

平成22年度 成果報告書

「パターン認識アルゴリズムに基づく高精度な創薬シード・リード化合物探索手法のシステム開発」

目次

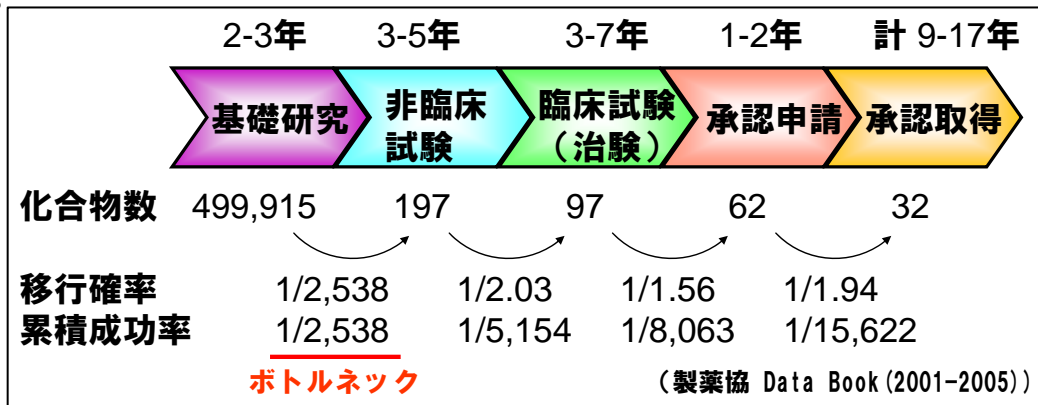
1	研究開発課題の背景	3
2	研究開発の全体計画	
2-1	研究開発課題の概要	5
2-2	研究開発の最終目標	6
2-3	研究開発の年度別計画	7
3	研究開発体制	
3-1	研究開発実施体制	8
4	研究開発実施状況	
4-1	GPCR 予測モデルによる予測精度の向上	9
4-1-1	GPCR 予測モデルの構築方法と検証	9
4-1-2	大規模相互作用データを用いた学習モデルの構築	11
4-1-3	β 2AR 及び NPY1R に対する活性予測と実証実験	12
4-1-4	CGBVS と既存手法 (LBVS・SBVS) との比較・精度検証	13
4-1-5	まとめ	16
4-2	イオンチャネル・キナーゼ予測モデル評価検討	17
4-2-1	イオンチャネルを標的とした化合物探索モデルの構築及び評価検証	17
4-2-2	イオンチャネル予測モデルの構築と実証実験	18
4-2-3	キナーゼを標的とした化合物探索モデルの構築と評価検証	18
4-2-4	キナーゼ予測モデルによる未知化合物の予測と実証実験	19
4-2-5	キナーゼ予測モデルによる選択性予測についての実験検証	20
4-2-6	まとめ	21
4-3	(A) 連結モジュール・入力変換部の作製	22
4-3-1	詳細仕様設計 (20 年度実施分)	22
4-3-2	詳細仕様設計 (21 年度実施分)	23
4-3-3	詳細仕様設計 (22 年度実施分)	26
4-3-4	まとめ	28
4-4	(B) 相互作用マシンラーニング予測モジュールの作製	28
4-4-1	詳細仕様設計 (20 年度分)	28
4-4-2	詳細仕様設計 (21 年度分)	30
4-4-3	詳細仕様設計 (22 年度分)	32
4-4-4	まとめ	35
4-5	(C) 連結モジュール・スコア統合部の作製	35
4-5-1	MOE 連結モジュール詳細仕様設計 (21 年度実施分)	35
4-5-2	大規模データへの対応 (22 年度実施分)	36
4-5-3	まとめ	36
4-6	総合システムの評価	36

4-6-1	CGBVS 及び SBVS による SRC キナーゼ阻害剤予測.....	36
4-6-2	まとめ.....	38
4-7	総括.....	38
5	参考資料	
5-1	研究発表・講演等一覧	39
5-2	産業財産権	40

1 研究開発課題の背景

(1) 医薬品開発の現状と問題点

医薬品開発には、膨大な時間と費用を要する。日本製薬工業協会のまとめ（2001～2005年）では、実に開発期間9～17年、開発費用1新薬あたり約500億円、医薬品開発の成功確率15,622分の1と見積もられている。特に、化合物ライブラリーから非臨床段階までの成功確率が0.04%と非常に低い値を示していることから、膨大な種類の化合物ライブラリーからヒット化合物を見つけ出す工程は、医薬品開発の効率化のボトルネックになっている（図1-1）。

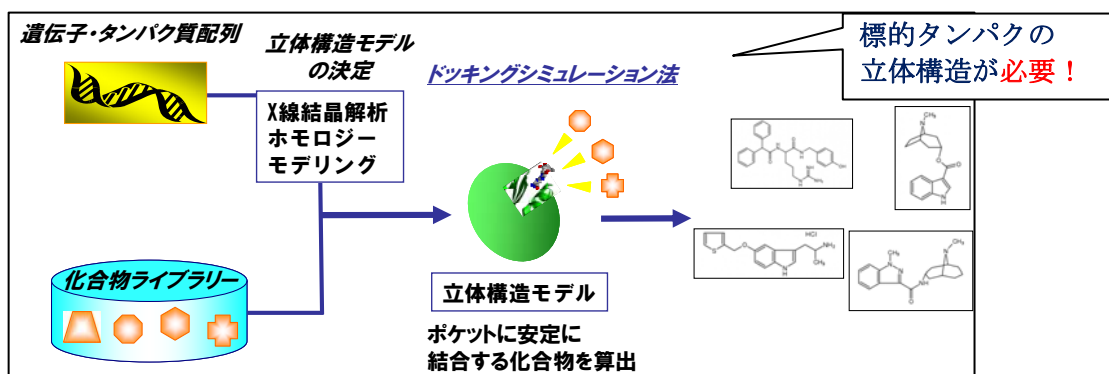


(図 1-1. 創薬プロセスの移行確率・累積性効率)

上記の工程を加速化する最も有力な方法として、計算機の中で化合物スクリーニングを行うインシリコスクリーニングが実践されている。また化合物探索の世界的プロジェクトとして、近年、「ケミカルゲノミクス・ケミカルバイオロジー」と称し各国が国策として取り組んでいる。

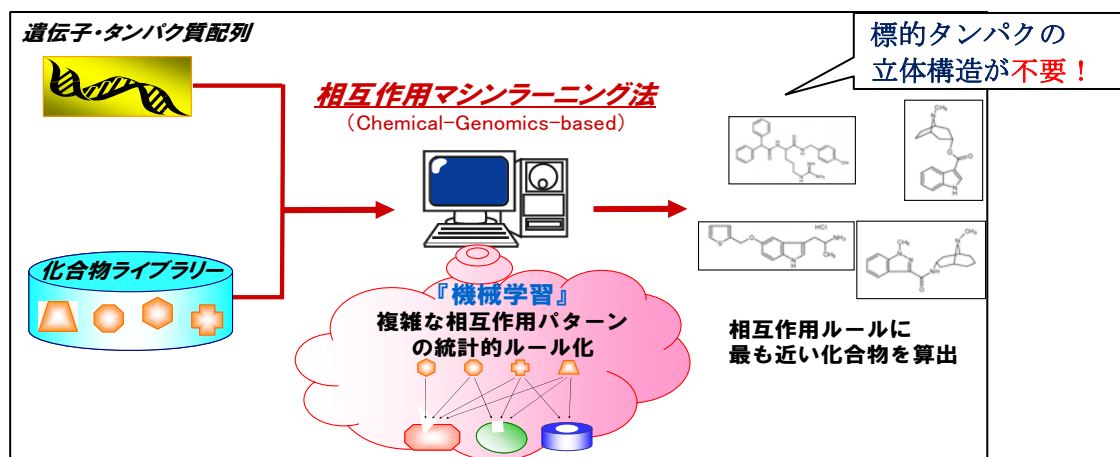
(2) 従来法の問題点と新手法による改善点

一般に、広く用いられているインシリコスクリーニング手法は、薬物とタンパク質の複合体立体構造モデルを科学的にシミュレートするドッキングシミュレーション法である（図1-2）。しかしながら、市販されている医薬品のターゲットタンパクの大半は、立体構造情報を得るのが難しいとされる膜タンパクであり、タンパク質の立体構造情報に依存しない新しい予測法の開発が現実問題としては重要な課題とされていた。



