定		12
---------------------	--	----

統合計算化学システム

MOE 2019.01 リリース

MOEは、さまざまな分子のモデリングとシミュレーションに対応する統合計算化学システムです。3月に最新バージョン MOE 2019.01 をリリースしました。MOE 2019.01 では、リガンドと受容体間の結合自由エネルギー計算、pHや構造変化を考慮したタンパク質物性推算、実験情報に基づく立体配座予測、柔軟な原子ラベル設定が新機能として追加されます。また、MOEsaic (SARアプリケーション)、抗体/タンパク質モデリング、エピトープマッピング、PSILOとの連携を中心に機能強化されます。プロジェクトデータベースも更新され、操作面においてもより便利に使いやすく改良されます。ここでは、MOE 2019.01 の代表的な新機能や機能強化を紹介します。

■ リガンドと受容体間の結合自由エネルギー計算

分子動力学 (MD) シミュレーションソフトウェアである Amber には、Thermodynamic Integration (TI) により結合自由エネルギー (ΔG) を推算する機能があります。新バージョンの MOE では、TI を使用して、受容体と異なるリガンド間の相対的な結合自由エネルギー差 ($\Delta\Delta G$) を簡単に評価するためのグラフィカルユーザーインターフェース (GUI) を提供します。

GUI において、受容体とそれに結合した複数のリガンド構造、計算条件 (圧力、温度、塩の種類と濃度)、MD シミュレーション時間を指定すれば、TI に必要なリガンドペアの決定、溶媒ボックスへの複合体の埋め込み、構造最適化計算、入力ファイルやスクリプトファイルの作成などを MOE が自動的にを行います。

計算結果の解析パネルでは、サンプリングされた TI の微分値やその平均/自己相関などを可視化できます。特に自己相関のプロット (図1) では、TI の微分値の安定性や収束の程度など、計算結果の質について評価できます。

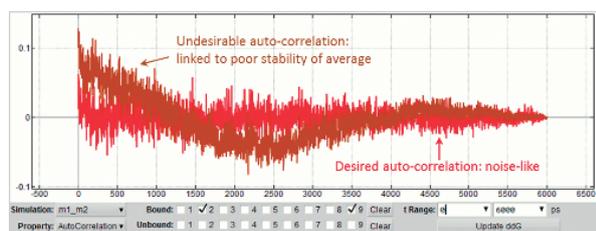


図1. TIの微分値の自己相関のプロット

■ MOEsaic アップデート

MOEsaic は化合物の構造データと活性データを整理し直感的な SAR 解析を可能にする実験研究者向けの MOE/web アプリケーションです。容易に MMP/R-group 解析が可能です。以下のような機能拡張が行われます (図2)。

■ 新規構造と活性データの追加

解析済みのデータセットに新規データを追加して、解析をはじめからやり直すことなく、MOEsaic の解析を続行できます。

■ 解析結果の共有

解析した結果の URL をメールで送ることにより、研究者間での MOEsaic の結果を共有できます。

■ カスタムプロパティモデルの追加

薬剤の評価基準として、Oral Drug、Sweet Spot、Golden Triangle の他に、ユーザー定義のモデルを追加できます。

■ バーチャル化合物の設計

スケッチャーで構築したバーチャル化合物をデータに追加して、MOEsaic の解析を続行できます。

■ インターフェースの改良

構造や各種プロットがより見やすくなります。プロットの拡大縮小に対応します。

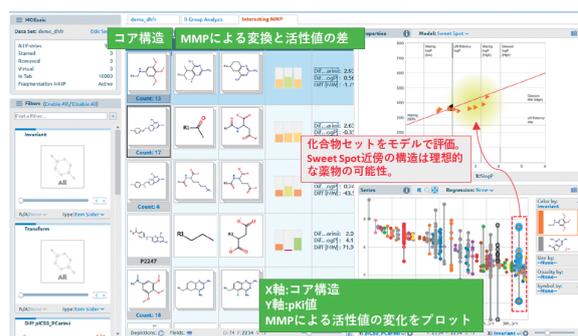


図2. 新しいMOEsaicのインターフェース。 Interesting MMP による網羅的な活性クリフの探索と化合物セットにおけるプロパティ評価をシームレスに実行。

■ 柔軟な原子ラベル表示

原子ラベルは以下のような機能追加が行われます。フォントの大きさ、スタイル、色、位置などを GUI から簡単に設定できるようになります。

■ HTML 対応

原子ラベルの文字について、太字、イタリック、上付き、下付き、フォントの種類、色の変更ができます。

■ 位置の調節

原子ラベルをマウスのドラッグ操作で任意の位置に移動できます。

■ テキストと背景の色

原子ラベルのテキストは、全体または指定された部分について任意の色を指定できます。背景の色を設定した場合、吹き出しが表示されます (図3)。

■ 特性値の挿入

GUI 上の分子に連動した原子、残基、分子鎖の特性

値を原子ラベルの文字列に挿入できます。特性値には、残基名、残基番号、元素名、原子名、部分電荷、キラリティー、温度因子、占有率などがあります。

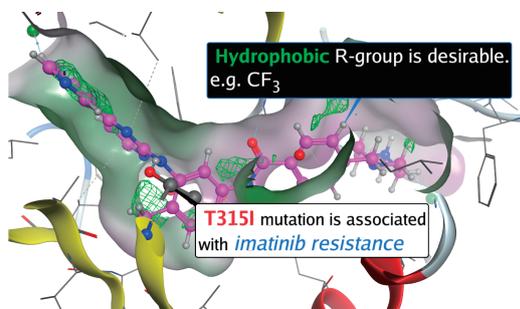


図3. 新しい原子ラベルの例

■ pHや構造変化を考慮したタンパク質物性推算

タンパク質の物性推算機能が強化され、タンパク質用の記述子において、そのpH依存性とLowModeMDによるアンサンブル平均の値を出力できます。残基ごとのプロパティの分布も計算でき、ある特定の残基におけるパッチ表面積の寄与なども出力できます(図4)。

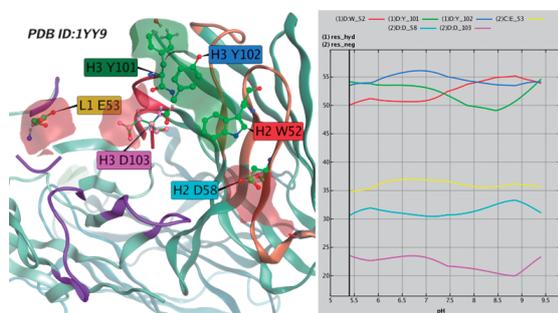


図4. 抗原との相互作用に関与するアミノ酸残基と分子表面パッチ(左図、緑: 疎水性パッチ、赤: 負電荷パッチ)と、残基ごとのパッチ表面積のpH依存性(右図のプロット色は左図のラベル色と関連)

