

## 世界最大の分子生成AIデータエコシステムによる日本の化学力強化に関する調査・実証

<b>実施者</b>	SyntheticGestalt株式会社
<b>概要</b>	<p><b>世界最大規模のFEPデータ基盤および結合親和性基盤モデルの構築とデータエコシステムの有効性実証。</b></p> <ol style="list-style-type: none"> <li>国内有数の製薬・農薬・化学企業と協力し、10万件以上のFEPデータ収集に向けた標的タンパク質・リガンド分子を決定。</li> <li>統計的信頼性を確保するための独立したFEP計算実施に基づく計算プロトコルの決定、及びベンチマーク検証を踏まえた調整を実施。</li> <li>100種類のタンパク質に対して10万件以上のFEP計算を実施。収集データを用いたタンパク質・リガンド分子ペア情報の基盤モデルを構築、従来の数万倍高速化でバーチャルスクリーニングを実現。</li> <li>SPR・ITC試験で1万件規模の実測データと照合し、データ品質を保証。</li> <li>基盤モデルを用いたバーチャルスクリーニングを実施し、データの利活用事例を示す。</li> </ol>

### 提案の背景・社会的現状

分子AI開発には課題が存在

- データ不足が最大障壁: 統一手法で収集された大量データが存在しない
- 高コストで実用化困難: FEP計算は高精度だが膨大な計算コストを要する
- 各機関データが活用不可: 測定条件が不統一のため機械学習に使用できない

SyntheticGestalt社の強みで競争力向上

- 世界トップクラスの国内化学・製薬企業との連携や世界最高水準の分子AI技術を活かせる
- データとアーキテクチャが性能を決定する分子AI分野で世界をリードできる可能性

### 実施内容

- 取得データ仕様決定
- データ計算プロトコル決定と調整
- 計算データ構築とAI予測モデル開発
- 実験評価
- AI予測モデルの実用性検証

### 社会実装の方法

- データセット販売:** AWS Data ExchangeやGoogle Cloud Marketplaceを通じた公平なサブスクリプション販売
- モデルの販売・開発:** データセットと同手法で展開・共同開発の2形態で展開
- 技術レポートの公開:** タンパク質ファミリー別の計算精度と適用可能範囲を体系化した技術知見を業界全体に公開

成果物の公開



データセット  
(公開)



モデル  
(公開)



レポート  
(公開)

# 世界最大の分子生成AIデータエコシステムによる日本の化学力強化に関する調査・実証

## データエコシステム図

■ : ステークホルダー

□ (点線) : 実証範囲

→ : データの流れ

## 補足