(図 1-2. 従来法：ドッキングシミュレーション法)

このような背景を鑑み、当社は、従来法であるドッキングシミュレーションとは全く概念の異なる独自の方法論（相互作用マシンラーニング法）の開発を行い、実用化に成功している（図 1-3）。



(図 1-3. 新規手法：相互作用マシンラーニング法)

この計算手法は、従来のインシリコスクリーニング技術であるドッキング計算とは異なり、立体構造情報を用いずに、ケミカルゲノクス情報（タンパク質－化合物の網羅的相互作用情報）のパターン認識に基づく機械学習アルゴリズム(サポートベクターマシーン)を用いて化合物予測を行う世界に類を見ない方法論であり、すでに、共同研究先である京都大学薬学研究科奥野研究室において、GPCR ファミリーについてその予測率と新規骨格発見能力の高さが証明されたところである。

当社は、これらの新規手法をもとにして平成 20 年 3 月 31 日に起業された企業であり、受託解析事業として事業を開始し、すでに製薬企業より数件の受注を行っている。しかしながら、これらは相応のスキルを有する研究者が、手動ベースで解析を行っているものであり、民間ユーザーが手軽に利用できるソフトウェアとしての提供が待たれている。また、普及のためには GPCR ファミリー以外の主要創薬ターゲットであるイオンチャネル・キナーゼファミリーへの手法適応の研究開発及び各々の学習モデルの精度向上が必要とされる。

2 研究開発の全体計画

2-1 研究開発課題の概要

① 創薬シード・リード化合物探索システム(相互作用マシンラーニングモジュール)のパッケージ化：

京都大学の特許技術を基本にした予測プログラムは、それぞれの計算ステップが断片化されているため専門研究者のマニュアル操作によって実行している。これらステップごとに断片化したプログラムをパッケージとして1本化する。

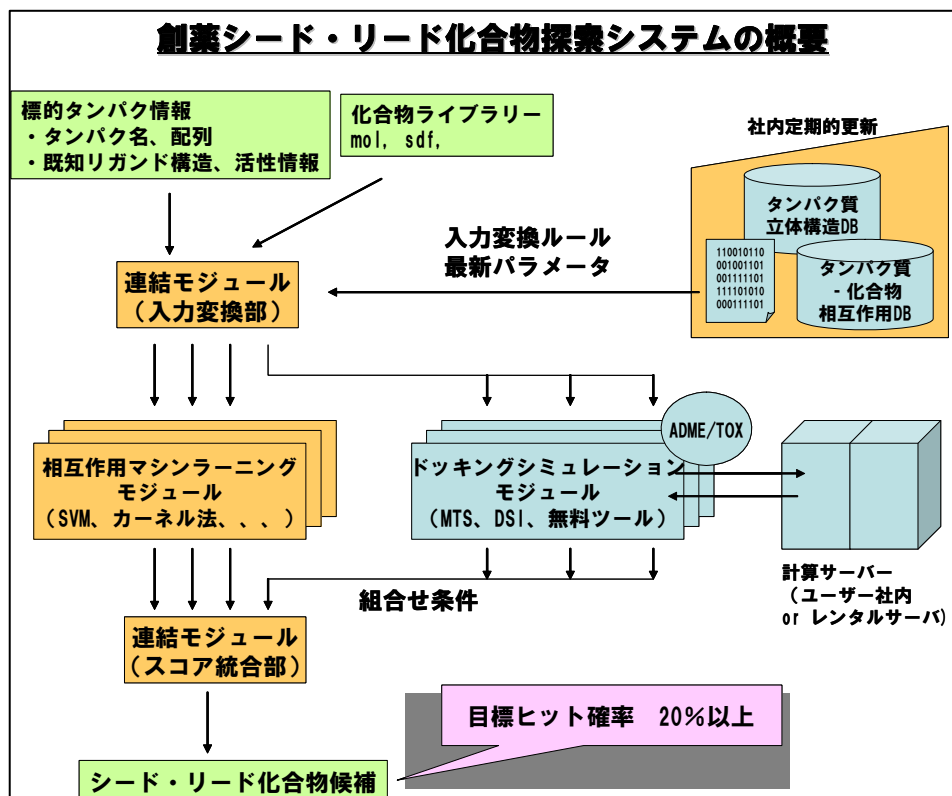


図 2-1. 創薬シード・リード化合物探索システムの概要

上図 2-1 に示すとおり、製品は、連結モジュール（入力変換部）、連結モジュール（スコア統合部）、相互作用マシンラーニング予測モジュールとドッキングシミュレーション予測モジュールから構成される。製品は、標的タンパク情報（タンパク名、配列、既知リガンド構造、活性情報）と予測対象の化合物ライブラリーの化学構造情報（mol, sdf 形式）をユーザ入力情報とし、標的タンパク質をターゲットとするシード・リード化合物候補を予測結果として出力する。

[各モジュールの作製について]

A. 連結モジュール・入力変換部の作製

連結モジュール・入力変換部は、ユーザ入力インターフェースとしての機能を持ち、標的タンパク質名などユーザ入力情報を各予測モジュールの入力フォーマット（例、相互作用パターンの教師セット、立体構造）へと変換するとともに、各予測モジュールへの並列ジョブ入力を実践する。ここで、変換処理を迅速に行うために、独自の変換ルールや、予測精度の精密化を目指した標的タンパクごとの最適パラメータの作成も適宜行う。

B. 相互作用マシンラーニング予測モジュールの作製

現在、開発代表者が実行している相互作用マシンラーニング予測プログラムは、それぞれの計算ステップが断片化されているため専門研究者のマニュアル操作によって実行している。これらステップごとに断片化されたプログラムをパッケージとして1本化する。

C. 連結モジュール・スコア統合部の作製

連結モジュール・スコア統合部は、予測モジュールの各ソフトで予測された予測スコアを組み合わせた総合スコアを算出し、総合スコアの高い順にシード・リード化合物候補を出力する。

② 外部ソフトとの連携のための連結モジュールの付加による統合システムの開発

当社の主要技術である相互作用マシンラーニング法は、既存手法に置き換わるというよりも、既存手法(ドッキングシミュレーション法等)と組み合わせることで、より高い予測精度を期待できる。本開発システムでは、これらの高い予測精度を示す複数の方法を組み合わせることにより、さらなる予測性能の向上を目指す。

パラメータ条件を変えながら計算を行うことにより、計算結果の正解傾向とパラメータとの因果関係やソフト間での予測傾向などを詳細に分析、性能分析を行った個々の予測ソフトの組合せを図ることにより、予測精度を向上させる最適な組合せ条件とパラメータを決定する。

これらによる精度向上方策を確立した上で、既存手法の予測ソフト(無償、市販製品)との組み合わせ予測を実現する連結モジュールを開発する。連結モジュールは、入力変換部とスコア統合部とに分ける。

2-2 研究開発の最終目標 (平成 22 年 9 月末)

- ① 創薬シード・リード化合物探索システム(相互作用マシンラーニングモジュール)のパッケージ化
 - ・ 各モジュール(連結モジュール・入力変換部の作製、相互作用マシンラーニング予測モジュールの作製)の完成。
 - ・ 目標ヒット率は μ Mオーダーで 10%。
- ② 外部ソフトとの連携のための連結モジュールの付加による統合システムの開発
 - ・ 代表的なドッキングシミュレーションソフトであるMOEとの連結化の完成。
 - ・ 目標ヒット率は μ Mオーダーで 20%。