■ MOE/web Database Browser

MOE/webにMDB、SDF、CSVを表示するためのアプリケーションが追加されます。低分子やタンパク質の三次元表示もサポートされ簡易的なビューアーとして使用できます(図5)。

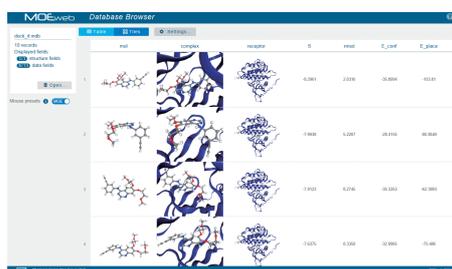


図5. NGLビューアーでの分子の回転が可能

■ NMRスペクトルからの配座予測

配座解析で得られた化合物の複数配座と、NMRにより得られた核オーバーハウザー効果(NOE)の結果を参照し、実験値に近い配座を予測できます(図6)。

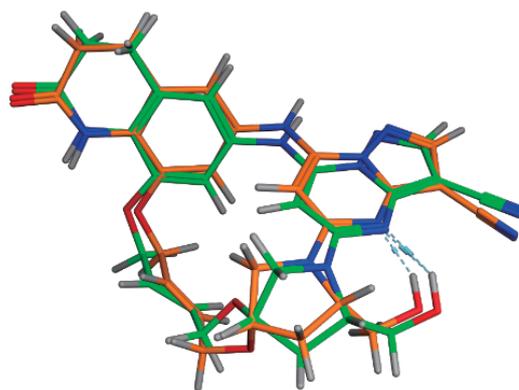


図6. 環状リガンドの配座予測。 橙: 溶液中のNOEの実験結果に近い配座、緑: 結晶構造。

■ PSILOとの連携

MOE 2019.01では、PSILOに登録されているタンパク質の立体構造に対して、類似ポケットや3D相互作用などのさまざまな検索をGUIから実行できます。ヒットした構造は、タンパク質の構造、リガンド、リガンド結合部位、3D相互作用検索のクエリーなどで重ね合わせてMOEに読み込めます。

■ その他の機能強化

■ タンパク質/抗体モデリング

MDB形式の複数のアミノ酸配列から、単一のテンプレートを用いた連続ホモロジーモデリングが可能です。バイスペシフィック抗体の構築がサポートされます。

■ エピトープマッピング

実験(例: 水素-重水素交換質量分析)から得られたエピトープ情報を用いて、エピトープマッピング機能のスコア計算に重み付けを行えます。

■ イオンチャンネルデータベースの追加

プロジェクトデータベースとして、イオンチャンネル(Na^+ , K^+ , Ca^{2+})が追加されます。

■ 原子選択とバブルヘルプ

マウスを用いて任意の距離の周辺原子を選択できます(図7)。原子の詳細情報をバブルヘルプで表示できます。

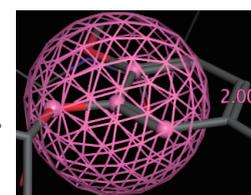


図7. 周辺原子の選択

■ まとめ

MOE 2019.01で強化される機能を用いて、より多彩なアプローチでの柔軟な分子設計が可能になります。この他にも多数の機能強化が行われます。

CDD Vault新機能紹介



CDD Vaultは、米国 Collaborative Drug Discovery (CDD) 社により開発された化合物情報とアッセイ情報を共有するクラウドサービスです。ハードウェアなどの準備や管理の手間をかけることなく、創薬研究データ管理システムをターンキーソリューションとして導入できます。優れた可視化機能を持つCDD Vaultは、製薬企業やベンチャー企業、大学、公的研究機関などが共同研究を進めるための最適なプラットフォームとして数多くの利用実績があります。また、ユーザーのニーズに応じて機能強化が常になされており、2018年にもさまざまな機能が追加されました。今回は2018年中に実装された主な新機能について紹介します。

■ CDD Vault とは

CDD Vaultは、構造式や物性値をはじめとした化合物情報およびその在庫情報、化合物ごとの数値や画像などのアッセイデータを管理できる研究情報共有システムです。また、オプションモジュールELNはCDD Vaultに登録する化合物の合成実験や各種アッセイの実験内容を記録する電子実験ノートです。低分子合成担当の化学者、アッセイ担当の生物学者、SAR解析担当の計算科学者が同じシステム上で情報を共有することで、組織や場所にとらわれない密なコミュニケーションを実現して研究業務を効率化できます。

CDD Vaultはクラウドサービスなので、随時行われる新機能の追加や既存機能の強化は自動的に反映され、ユーザーの作業は不要です。CDD Vaultの機能強化はCDD社のウェブサイト¹⁾に公開されています。2018年に実装された主な機能強化は次のようなものが挙げられます。

- ・CDD VaultとELNモジュールの連携強化
 - ・ELNへのアッセイアノテーション付与
 - ・生物製剤情報の管理
 - ・数値計算時に不等号を適切に保持
(例えば、 IC_{50} が" $<10^{-8}M$ "なら pIC_{50} を" >8 "と算出)
 - ・構造式の列を表の左端に固定
 - ・レコードのエクスポート機能の強化
 - ・検索時に指定できるキーワード数の増加
- ここでは、特に重要な機能強化である最初の2点について紹介します。

■ CDD Vault と ELN モジュールの連携強化

ELNモジュールはCDD Vaultとは独立した形で当初実装されていましたが、数度の新機能追加によりCDD Vaultとの連携が順次強化されてきました。

■ ELNモジュールからCDD Vaultを参照

ELNのページに実験内容を記載する時に、CDD Vaultに登録された化合物やアッセイを化合物名やプロトコル名といったIDや別名で検索してリンクを作成できるようになりました。リンクはIDのほかに指定したテキストや化合物の構造で表示できます(図1)。

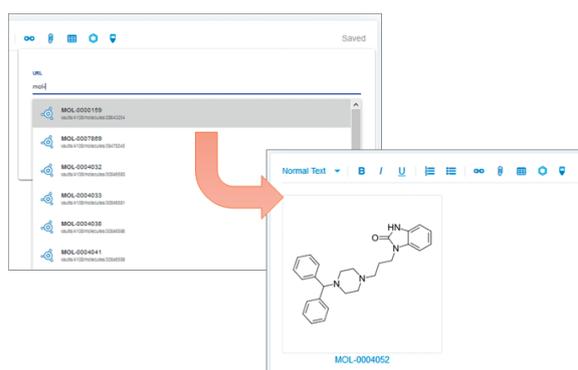


図1. ID(化合物名)で化合物の構造を呼び出し

■ ELNモジュールからCDD Vaultに化合物のバッチを登録

ELNのページに記入した構造式や反応式から、CDD Vaultに化合物のバッチを登録できるようになりました。こうして登録されたバッチはELNのページとひも付けられ、相互にリンクされます。

■ ELNへのアッセイアノテーション付与

ELNのページにバイオアッセイのアノテーションを付与できるようになりました。BioAssay Ontology²⁾や Drug Target Ontology³⁾などの定評あるオントロジーに基づいたアノテーションを選択できます。また、自由に記載することもできます。

付与したアノテーションに基づいてアッセイの簡単な概要を自動的に生成できます。生成した概要をより詳細な説明文のひな型として利用することで、複数のアッセイで説明文の体裁を統一できます。

■ トライアル

CDD Vaultの全ての機能は30日間無償でトライアルが可能です。ご希望の際は弊社ウェブサイトのお問い合わせフォームよりお申込み下さい。

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

- 1) <https://www.collaborativedrug.com/blog/recent-vault-updates/>
- 2) <http://bioassayontology.org/>
- 3) <http://drugtargetontology.org/>

MN/AM CORINA Symphony 1.1 機能紹介

配座発生/プロファイリング/データセット管理

CORINA Symphonyは、*in silico* の探索と最適化のために分子データセットを管理、操作、プロファイルするためのケモインフォマティクス・アプリケーションです。CORINA Symphonyは、CORINA Classicの機能による分子構造の標準化と三次元化の計算に加え、記述子計算とToxPrint¹⁾計算がグラフィカルユーザーインターフェース(GUI)のワークフローを用いて実行できます。今回は、昨年10月にリリースされたCORINA Symphony 1.1の主な機能を紹介します。

■CORINA Symphony 1.1概要

CORINA Symphonyは、CORINA Classicのアルゴリズムを利用して分子構造の標準化と三次元化をGUIのワークフローから実行できるように拡張したソフトウェアです。

CORINA Classicは、有機化合物の三次元構造を高速かつ高精度に生成することで定評がありますが²⁾³⁾、バッチ計算専用で、ターミナル画面からのコマンド入力による操作が必要となります。CORINA SymphonyはGUIのワークフローを用いて計算できるので、コマンド操作に慣れない方でもGUIを用いて手軽に計算が行えます。CORINA Classicの機能に加えて、記述子計算、ToxPrint計算、データ管理の機能が搭載されています。

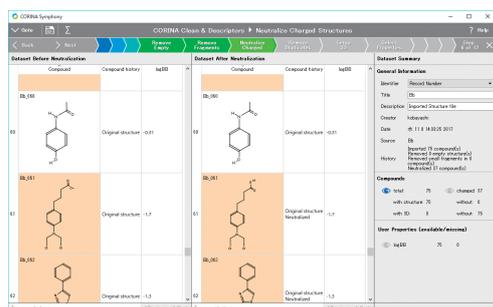


図1. ワークフロー画面(左: 処理前の分子, 右: 処理後の分子)

■分子構造標準化

分子構造を標準化します。フラグメントの除去/水素原子の付加/分子構造の中性化/重複分子の削除の処理が行えます。

■分子構造三次元化

分子構造の三次元化を行います。慣性モーメントをそろえるように分子を配置することができます。

■記述子計算

分子の構造および物理化学的特性を表現する一般的な記述子、形状および大きさの記述子、ベクトルに基づく記述子を提供しています。50種類以上の記述子が計算可能です。

CORINA Symphony 1.1では新たに量子化学計算ソフトEMPIREに基づく記述子が計算できるようになりました。EMPIREはCepos InSilico GmbHが開発する半経験的分子軌道法のソフトウェアです⁴⁾。MNDO、MNDO/d、AM1、AM1*, PM3、RM1を用いた計算が行えます。EMPIREを

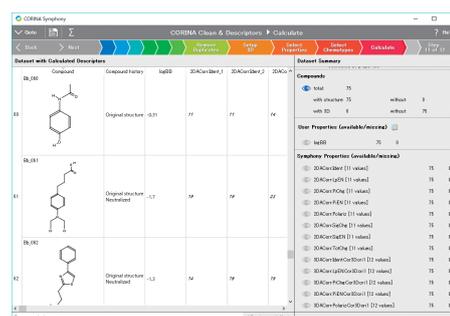


図2. 計算結果表示画面(左: 分子構造と特性値のテーブル, 右: 計算結果の表示と制御部)

わせて利用することで、量子力学的な特性も計算することができます。

■ToxPrint 計算

プロファイリングとフィンガープリント計算用のToxPrintケモタイプ(フラグメント)ライブラリーを内蔵しています。ToxPrintのケモタイプは、Chihae YangとAltamira LLCによって開発され、米国食品医薬品局(FDA)の下部組織である食品安全・応用栄養センター(CFSAN)の食品添加物安全局(OFAS)で利用されています。ToxPrintは、700以上の化学種を計算できます。

■データセット管理機能

化合物に関する物性値や記述子などのデータおよび関連情報を保存・管理するためのデータベース・バックエンドを提供します。

- インポートまたは計算されたデータをテーブル表示し、視覚的に管理および解析
- 表セルの編集や特性での並べ替え
- 外部スプレッドシート・アプリケーションへのコピー&ペースト
- 分子構造や計算結果をSDF、CSVファイルなどで出力

- 1) Yang, C. *et al. J. Chem. Inf. Model.* **2015**, *55* (3), 510–528.
- 2) Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
- 3) Schwab, C.H. *Drug Discov. Today Technol.* **2010**, *7* (4), e245–e253.
- 4) 別売(アカデミックフリーライセンス)
<http://www.ceposinsilico.de/products/empire.htm>

KNIMEによるドッキング計算解析事例



BioSolveIT社から、さまざまな創薬支援ツールをKNIME¹⁾上で利用するためのKNIME Interfaces (23種の解析ノードセット)の新バージョンがリリースされました。KNIMEのワークフローからStructure-Based Drug DesignのためのプラットフォームLeadITやメディシナルケミスト向けのリード化合物最適化ツールであるSeeSAR等のBioSolveIT社のツールを使った解析が可能です。ここでは、KNIMEを使ったドッキングシミュレーションの解析例を紹介します。

■ BioSolveIT KNIME Interfacesで利用できる機能

KNIMEは独Konstanz大学で開発されたワークフロー作成ツールです。直感的で操作しやすいグラフィカルユーザーインターフェース(GUI)を備え、処理の追加、削除、順番の入れ替えが容易です。ワークフローを使うことで処理の自動化が実現できます。KNIME Interfacesでは、23種類のノードが提供されており、BioSolveIT社の創薬支援ツールであるSeeSAR, LeadIT(FlexX, ReCore), CoLibri, FTrees, Naomiが利用可能です(図1)。