2-3 研究開発の年度別計画

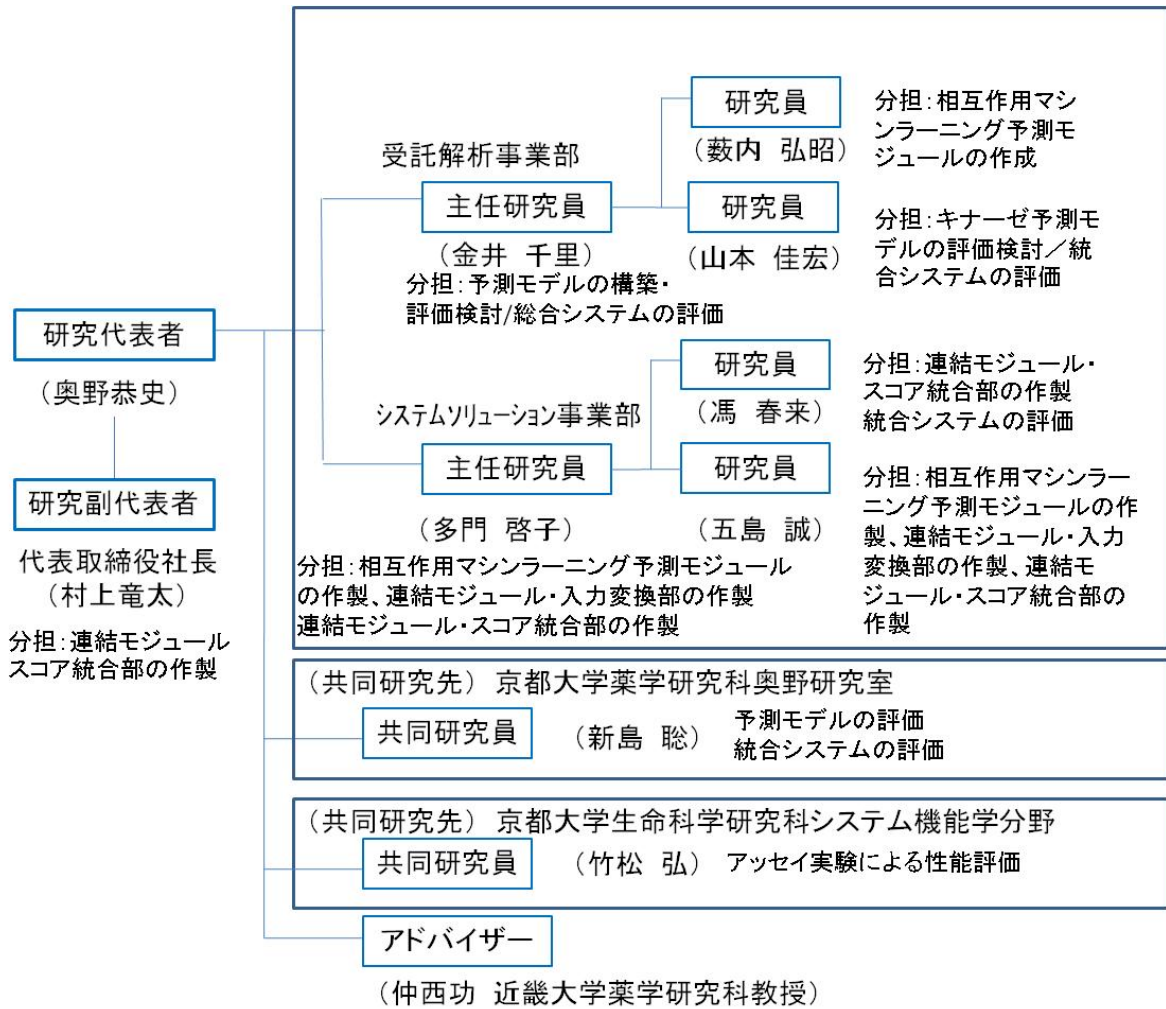
金額は非公表

研究開発項目	20年度	21年度	22年度	計	備考
パターン認識アルゴリズムに基づく高精度な創薬シード・リード化合物探索手法のシステム開発					
①創薬シード・リード化合物探索システム(相互作用マシンラーニングモジュール)のパッケージ化	—	—	—	—	
②外部ソフトとの連携のための連結モジュール(スコア結合部)の製作		—	—	—	
間接経費額(税込み)	—	—	—	—	
合計	—	—	—	—	

- 注) 1 経費は研究開発項目毎に消費税を含めた額で計上。また、間接経費は直接経費の30%を上限として計上(消費税を含む)。
 2 備考欄に再委託先機関名を記載
 3 年度の欄は研究開発期間の当初年度から記載。

3 研究開発体制

3-1 研究開発実施体制



4 研究開発実施状況

4-1 GPCR 予測モデルによる予測精度の向上

4-1-1 GPCR 予測モデルの構築方法と検証

1. 化合物-GPCR 相互作用データの収集及び交差検定法によるモデルの評価

相互作用マシンラーニング法 (CGBVS: Chemical-Genomics-based-Virtual-Screening) を実行するには、多くの「化合物 - タンパク質間相互作用データ」が必要であるが、それらのデータの大部分は製薬企業や各研究室の極秘資料となっている。そこで、当社はまず、文献などから利用可能な化合物-タンパク質相互作用データを収集、整備することを始め、データ収集の対象とするタンパク質として、創薬標的として最も注目されている遺伝子ファミリーのひとつである GPCR (G タンパク質共役型受容体) に的を絞った。

さらに、既存のスクリーニング手法との比較検討を行うため、今回開発した CGBVS と LBVS (リガンド類似性に基づくスクリーニング法) との予測性能を比較した。収集した化合物-GPCR 相互作用の一部を用いて、負例を交換しながら 5 分割交差検定法 (fivefold cross-validation) を試行した。ただし、負例については、データが入手できなかったため、ここでは化合物-タンパク質ペアをランダムに組み合わせ生成した。交差検定の結果、最近傍法を用いた LBVS では $84.4 \pm 0.3\%$ 、CGBVS では $91.6\% \pm 0.2\%$ の相互作用を正しく予測した。また、ROC 曲線からも、CGBVS の予測性能の高さが確認された (右図 4-1)。したがって、ケミカルゲノクス情報の活用がリガンド予測性能の向上につながったといえる。なお、予測に要する計算的負荷が一般的な LBVS によるものとほとんど変わらないことから、CGBVS は大規模スクリーニングに十分適用可能との結果が得られた。

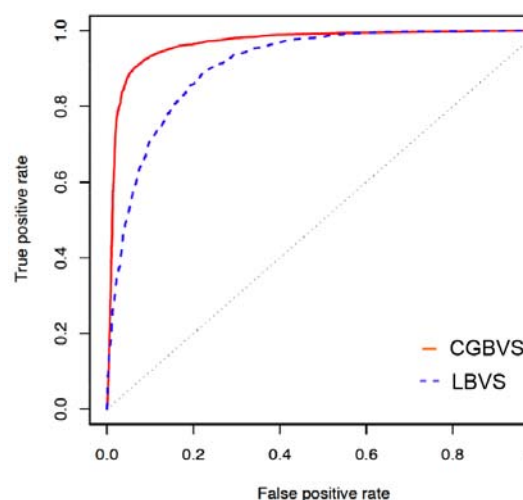


図 4-1. ROC 曲線による相互作用予測性能の比較。赤い実線が CGBVS、青い破線が最近傍法を用いた LBVS を示す。

2. ヒト $\beta 2$ アドレナリン受容体 ($\beta 2AR$) リガンド予測への適用

当社が収集した化合物-GPCR 相互作用情報は、今までの研究で「強く結合する」と知られている情報のみであり、その他の大部分の化合物-GPCR ペア間の相互作用の有無は不明である。そこで、CGBVS が期待に反して相互作用すると予測したペアは本当に相互作用しないのか、という疑問解決のための *in vitro* での実験により、予測された相互作用スコアと実際の相互作用の有無との関連性を確認した。