図1. KNIME Interfacesで提供されるノード

■ KNIME Interfacesによるワークフロー例

ここではLeadITとSeeSARのノードを活用したドッキング計算のワークフローについて説明します。このワークフローでは、ドッキングに使うリガンド構造の取得、受容体の準備、LeadITによるドッキング計算、SeeSARによるスコア計算と結果の可視化を行います(図2)。

1. リガンドの準備

PDB SMILES Query ノード(Vernalis社)を用い、ステロイド系抗炎症薬であるデキサメタゾンのSMILESを入力として類似構造をPDB(www.rcsb.org)ウェブサービスから取得します。続くノードでリガンドIDに基づく重複除去、Generate 3D Coordinate ノードによる分子構造の3次元化を行います。ここではデキサメタゾンとの類似度が0.7以上の分子構造が5つ得られました。

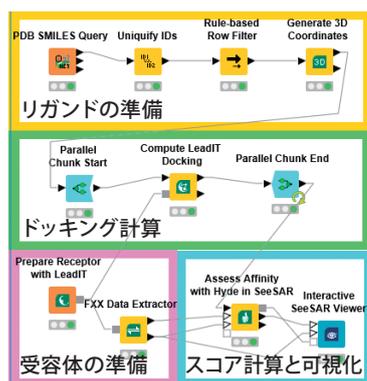


図2. ドッキング計算のワークフロー例

2. 受容体の準備

Prepare Receptor with LeadIT ノードの設定パネルでLeadITのGUIを呼び出し、受容体領域やリガンド結合サイトの設定を詳細に行うことができます。ここでは、グルココルチコイド受容体(PDB ID: 3MNP)とそのリガンド(デキサメタゾン)を用いました。

3. ドッキング計算

Parallel Chunkノードを使い、ドッキング計算を行うCompute LeadIT Docking ノードの同時実行する計算数を所有するライセンス数に応じて指定します。

4. スコア計算と結果の可視化

Assess Affinity with Hyde in SeeSAR ノードを用いて、リガンド-受容体間の結合自由エネルギー(ΔG)をHydeスコアにより評価します。ここで計算した結果はKNIMEの標準のテーブルとして出力できるため、得られた結果を他の解析ノードに対する入力として利用することも可能です。続いてInteractive SeeSAR Viewer ノードでリガンド中の原子ごとのHydeスコアを可視化できます。今回の結果では、グルココルチコイド受容体のリガンドとして、正解であるデキサメタゾンがスコア、ドッキング構造のいずれにおいても正しく評価され、本ワークフローの有用性が示されました。

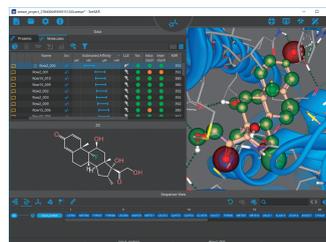


図3. SeeSARによる結果の可視化。リガンドの原子ごとの ΔG の寄与がマイナスであれば緑、プラスであれば赤に色付け。

■ まとめ

BioSolveIT社ではKNIME Interfacesとして23種類の解析ノードを提供しており、同社の創薬支援ツールによる計算手順のワークフロー化が実現できます。KNIMEワークフローにすることで、他の解析ツール(MOE, R, Pythonなど)との連携が容易になり、分子モデリングや機械学習と組み合わせた解析を行うことができます。

1) <http://www.knime.org/>

BioInfoGate社製OFF-Xは、医薬品のターゲット(分子作用機序)／医薬品分子／有害事象で分類された包括的に他に類を見ない医薬品安全性情報のクラウドサービスです。2019年3月末時点で、7千件以上の治療ターゲットおよび6千件の関連タンパク質について約27万件のアラート情報が収録されています。OFF-Xの新たなナレッジ・エリアとして、医薬品を閲覧するためのドラッグ・エリアが新たにリリースされました。ドラッグ・エリアはターゲット・エリアおよび有害事象エリアと完全に統合され、医薬品からターゲット(作用機序)と有害事象の関連性を迅速に把握できるようになりました。また、ターゲット・エリアには、50万件におよぶ参考文献をターゲットごとに整理して参照できる機能が追加されました。OFF-Xのコンテンツは、オフ・ターゲットを特定するための副次的薬理スクリーニング戦略を策定する上でとても有用です。ここでは、副次的薬理スクリーニングの支援および新機能であるドラッグ・エリアと文献の参照機能について紹介します。

■ 副次的薬理スクリーニングの支援

安全性の高い薬理的な戦略を立案するためには、開発中の医薬品に関わる可能性があるすべての安全性関連のターゲットを特性し、毒性や安全性の問題を引き起こすかもしれないオフ・ターゲット(例えば、薬物誘発性肝障害に係るBSEP阻害など)について状況を常に把握していることが求められます。

■ MRGPRX2

MAS 関連 GPR ファミリーメンバー X2 (MRGPRX2 : MAS related GPR family member X2) は、さまざまな有害事象に関与していることで最近注目されています。MRGPRX2は、自然免疫応答を調節する肥満細胞の活性化に関与しており、偽アレルギー反応(アナフィラキシー様反応)、蕁麻疹、喘息および喘息増悪、かゆみ、紅斑を伴った炎症状態に関連しています。また、MRGPRX2は、ニューロペプチド、システインプロテアーゼ、抗菌ペプチドおよびカチオン性タンパク質を含むさまざまなリガンド群と相互作用することが知られています(図1)。そのため、MRGPRX2は、低分子化合物を含むさまざまな種類の薬物とも相互作用することが予想できます。このことから、MRGPRX2は副次的薬理における重要なターゲット候補であるとみなせます。

このように安全性に関して注目すべきターゲットがある場合には、関連する有害事象、有害事象を引き起こした化合物の情報、さらには、ターゲットや安全性に関連するキーワードを含む文献などをOFF-Xで迅速に調査できます。

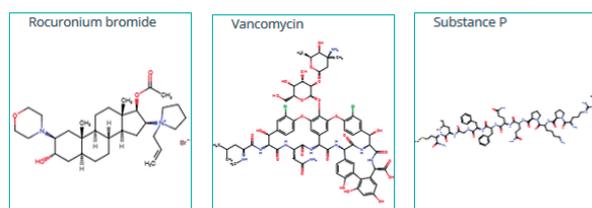


図1. MRGPRX2と相互作用する化合物の例

■ 副次的薬理スクリーニング

副次的薬理スクリーニングは、新薬の開発に伴う安全性

リスクを予測する上で効率的な実験手法です。OFF-Xは、副次的薬理スクリーニング戦略の立案を支援します。戦略を立案する上で、OFF-Xに登録されているターゲットのアラート情報と医薬品のオフ・ターゲット情報は、安全性関連の潜在的なターゲットの特定に貢献します。また、日々更新されるアラート情報を監視することにより、対象のターゲットの有害事象に関連する最新情報を入手できます。