まず、ヒト $\beta 2$ アドレナリン受容体 ($\beta 2AR$) を標的 GPCR とし、構築した学習モデルを用いてリガンド予測を行った。リガンド予測の対象化合物は、他 GPCR のリガンドとして知られている 826 化合物とした。そして、CGBVS により予測された $\beta 2AR$ リガンド候補上位 50 の化合物のうち、文献・特許調査により 14 種の化合物について $\beta 2AR$ との相互作用に関する報告を確認した (図 4-2 左上)。さらに、残りの相互作用不明な化合物のうち、入手可能な 21 種類について *in vitro* 結合阻害実験を行ったところ、17 種類の化合物が相互作用 ($10^{-5}M < IC_{50} < 10^{-3}M$) を示した (図 4-2 右上)。結合阻害実験のヒット率は 81% (17/21) にのぼり、ここにおいても高い予測的中率が示された。

一方、下位 50 化合物については、 $\beta 2AR$ リガンドとして報告されているものは文献およ

び特許調査では確認されなかった (図 4-2 左下)。さらに、残りのうち入手可能な 9 化合物についても結合阻害実験を行ったが、2 個の化合物が同程度の強さの相互作用を示したものの、残りの 7 化合物は相互作用を示さなかった (図 4-2 右下)。

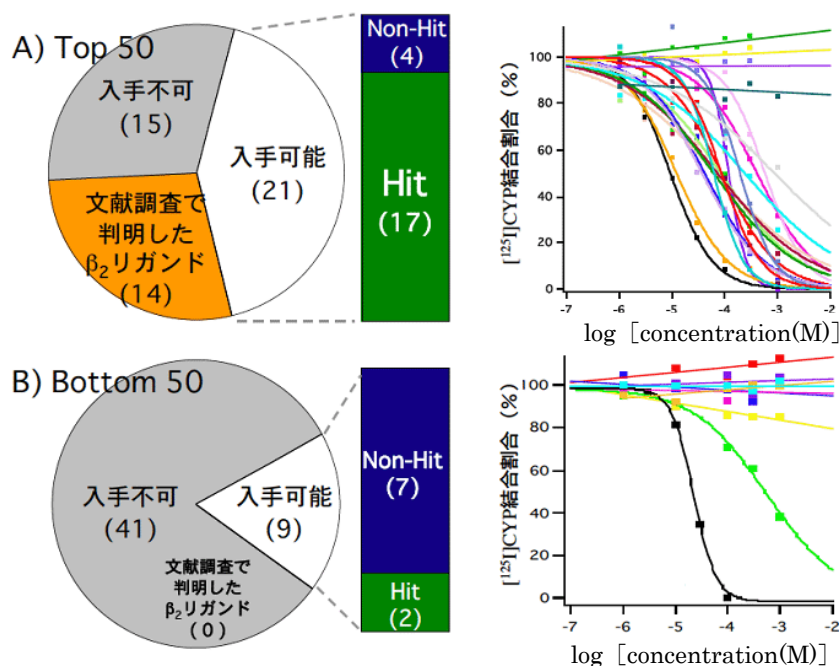


図 4-2. β 2AR リガンド予測結果の検証

予測スコア上位 50 (A) と下位 50 (B) の化合物に対する調査・実験結果を示している。

左: 文献調査および実験検証で判明した化合物の内訳

右: $[^{125}I]$ -シアノピンドロールに対する結合阻害曲線。縦軸が阻害の割合、横軸が各化合物の濃度を示す。

これらの予測結果を LBVS によるものと比較したのが下図 (図 4-3) である。結合阻害実験で相互作用を確認した化合物の半数近くは、LBVS ではスコアが低かった。実際に、これらの化合物は、典型的な β 2AR 作動薬の構造 (カテコラミン骨格、イソプレナリン誘導体) および β 2AR 拮抗薬の構造 (アリルアルキルアミン誘導体) とは異なる多様な骨格を持っており、化合物の構造類似性に基づく従来の方法では発見が困難なリガンド群である。

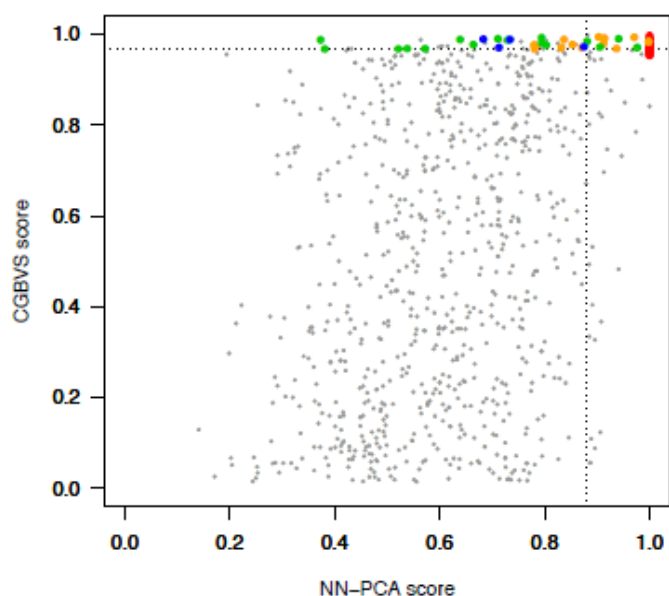


図 4-3. CGBVS と LBVS による β 2AR リガンド予測結果の比較

各点が化合物であり、縦軸は CGBVS、横軸は LBVS (主成分空間における最近傍法) による相互作用予測スコアを示している。文献調査および結合阻害実験の結果を以下の色分けにて表示した。

- 既知 β 2AR リガンド
- 文献調査 β 2AR リガンド
- 相互作用あり (結合阻害実験)
- 相互作用なし (結合阻害実験)

なお、図中の点線は、各手法において上位 50 位のスコアを示す。

また、これらの化合物の中には、ニューロペプチド受容体アンタゴニストなど、従来はペプチド受容体に作用する化合物として知られていたものも含まれており、遠縁にあたる β 2AR とも相互作用することが今回の結合阻害実験により確認された。以上から、ケミカルゲノミクス情報が、リガンド予測精度の向上のみならず、新規骨格を持つリガンドの検出にも有用であることが示唆された。

3. GPCR カーネルの改良

さらなる相互作用予測性能の改善を図るため、GPCR カーネルの改良を行った。上記モデルでは、GPCR のアミノ酸配列全長に存在する部分文字列を数え上げることにより、GPCR のベクトル化を行っていた。しかし、他のベクトル化手法も CGBVS に適用可能である。そこで、種々のベクトル化手法を CGBVS 法に適用することで、予測性能の向上を目指した。

方法として、まず、データベースに登録されているヒト GPCR のアミノ酸配列を以下の手法で数値ベクトルに変換した：[手法 1]部分文字列を数える、[手法 2]アミノ酸モチーフ配列の有無を数える、[手法 3]多重アライメントで並んだアミノ酸を数値に変換する。次に、各 GPCR ベクトルを用い、それぞれ相互作用学習モデルを構築した。学習モデルの予測性能評価には、交差検証法を用いた。

結果として、いずれのベクトル化手法を用いても、化合物類似性に基づく従来のリガンドスクリーニング手法よりも高い予測性能（正確度で 5-10%の上昇）を示した（図 4-4）。予測的中率（正確度）としては、[手法 3]が最も高く、88.8%であった。また、ベクトル化手法によって予測結果が異なったことから、これらの手法を相補的に組み合わせることで、さらに高性能なモデルを構築できる可能性が示唆された。

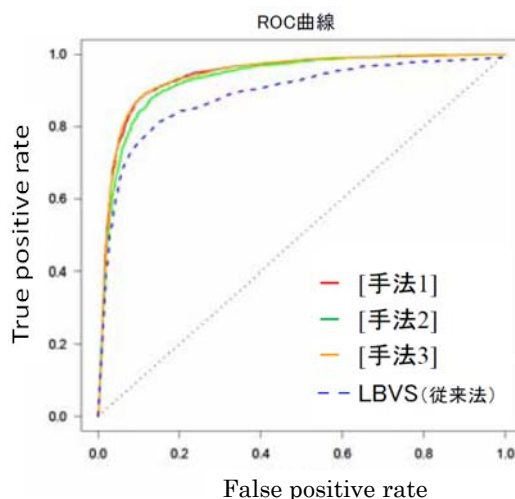


図 4-4. 手法 1~3 と LBVS との比較

4-1-2 大規模相互作用データを用いた学習モデルの構築

以上の結果により、新規リガンド予測における CGBVS の有用性を確認することができた。しかし、これまで学習モデル構築のために利用した相互作用情報は、収集したデータの一部（約 5000 件）にすぎない。そこで、次に、収集したデータを最大限に活用するため、100 万件もの相互作用情報を入力とした学習モデルの構築を試みた。

まず、学習モデルを構築する際、実時間で収束可能な計算にするため、数値ベクトルのバイナリ化、集団学習の適用、負例セットの削減など、構築方法の改良を行った。次に、実際に計算サーバにてこの大規模学習モデルを構築した。

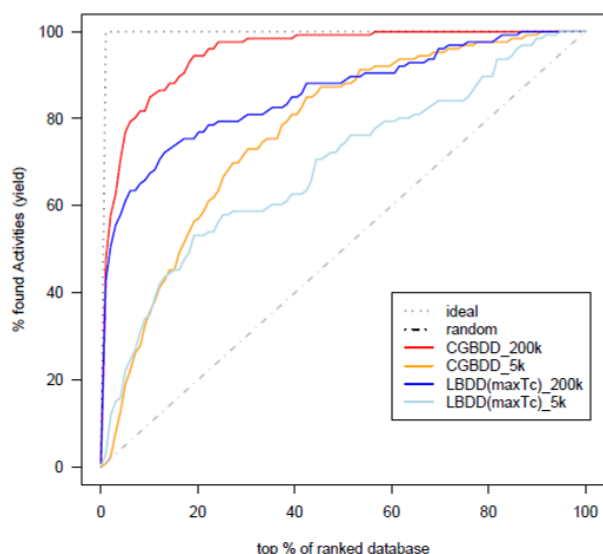


図 4-5. CGBVS（20 万相互作用学習モデル）の評価

結果、GPCR の 100 万相互作用情報から 20 万相互作用を抽出した学習モデルを作成し、次に示す各種法との比較評価を行った（左図 4-5）。20 万相互作用情報を用いた CGBVS（赤線）、LBVS（青線）、に加えて、当社設立当初の 5,000 相互作用による初期学習モデルによる CGBVS（オレンジ線）、LBVS（水色）との予測精度比較を行った。

結果、大規模な SVM 機械学習計算を可能とした 20 万相互作用学習モデルは、初期学習モデルを凌ぐ予測精度を示している。

つづいて、性能評価のため、糖尿病に関する 3 種類の GPCR のリガンド予測に適用した（予測対象は、いずれも未学習の化合物）。予測の結果、いずれも LBVS と同等以上のリガンド予測性能を示した。しかも、ここでも CGBVS は、LBVS では検出できないような新規骨格を持つリガンドを正しく予測している（図 4-6）。

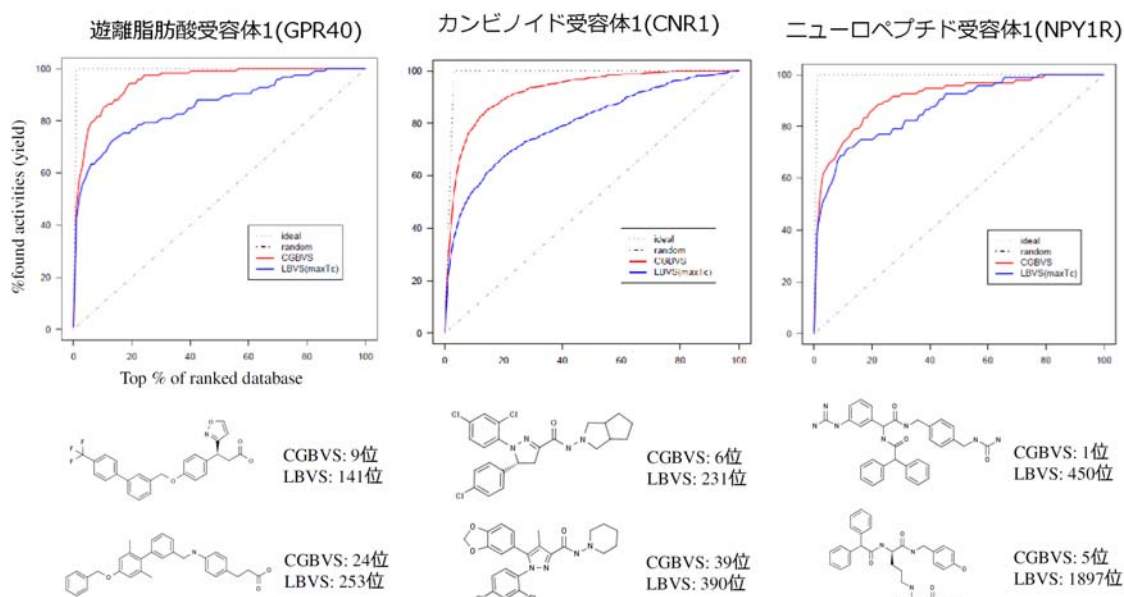


図 4-6. CGBVS（20 万相互作用学習モデル）と LBVS によるリガンド予測評価

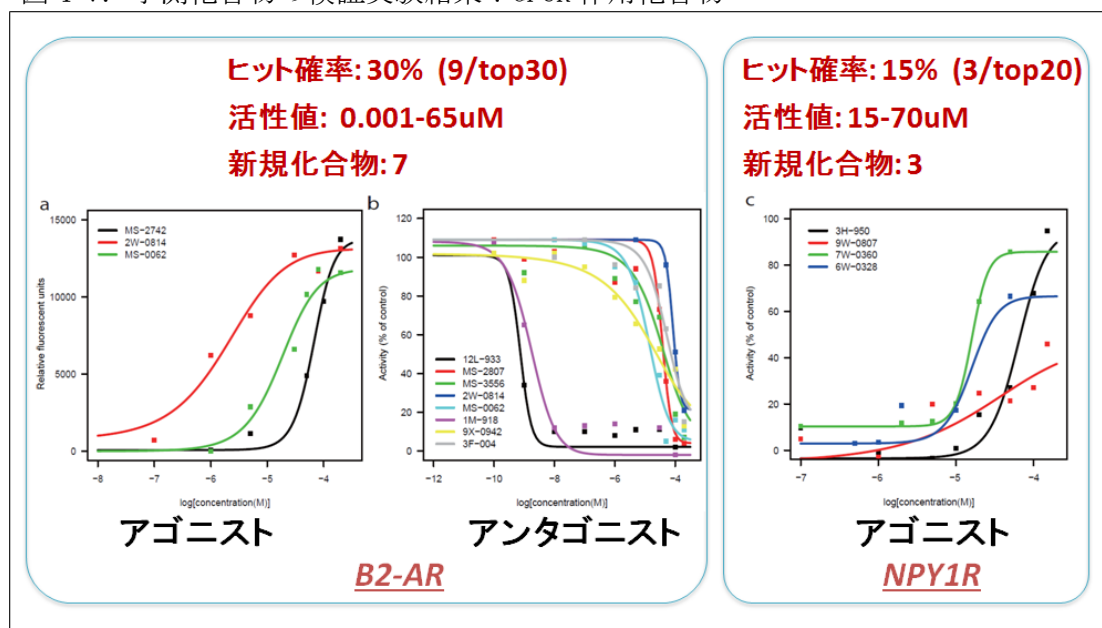
4-1-3 β 2AR 及び NPY1R に対する活性予測と実証実験

次に、4-1-2 で構築した大規模データを用いたモデル構築の方法論を用いて、相互作用マシニング法を GPCR の β 2 アドレナリン受容体 (β 2AR) を標的としたインシリコスクリーニングに適用し、活性予測と実証実験を行った。