■ OFF-Xの新機能

■ ドラッグ・エリアの追加

ドラッグ・エリアでは、2019年3月時点で、前臨床、臨床開発、発売後の医薬品を対象とした約1万化合物について安全性のアラート情報が登録されており、医薬品名を検索してアラート情報を閲覧できます。ドラッグ・エリアの基本情報には、医薬品の作用機序としてオン・ターゲットの作用機序である薬効薬理(Primary Pharmacology)と、OFF-Xのアラート情報から得られたオフ・ターゲット作用機序として副次的薬理(Secondary Pharmacology)が表示されます。医薬品のアラート情報は有害事象または日付により分類して閲覧できます。また、同じターゲットの医薬品における有害事象をテーブル形式で比較できます。ドラッグ・エリアとターゲット・エリアおよび有害事象エリアはお互いにリンクされています。ターゲット・エリアから医薬品を選択することによりドラッグ・エリアに移動することができます。有害事象エリアでは、有害事象に関連する化合物一覧をテーブル形式で参照できます。

■ 文献の参照機能

ターゲット・エリアには、OFF-Xに登録されている安全性アラート情報に加えてBioInfoGate社独自のテキストマイニング手法で得られた追加の参考文献(約50万件)を参照できる機能(Additional References)が追加されました。

登録される参考文献は日々更新されます。参考文献はターゲット/遺伝子/化合物名や安全性/毒性に関するキーワードを含んでおり、潜在的な安全性に影響の可能性のある内容も含まれています。この情報は、OFF-XのオプションモジュールであるOFF-X Safety Mapsにもリンクされています。

次世代シーケンサーデータ解析ソフトウェア

Flow : Single Cell RNA-seqデータ解析

Partek Flow (以下、Flow) は次世代シーケンサーデータ解析ソフトウェアです。次世代シーケンサー (以下、NGS) から出力されたリード配列を読み込んで、RNA-seqやDNA-seqのデータを解析することが可能です。Flowはリード配列からバーコードやUMIのトリミング、参照ゲノム配列へのアラインメント、UMIのデデュプリケーション、バーコードの絞り込み、細胞ごとの遺伝子カウントなど、Single Cell RNA-seqの前処理に必要な機能を搭載しています。また、細胞の絞り込み、遺伝子の絞り込み、PCAやt-SNEによる次元削減、クラスタリングによる細胞の分類、細胞群ごとの遺伝子発現解析、特徴遺伝子の抽出、生物学的解釈など、Single Cell RNA-seqのデータ解析を行うための豊富な機能を搭載しています。今回はFlowを用いたSingle Cell RNA-seqデータ解析について紹介します。

■ Single Cell RNA-seqとは

Single Cell RNA-seqはNGSを用いて細胞1つ1つの遺伝子発現を解析する手法です。マイクロ流路や液滴(ドロップレット)を使って1細胞ごとのcDNAライブラリーを調製する技術が発達してきたことにより、近年、がん・免疫・幹細胞・発生・脳神経など幅広い分野で利用が進んでいます。多数の細胞の平均値としての遺伝子発現を解析するRNA-seqに比べて、Single Cell RNA-seqは細胞集団の不均一性や細胞のクローン進化などの研究に活用されています。

Single Cell RNA-seqでは、cDNAライブラリーの調製手法によって異なりますが数百から数万の細胞に由来するリード配列が出力されます。そのため、1回のランで多数の細胞をまとめて解析するためのバーコードやPCRによる増幅バイアスの影響を取り除くためのUMI (Unique Molecular Identifiers)などの特殊な塩基配列がリード配列に含まれます。Single Cell RNA-seqのデータ解析では、バーコードやUMIを除去して細胞ごと転写産物ごとに解析できるようにリード配列を処理する必要があります。

Single Cell RNA-seqの解析目的には、細胞集団の組成や各細胞型のマーカー遺伝子を調べる多クラスデータ解析、細胞系譜や遺伝子発現制御ネットワークを調べる擬時間解析、細胞の空間配置や遺伝子の空間発現パターンを調べる空間再構築解析などがあります。多クラスデータ解析や擬時間解析では細胞当たり数千遺伝子の発現量を2次元や3次元に圧縮することで細胞集団や細胞系譜を可視化します。2次元や3次元のプロットに表示された細胞を手作業や複数のクラスタリング手法によりクラスターに分類して細胞集団や細胞系譜を同定します。クラスターごとのマーカー遺伝子を抽出してGO解析やパスウェイ解析を行うことで生物学的解釈が得られます。

■ バーコードやUMIの除去

Single Cell RNA-seqでは、第一鎖と第二鎖にプロトコルごとに異なるバーコードやUMIなどの特殊な塩基配列を含みます。Flowは主要なプロトコルのリード配列の構造およびバーコードやUMIの配列がライブラリーとして提供されているので、簡単にバーコードやUMIを取り除くことができます。

2019年3月現在、10x Chromium Single Cell 3' v2/v3、1CellBio inDrop、Drop-seq、Fluidigm C1 mRNA-seq HT IFC、Rubicon Genomics ThruPLEX Tag-seq、BioRad SureCell WTA 3'のプロトコルに対応しています。

■ UMI デデュプリケーション

Single Cell RNA-seqでは、1細胞にごくわずかに含まれるRNAを検出するためにRNAをcDNAに逆転写した後でPCRやIVTを使って増幅します。その際、RNAの長さやGC含量などによって増幅されやすさが異なるためPCRバイアスが生じます。

PCRバイアスによる発現定量性の問題を回避するためにcDNAには1分子ごとに異なる6~10塩基のUMIと呼ばれるランダムな塩基配列が付加されます。1遺伝子につき何種類のUMIが対応しているかをカウントすることで、PCRバイアスによらず細胞内でのRNAのコピー数を正確に見積もることができます。Flowは数クリックで簡単にUMIの重複を除去して遺伝子ごとのUMI数をカウントできます。

■ バーコードの絞り込み

ドロップレットやナノウェルを利用したcDNAライブラリーの調製では、ドロップレットやナノウェル中に1個の細胞とバーコード配列を持ったプライマーのついたビーズを含むように調製します。その結果、1実験で1つのFASTQファイルが生成され、FASTQファイル内に含まれるリード配列がどの細胞に由来するかをバーコードで識別します。

大部分のドロップレットやナノウェル中に1個の細胞が取り込まれるように細胞の濃度を調製しますが、複数の細胞が取り込まれたり1個の細胞も取り込まれなかったりするドロップレットやナノウェルも存在します。Flowは横軸にリード数の降順に並べたバーコード、縦軸にリード数を表示するグラフ(図1)を使って細胞が取り込まれなかったバーコードを除去することができます。バーコード当たりのリード数が急減している部分(図1の中央やや右)に境界線が自動的に引かれます。境界線より左のバーコードは細胞が取り込まれたドロップレット、右のバーコードは細胞が取り込まれなかったドロップレットです。