まず、317 種類の GPCR のタンパク質配列とそれらに作用する 866 種の化合物の化学構造からなる 5207 通りの既知の相互作用ペアを計算機に機械学習させ、GPCR と化合物との既知相互作用に内在する統計的保存パターンを抽出し、それらに基づく予測モデルを構築した。そして、この予測モデルを用いて、Bionet 社化合物ライブラリーの 11,739 化合物から、 β 2AR と相互作用する可能性の高い化合物の絞込みを行った。11,739 化合物のうち、活性スコアに基づいて上位 30 個の化合物について、アゴニストおよびアンタゴニスト活性測定を行った。

その結果、図 4-7 左に示すように、上位 30 化合物中、9 個の化合物が活性を示し（ナノモルオーダーの非常に高活性な化合物を含む）、実に 30% という驚異的なヒット確率を示した。また、7 個の化合物は新規構造であり、新薬開発の特許戦略上、最も重要である新規化合物の探索にも本手法は成功している。

図 4-7. 予測化合物の検証実験結果：GPCR 作用化合物



また、上記の β 2AR 以外に、ペプチド系 GPCR の一つであるニューロペプチド Y 1 受容体 (NPY1R) を標的としたインシリコスクリーニング計算に適用し、Bionet 社化合物ライブラリーの 11,739 化合物の活性予測を行い、活性スコアに基づいて上位 20 個の化合物について In vitro 活性実験を行った。その結果、上位 20 化合物中、3 個の化合物が活性を示し (15% のヒット確率)、いずれも非ペプチド系の新規骨格を有する化合物であった (図 4-7 右)。

4-1-4 CGBVS と既存手法 (LBVS・SBVS) との比較・精度検証

1. ケミカルゲノミクス情報の収集・整備

CGBVS および LBVS (類似化合物探索法) による β 2AR リガンド予測を行うため、GLIDA データベースから、G タンパク質共役型受容体 (GPCR) とそのリガンドに関する相互作用の情報 5206 件を収集した。この情報には、ヒト β 2AR の相互作用 40 件が含まれている (図 4-8)。

図 4-8. ヒト β 2AR タンパク配列情報

```
DYKDDAMGQPGNGSAFLAPNRSHAPDHVDTQQRDEVVVVGMGIVMSLIVLAIVFGNVLVITAIKFERLQVTNYFITSLACADLVMGLAVVPFGAAHILMKM
WTFGNFWCFWTSIDLVCVTASIE TLCVIAVDYFAITSPFKYQSL LTKNKARV IILMVWIVSGLTSLFPIQMHWYRATHQEA INCYAEETCCDFFTNQAYAIAS
SIVSFYVPLVIMVFVYSRVFQEAQRQLNIFEMLRIDEGRLKIKYKTEGYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGIL
RNAKLKPVYDSLDAVRRALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITFRTGTWDAYKFLCKEHLKALTLGIIMGTFTLC
WLPFFIVNIHVHIQDNLIRKEYIILLNWI GYVNSGFNPLIYCRSPDFRIAFQELLCLRRSSLKAYGNGYSSNGNTGEQSG
```

2. 化合物、タンパク質の記述子計算

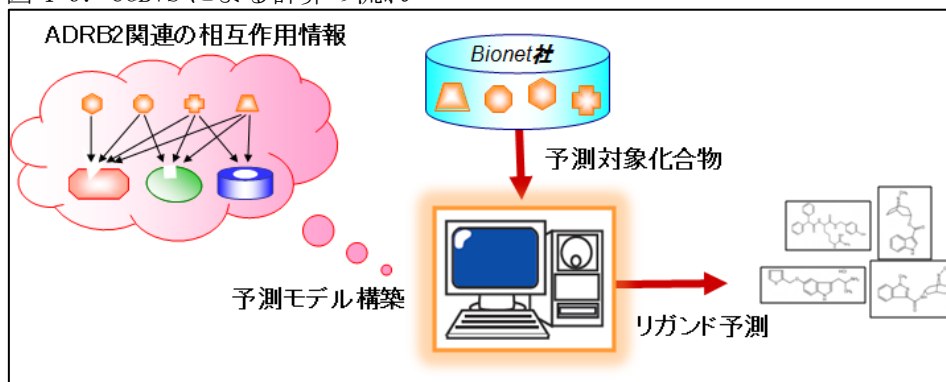
収集した相互作用をベクトルとして表現するために、各化合物の化学構造、各タンパク質のアミノ酸配列について、その属性 (記述子と呼ぶ) を計算した。なお、化合物記述子には、物性や構造的特徴を数値化する商用ソフト DragonX を使い、タンパク質記述子にはアミノ酸組成頻度を数え上げるプログラムを用いた。

また、Bionet 社の化合物ライブラリについて、同様にして記述子計算を行った。このとき、ライブラリ化合物間で構造類似性検索を行い、著しく化学構造の類似したものを取り除いた。その結果、11500 手類の化合物を予測対象として用いることにした。

3. CGBVS による β 2AR リガンド予測

代表的な学習アルゴリズムであるサポートベクターマシン (SVM) を用いて、化合物-タンパク質相互作用の有/無を判別した。具体的には、正例 (相互作用する化合物-タンパク質ペア) および負例 (相互作用しない化合物-タンパク質ペア) に対応する記述子をそれぞれ組み合わせて特徴ベクトルを構成し、SVM を用いて学習モデルを構築した。続いて、この学習モデルを用い、Bionet 化合物- β 2AR ペアに相当する新しいベクトルが相互作用有/無 のどちらのクラスに属するか予測を行った (図 4-9)。

図 4-9. CGBVS による計算の流れ



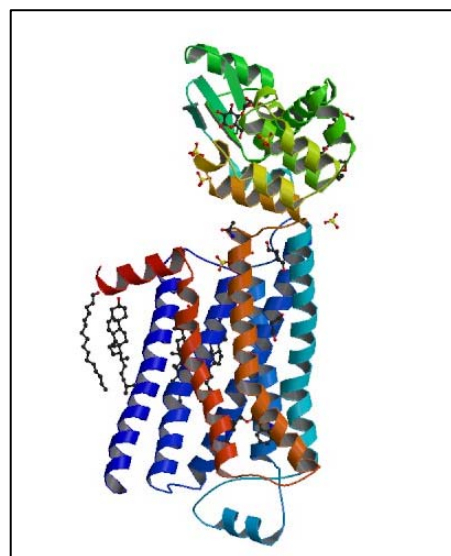
4. LBVS (類似化合物探索法) による β 2AR リガンド予測

化学記述子によって定義される空間にて近傍に位置する化合物は、類似する生物活性を示すことが知られている。我々は、化学記述子から計算される主成分空間上での既知リガンドからの距離を尺度として、予測する Bionet 化合物- β 2AR 間のスコアを算出することにした。具体的には、各予測化合物について、すべての ADRB2 既知リガンド 40 種類に対して、主成分ベクトルのピアソン相関係数を計算し、その最大値を LBVS スコアとした。なお、主成分軸には、累積寄与度が 80% に達するまでの 30 軸を用いた。

5. SBVS (ドッキング法) による β 2AR リガンド予測

近年同定された高解像度の β 2AR の結晶構造 (PDB 番号: 2RH1 : 図 4-10) を利用して、 β 2AR-各化合物間の結合時の自由エネルギーを計算した。まず、タンパク質構造の前処理として、ソフトウェア MOE を用い、重原子を固定した状態で水素原子の座標を熱力学的に安定な位置に配置した。そして、ドッキングシミュレーションソフト GOLD を用いて、結晶構造上でリガンド (carazolol) 結合部位にて各化合物の座標を繰り返し配置させ、フィット関数の小さいものから 10 個選んだ。なお、 β 2AR の残基 Asp113 および Asn312 については、先行研究によってリガンド- β 2AR 間に水素結合を形成すると判明しているため、ドッキングの拘束条件に追加して計算を行った。そして、これら 10 種類の配置についてスコア関数 Chemscore を計算し、その最小値を自由エネルギーとしてスコア化した。

図 4-10. β 2AR の結晶構造



6. β 2AR リガンド予測に対するカルシウムアッセイ実験

上記 CGBVS、LBVS、SBVS の予測スコア上位 20 化合物をそれぞれ購入し、in vitro で活性確認の実験を行った。リガンドで刺激された β 2AR は、Gq タンパク質を介して、細胞内

のカルシウム濃度を上昇させることが知られている。そこで、試験化合物の β 2AR 活性を確認するため、まず、CHO-K1 細胞から、 β 2AR を強制的に発現した安定細胞を作製した。そして、FLIPR Calcium Assay 4 Kit (Molecular Devices, Sunnyvale, CA)を用いて、各化合物投与による蛍光(Ex 485 nm, Ex 525 nm)の変化を測定し、ADRB2 活性を確認した(使用機器:図 4-11)。 β 2AR アンタゴニスト効果の測定時には、あらかじめ、既知 ADRB2 アゴニストであるイソプロテレノール(50nM)で刺激し、10分後に試験化合物を投与してその阻害効果を確認した。これらの実験を同一条件下で4回測定し、統計的に有意($P < 0.05$ 、対応のないT検定、 $n = 4$)なものについては、活性の強弱を確認するため、複数の濃度点(1.0-200 μ M)で測定し、50%効果(阻害)濃度を求めた。

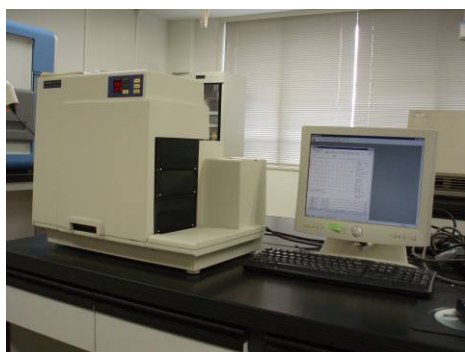


図 4-11. 測定装置「FLEX station」

用途：細胞内のカルシウム濃度を蛍光強度として測定する

7. 実証実験の結果

スコア上位化合物について、カルシウムアッセイにて各々の手法で ADRB2 活性が確認できたものを表 4-1 に示す。表中、太文字が各手法にて top20 に入った化合物を表しており、グレー色付けしたセルは参考値までに各手法でのスコア順位を記載している。

表 4-1 スコア順位と β 2AR 活性

ID	CGBVS 順位	LBVS 順位	SBVS 順位	EC ₅₀ (μ M)	IC ₅₀ (μ M)
NS-00895987	1	309	707	-	0.0007
NS-00898737	14	1616	1096	-	0.0017
NS-00902139	4354	8163	26	< 1 μ M	30
NS-00902920	10	322	-	1.9	83
NS-00901956	1130	4468	18	-	3
NS-03595378	1706	20	-	-	3
NS-00916999	4565	9726	15	-	3
NS-00926838	7510	7613	12	-	3
NS-00925291	13	3945	5620	17	13
NS-00923997	26	2271	6840	-	13
NS-00909249	9241	1790	10	-	25
NS-00927450	6	3	5616	-	33
NS-00198632	4	238	2248	-	37
NS-00903394	27	124	-	-	48
NS-00106740	53	14	6044	-	60
NS-00926872	8	6825	1836	65	-
NS-03596918	1471	3886	19	100	-

CGBDD スコア上位 20 化合物のうち、ADRB2 活性が確認できたものは 7 個(うち 3 つはアゴニスト活性)であった。これらのうち、NS-0895987 および NS-00898737 は非常に強力なアンタゴニスト活性(IC₅₀=0.7nM, 1.7nM)を示した。また、他の化合物については、活性

自体は弱いものの、既知の ADRB2 リガンド骨格を持たないため、新規骨格化合物といえる。

一方、LBDD スコア上位 20 化合物のうち、ADRB2 活性が確認できたものは 3 個（アゴニスト活性なし）であった。

また、SBDD スコア上位 20 化合物のうち、ADRB2 活性が確認できたものは 6 個（うち 2 つはアゴニスト活性）であった。このうち一つは、強いアゴニスト活性 ($EC_{50} < 1 \mu M$) を示した。

8. 結果の考察及び実証実験の達成度

以上の結果より CGBVS は、化合物ライブラリーを対象としたリガンド予測計算においても、従来法を超えるリガンド発見能力を有することが示された。特に、低計算コストで新規骨格化合物の発見に繋がったことから、CGBVS は新薬開発におけるニーズにぴったり適合した手法といえる。特に注目すべきは、それぞれのスコア順位を比較したときに、その相関性が弱く、いずれの手法も互いに相補的な結果となっている点である。これは、CGBVS が、従来法では発見できなかったリガンドを射程圏内に捉えていることを意味している。すなわち、これら 3 種のスクリーニング手法の使い分けにより、より合理的な新薬開発が可能となると考えられる。

4-1-5 まとめ

これらの結果により、CGBVS は大規模スクリーニングに十分適用可能であることが確認され、リガンド予測精度の向上のみならず、新規骨格を持つリガンドの検出にも有用であることが確認できた。さらには、CGBVS と既存手法 (LBVS・SBVS) との比較・精度検証により、既存手法との関係性は相補的であり、組み合わせることでより精度の高い結果を得ることが出来る可能性を検証することが出来ている。

これらの結果、GPCR の予測モデルのモデル構築方法の検討、精度評価については、実施計画上の目標を到達している (μM オーダーで 10%以上)。

4-2 イオンチャネル・キナーゼ予測モデル評価検討

4-2-1 イオンチャネルを標的とした化合物探索モデルの構築及び評価検討

1. データの収集

イオンチャネルは、生体膜間にイオンを透過させる機能をもつタンパクの総称であり、GPCR と同様な膜貫通タンパク質の一種である。細胞の内外における各種イオンの濃度あるいは膜電位の維持、神経細胞など電気的興奮性細胞での活動電位の発生、シグナル伝達などに関与する。イオンチャネルの開閉の制御様式には幾つかあるが、主に電位依存性とリガンド依存性の2種類がある。電位依存性とは、膜電位の変化に応じてチャネルが開閉し、イオンの濃度勾配からなる駆動力により特定のイオンを選択的に透過させるものである。一方、リガンド依存性とは、受容体とイオンチャネルが共役している構造で、リガンドが受容体に結合することでチャネルが開いてイオンを透過させるものである。

電位依存性とリガンド依存性のイオンチャネルについて、公的データベース(Integrity)からリガンド情報の収集を行った。総リガンド数は15792化合物であった。

収集した化合物から学習モデルに使えるものを選び、それら化合物の相互作用情報を調べた結果、15213化合物であった(下図4-12参照)。これらの相互作用情報を用いて学習モデルの構築・評価検討を行った。

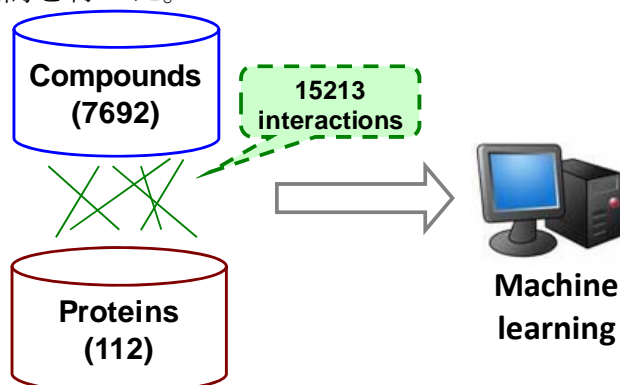


図4-12. イオンチャネルにおける相互作用情報数

2. 学習モデルの構築

調査したリガンド-タンパクの相互作用情報から相互作用マシンラーニング(SVM)モデルを構築する際に用いる化合物とタンパクのベクトル化手法を複数組み合わせで検討した。化合物のベクトル化に関しても複数の手法を用いて検討した。