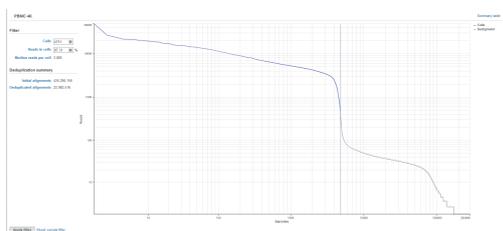


図1. バーコードの絞り込み(細胞を含まない液滴の除去)

■細胞や遺伝子の絞り込み

前処理を終えて細胞当たり遺伝子当たりのカウントマトリックスが得られたらQA/QCを実施します。QA/QCにより細胞当たりのカウント数、細胞当たりの検出遺伝子数、細胞当たりのミトコンドリア由来のリード割合がバイオリンプロットで表示されます(図2)。これらのプロットにより複数の細胞が取り込まれたドロプレットを除外するなどの絞り込みができます。

Single Cell RNA-seqでは細胞当たり3,000~10,000個の遺伝子が検出されます。そのため、カウントマトリックスにはほとんどの細胞でカウント数が0の検出されない遺伝子も含まれます。Flowはノイズ低減フィルター、統計フィルター、遺伝子リストフィルターにより遺伝子の絞り込みができます。例えば、ほとんどの細胞でカウント数が0の遺伝子はノイズ低減フィルターにより除去されます。

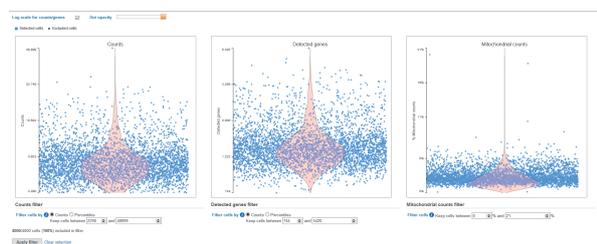


図2. Single Cell QA/QCの結果

■t-SNEプロットによる次元削減

細胞や遺伝子の絞り込みおよび正規化などを行ったカウントマトリックスは次元削減を行い細胞集団や細胞系譜を可視化します。FlowはPCA(主成分分析)とt-SNE(t分布型確率的近傍埋め込み)の2種類の次元削減の手法を搭載しています。PCAは線形変換系手法なので、非線形構造を持つデータに対して適切な低次元表現が得られない、局所的なデータ間の関係は信頼性が低い、といった問題があります。そのため、Single Cell RNA-seqではt-SNEプロットが広く利用されています。図3は星状細胞腫の3次元t-SNEプロットです。あらかじめ分かっているマーカー遺伝子BCANの発現量で色付けしています。BCANの発現により2つの細胞集団に分割されていることがわかります。t-SNEプロットでは投げ縄ツールを使ってフリーハンドで細胞集団を選択できます。

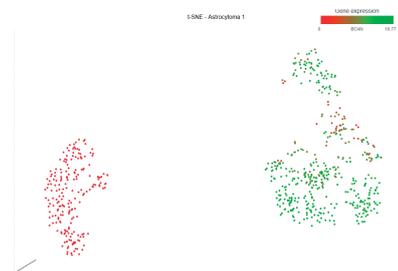


図3. t-SNEプロット

■クラスタリング

正規化後のカウントマトリックスをクラスター解析することで細胞集団や細胞系譜を同定します。Flowは階層型クラスタリング、k平均法クラスタリング、グラフに基づくクラスタリングの3種類のクラスタリング手法を搭載しています。Single Cell RNA-seqでは細胞間の類似グラフを利用したグラフに基づくクラスタリングが広く利用されています。グラフに基づくクラスタリングでは解析に用いる主成分の数をパラメーターとして設定します。解析に用いる主成分の数は主成分分析の結果として得られるスクリープロットで各主成分の固有値の大きさから決定します。図4は末梢血単核細胞4,000個をグラフに基づくクラスタリングでクラスター解析した結果を3次元t-SNEプロットで表示しています。11個のクラスターにきれいに分割されていることがわかります。また、各クラスターに特徴的な上位10個のマーカー遺伝子の表も合わせて出力されます(表1)。

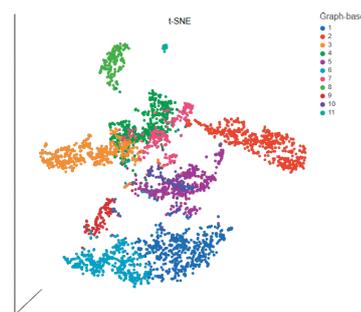


図4. クラスタ解析結果のt-SNEプロット

表1. 各クラスターのマーカー遺伝子

Top features	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11
1	LEF1	KIFAC	STROA2	TRGC2	IL7R	TRAC	CPVL	KLRF1	NOG	CDKN1C	LILRA4
2	NOG	WPRE33	S100A8	CD3E	AZP3	LEF1	LGALS2	SPON2	LEF1	TESK4	TPR2
3	CD37	IGHM	VCAN	CST7	TNFRSF4	NOG	CSTA	CLIC3	PRKCO	AS1	TOF7L2
4	AP001324.1	CD79A	CD14	NRG7	IL32	CD37	IGSF6	FGFBP2	TRAC	FCGR3A	LINC00996
5	RPS6	FCER2	LGALS2	G2MA	TRAC	CD3E	PLBD1	GZMB	GIMAP7	SMM25	EMASE1L3
6	TRAC	IGHD	LYZ	GZMK	ITGB1	TCF7	MNDA	GRLY	MAL	CSF1R	PTCRA
7	TCF7	BANK1	CSTA	LYAR	CD2	CD3D	CFP	KLRD1	CD3E	AC104809.2	SMPD3
8	CD7	LINC00926	MNDA	CTSW	CD48LG	RPL39P4	ALDH2	PRF1	TREB2	HMOX1	GAS6
9	TRAC2A	MSI41	CYBB	KLRF1	TRACD	ALM0873.1	FCM1	TTCC8	LEHB	SSG12.10	JCH4H1
10	CD3E	HLA-DQB	FCM1	B2M	OPTN	LDHB	NCF2	KLZ	LDLRAP1	MS4A7	PLD4

■ご評価

Flowは無償トライアルにより4週間ご評価いただけます。Single Cell RNA-seqのデータ解析を行いたい方はぜひFlowをお試し下さい。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。

材料設計支援統合システム

ハイブリッド汎関数による高精度なバンド構造評価

MedeA-VASPで実装されているHSE06ハイブリッド汎関数を用いてInAsのエネルギーバンド構造を評価します。結果は0.35 eVのバンドギャップを与え実験値と一致します。また最も低い伝導帯中の Γ 点とL点とのエネルギー間隔(Γ -L separation)は、1.11 eVになると算出されています。以前のハンドブックでは0.74 eVという値でしたが、最近の実験では、 1.1 ± 0.05 eVと報告されており、この測定値は事例の結果とよく一致します。

本稿では、III-V族化合物半導体で電子・光学デバイスのバンド構造エンジニアリングに必要となる、固体エネルギーバンド構造の詳細を計算するため、Hartree-Fock-DFTハイブリッド汎関数であるHSE06を用いた事例について報告します。

■ 結晶構造

InAsの単位格子を、MedeAで利用可能なデータベースで検索します。このIII-V族化合物半導体の構造は、格子長が報告値で6.0584~6.060 Åとなる立方晶の閃亜鉛鉱型構造の結晶です。本事例のシミュレーションでは、過去の計算と直接比較することが可能となるよう¹⁾、Thompsonらの報告にある6.0588 Åを使用します²⁾。交換相関汎関数PBEsol³⁾を用いた構造最適化では、実験結果よりわずかに0.6%だけ大きい格子定数となります。

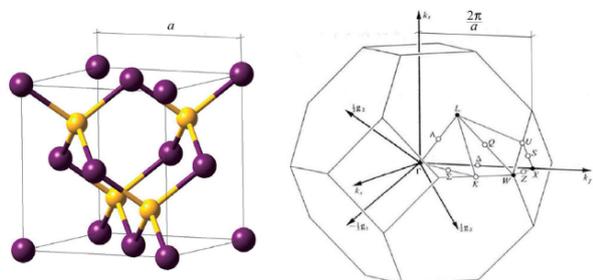


図1. 左図: InAsの結晶構造。右図: 対応する第1ブリルアンゾーン。高対称点および線は標準表記に従ってラベル付け。

■ バンド構造計算

エネルギーバンド構造は、MedeAソフトウェア環境に搭載されているVASPで計算しました。Heyd-Scuseria-Ernzerhof (HSE06)のハイブリッド汎関数⁴⁾を使用して、平面波カットオフ値およびk-メッシュはMedeAのデフォルトパラメーターが適用されています。また、非局所項については汎関数PBEsolを用い³⁾、バンド分散全体のプロット数は160に設定しました。その結果を図2に示しました。

エネルギー準位 -17 eVと-16 eVのあいだにあるフラットなバンド構造は、局在化されたIn $4d^{5/2}$ および $4d^{5/2}$ の準内殻軌道によって形成され、より高い準位の

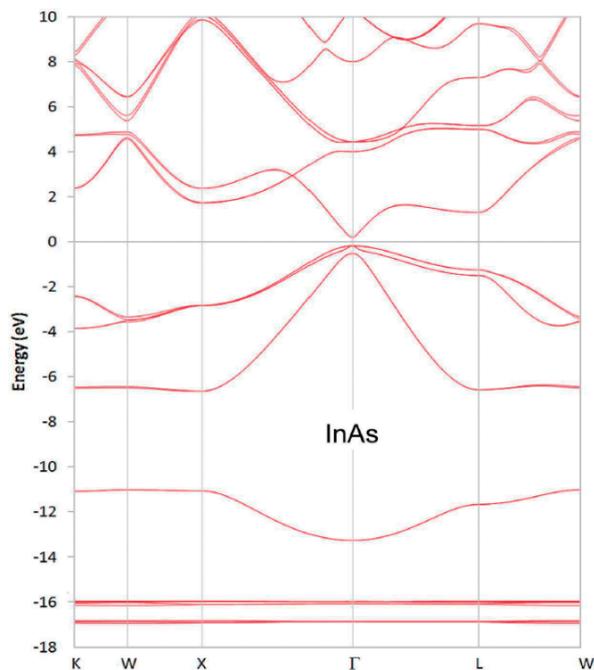


図2. InAsのエネルギーバンド構造(ハイブリッド汎関数 HSE06 とスピン-軌道相互作用相対論的ハミルトニアンを用いて算出)

すべてのバンド構造は、両方の原子種のs軌道とp軌道の重なりから生じています。図2のバンドギャップ近傍の詳細と実験データとの比較を図3に示しました。実験的な測定によるバンドギャップの値 $E_g = 0.35$ eVと比較すると、計算による値は見事に一致しています。また、計算と測定から得られたスピン-軌道相互作用による分裂 E_{SO} は、 Γ 点でかなり良好な一致を示しています。

それとは対照的に、参考文献⁶⁾で報告されているX点とL点の谷底の位置では、計算されたデータと著しい不一致があります。事実、以前のハンドブックの値では Γ -L separation、すなわち伝導帯の底とL点最小 E_L とのエネルギー差は、0.74 eVと報告されており⁵⁾、Ioffe Institute⁶⁾によって与えられたものと非常に近い値、つまり、

$E_L - E_g = 1.08 \text{ eV} - 0.35 \text{ eV} = 0.73 \text{ eV}$ (図3の上部参照) ですが、本稿の計算では、

$$E_L - E_g = 1.46 \text{ eV} - 0.35 \text{ eV} = 1.11 \text{ eV}$$

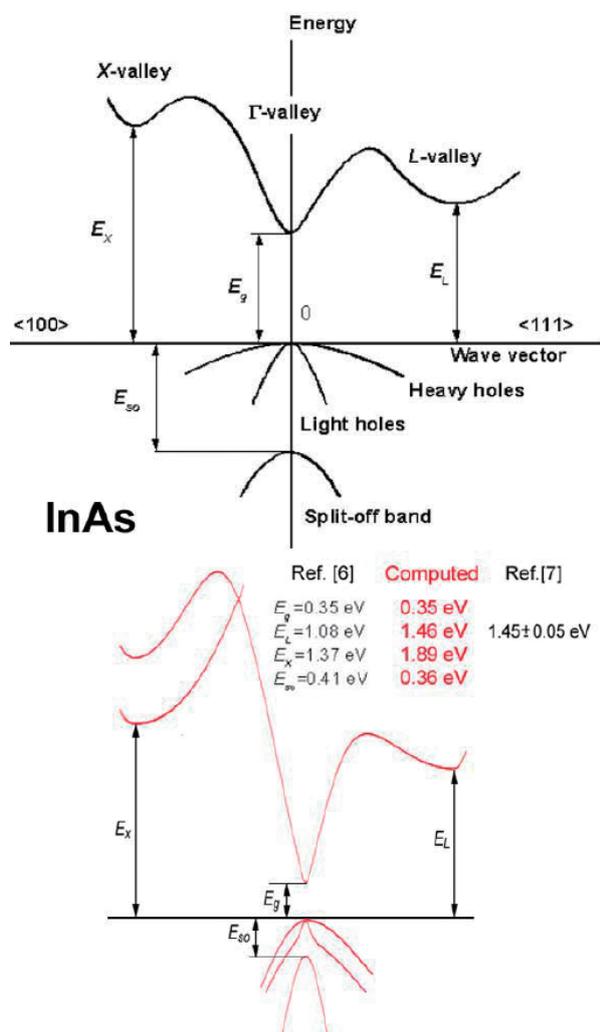


図3. エネルギーギャップ近傍のエネルギーバンド構造の詳細。Ioffe Institute のウェブサイト⁶⁾で公表されたデータと計算されたデータおよび最近の測定値⁷⁾との比較。X点およびL点における谷底の位置の著しい違いに注意。

を与えます。

これまでのScreened Exchange法によるFLAPWを用いた計算では、 Γ -L separationとして1.21 eVという値が得られていますが、こちらは本稿の結果に近い値です。興味深いことに、Characheらによる実験から、測定値 1.1 ± 0.05 eV⁷⁾が得られており、これは本稿の結果とよく一致します。

結論として、異なる計算手法と異なる交換相関汎関数を用いた2つの独立した計算と、より近年の実験データでは、 Γ -L separationの値として1.1~1.2 eVの値が得られており、以前ハンドブックによって報告されて

いた値0.74 eVの妥当性について疑問を投げかけています。おそらく Γ -X separationの値についても、Ioffe Instituteの図3に示されたデータより本稿の計算結果の方が、より信頼性が高いと推察されます。

まとめ

本稿ではInAsのエネルギーバンド、特にその微細構造を例として、精度の高い第一原理電子状態計算を用いることで、実験に匹敵する精度と信頼性を持つ材料特性のデータが得られることを示しました。今回示したような計算は、系統的に広範な系に適用することができます。また、経験的なパラメーターに依存しないため、計算結果は高い予測能力を持つと期待されます。第一原理計算の適用範囲はエネルギー状態の評価に限定されません。振動や応力などの機械特性の評価に用いることもできますし、計算対象とする材料も合金をはじめ表面・界面のモデルなど、幅広く対応可能です。実験では評価が困難で測定に時間のかかるものやコストの高いものを代替するポテンシャルを秘めています。

参考文献

- 1) C.B. Geller, W. Wolf, S. Picozzi, A. Continenza, R. Asahi, W. Mannstadt, A.J. Freeman, and E. Wimmer, *Appl. Phys. Lett.* **79**, 368 (2001)
- 2) A.G. Thompson, J.E. Rowe, and M. Rubenstein, *Journal of Applied Physics* **40**, 3280-3288 (1969); Pauling file #530616
- 3) J.P. Perdew, A. Ruzsinsky, G.I. Csonka, O.A. Vydrov, G.E. Scuseria, L.A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.* **100**, 136406 (2008)
- 4) J. Heyd, G.E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **124**, 219906 (2006)
- 5) *Handbook Series on Semiconductor Parameters*, edited by M. Levenshtein, S. Rumyantsev, and M. Shur, Vol. 1, World Scientific, New York (1996)
- 6) <http://www.ioffe.rssi.ru/SVA/NSM/Semicond/InAs/Figs/721.gif>
- 7) G.W. Charache, D.M. DePoy, J.E. Raynolds, P.F. Baldasaro, K.E. Miyano, T. Holden, F.H. Pollak, P.R. Sharps, M.L. Timmons, C.B. Geller, W. Mannstadt, R. Asahi, A.J. Freeman, and W. Wolf, *J. Appl. Phys.* **86**, 452 (1999)

学会・展示会出展およびセミナー開催予定

今年度上半期の弊社の主な学会・展示会出展およびセミナー開催の予定を紹介します。学会・展示会では弊社取り扱い製品の説明やデモンストレーションを行います。製品についてのお問い合わせや、実際の操作をご覧になりたいなど、何かございましたら、ぜひ弊社ブースにお立ち寄りください。また、セミナーについては、弊社社員による製品紹介や、開発元や研究者の方による招待講演などを予定しています。セミナーは、ユーザーに限らずどなたでもご参加できます[※])ので、どうぞお気軽にご参加ください。

■ マテリアルサイエンス分野

学会・展示会出展

- ・ナノ学会第17回大会
会期：2019年5月9～11日
会場：かごしま県民交流センター
展示製品：計算科学統合プラットフォーム MedeA、AMS、SciMAPS等
- ・第68回高分子学会年次大会
会期：2019年5月29～31日
会場：大阪府立国際会議場
展示製品：計算科学統合プラットフォーム SciMAPS、高精度分子シミュレーション支援ソフトウェア Direct Force Field等
- ・日本コンピュータ化学会2019年春季年会
会期：2019年6月6～7日
会場：東京工業大学 大岡山キャンパス
展示製品：量子化学計算ソフトウェア Gaussian、AMS、TURBOMOLE、Molpro、MOPAC2016等

■ ライフサイエンス分野

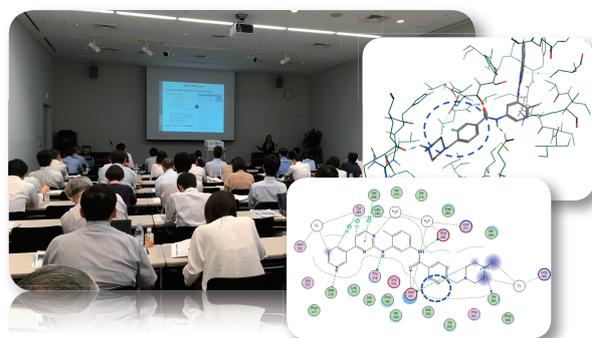
学会・展示会出展

- ・第19回日本蛋白質科学会年会
会期：2019年6月24～26日
会場：神戸国際会議場 神戸国際展示場
展示製品：統合計算化学システム MOE、タンパク質立体構造データベースシステム PSILO、BioSolveIT 社創薬支援ツール等

- ・第1回 ファーマラボ EXPO 医薬品研究開発展
会期：2019年7月3～5日
会場：東京ビッグサイト 青海展示棟
展示製品：MOE、PSILO、研究情報共有システム CBIS、ターゲット・作用機序予測プラットフォーム Chemotargets CLARITY、安全性評価とリスク評価のためのプラットフォーム ChemTunes & ToxGPS等

公開セミナー

- ・MOEフォーラム2019
日時：2019年7月24日(水) 10:00～17:30
会場：大手町サンケイプラザ
開発元によるMOEの最新情報の発表紹介、研究者による応用事例紹介の招待講演を予定しています。



■ 最新情報はこちら

<https://www.molsis.co.jp/news/seminar/>

※) トレーニングコースはユーザー限定の場合があります。詳細は弊社までお問い合わせください。

- ・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。その他に発送をご希望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. 9, 2019
2019年4月1日発行
発行所 株式会社モルシス
Copyright © 2019 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