3. 学習モデルの評価

化合物とタンパクのそれぞれのベクトル化手法を組み合わせで作成した学習モデルの性能評価を行った。評価は、学習に用いた化合物とは重複していない既知リガンドセットをそれぞれの学習モデルにて予測してROC曲線を描くことによって行った。

結果、化合物側のベクトル化手法はMACCSかCATSが良いことが分かった。タンパク側についてはmism30を使えば、MACCS、CATSのどちらにも共通して良い成績であることが分かった。MACCSとCATSはそれぞれ異なった視点から化合物をベクトル化する手法なので、場合によって使い分けて実際の化合物予測に用いることで良い化合物が見出せる可能性がある。

4-2-2 イオンチャネル予測モデルの構築と実証実験

1. イオンチャネルデータベースの評価

モデル構築の素材となるデータベースについて、商用データベースである AUREUS Pharma 社のデータベースと当社データベースを統合・整理し、データベースの評価を行った。

2. イオンチャネル予測モデルの構築と実験検証

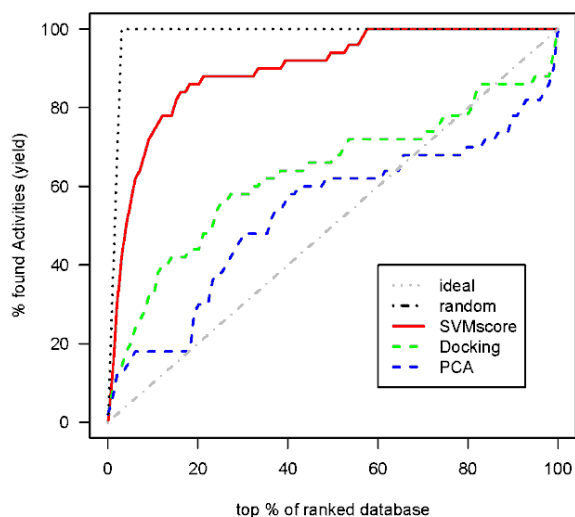
イオンチャネルリガンドデータベースからデータサンプリングを行い、相互作用マシンラーニング法により学習モデルを作成した。

さらに、構築した予測モデルのうち、化合物記述子に MOE2D を用いたものを用いて、計算による活性予測を行い、アッセイ実験による精度評価を行った。上位 50 化合物について活性評価を行ったところ、13 化合物のヒットが確認された (<5 μ M 以下)。ヒット率としてはおよそ 26% (13/50) を達成できている。

4-2-3 キナーゼを標的とした化合物探索モデルの構築及び評価検討

キナーゼは現在知られている薬物のうち、その 10% が作用することが報告されているタンパク質で、非常に重要な創薬ターゲットの一つである。ここでは、機械学習モデルの一つである SVM を用いて化合物とキナーゼの相互作用を予測するモデルを構築した。

次に学習データとは異なるデータセットを用いて、正しく予測ができるのかどうかを評価するために、外部データの収集を行った。今回は DUD(<http://dud.docking.org/r2/>)という



データベースより、3 種類のキナーゼ (CDK2, EGFR, FGFR1) のリガンド情報を収集した。また、本研究では、SVM の性能を比較するために、ドッキングシミュレーション法およびリガンドベースの主成分分析(PCA)も同時に行った。評価手法には **Enrichment Curve** を用いた。この評価方法は、グラフが上に位置するほど予測性能が良いことを表す。CDK2 の予測に関しては、ドッキングシミュレーション法が PCA よりも良い結果となったが、SVM を用いると、それら以上の予測精度が得られるという結果になった(左図 4-13)。

図 4-13. CDK における予測性能比較

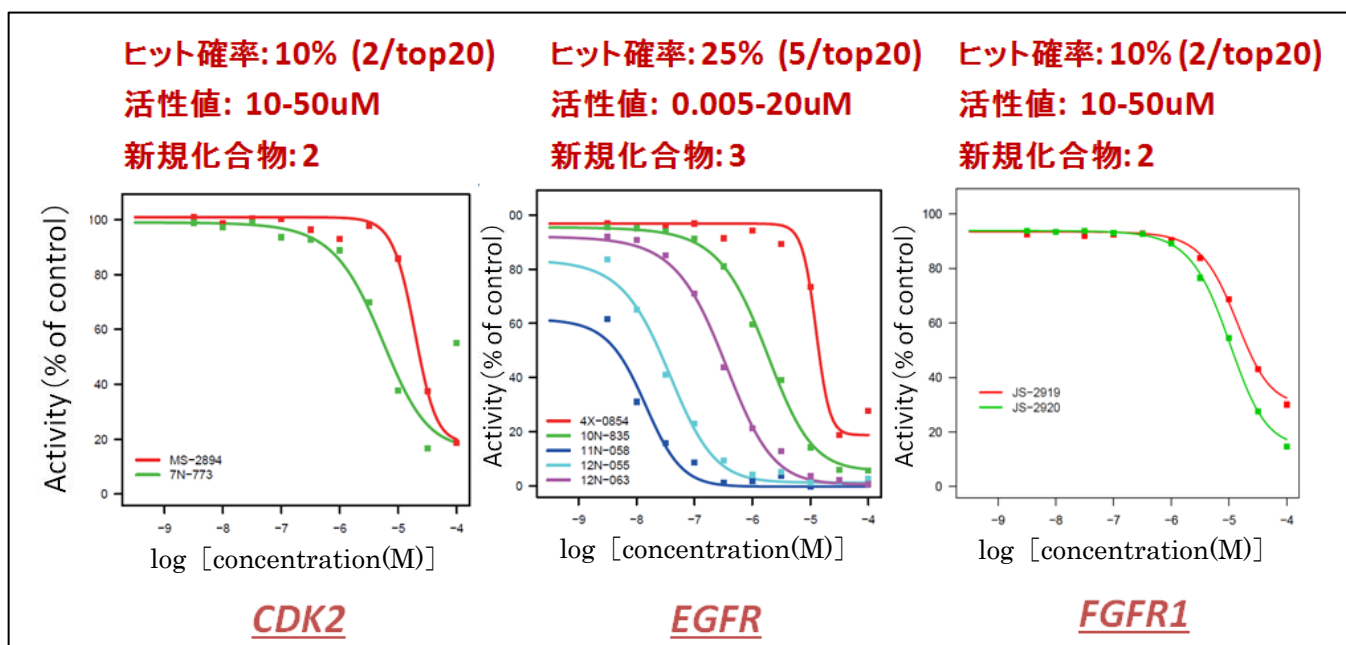
EGFR および FGFR1 の予測では、ドッキングシミュレーション法ではほとんど予測ができず、リガンドベースの PCA ではある程度予測できることが明らかとなったが、SVM ではそれ以上の非常に高い精度で予測をすることが可能であった。

4-2-4 キナーゼ予測モデルによる未知化合物の予測と実証実験

次に、キナーゼに対して相互作用するかどうか全く未知の化合物に対する予測を行った。予測をする化合物には Bionet 社から販売されている 11739 化合物を用いた。予測対象のキナーゼは、DUD での検証と同じ CDK2、EGFR、FGFR1 の三種類とした。本手法との比較のため、Ligand-based のモデル (主成分分析: PCA) およびドッキングシミュレーション法による予測も行った。これらの学習モデルにおける結果から、各キナーゼに対して予測スコアの高かった上位 20 化合物 (スコアトップ 20 化合物 x 3 methods x 3 kinases = 計 180 化合物) を選び、アッセイ実験を行った。

図 4-14 に各々のキナーゼに対するヒット率及び活性値、ヒット化合物のうち新規化合物であるものの数を示す。いずれにおいても、10%以上のヒット率と新規骨格の発見能力を示すことが確認された。また、ヒット化合物のうち、新規化合物についてもそれぞれ 2 ないしは 3 個が見出されている。

図 4-14. 予測化合物の検証実験結果：キナーゼ



4-2-5 キナーゼ予測モデルによる選択性予測についての実証実験

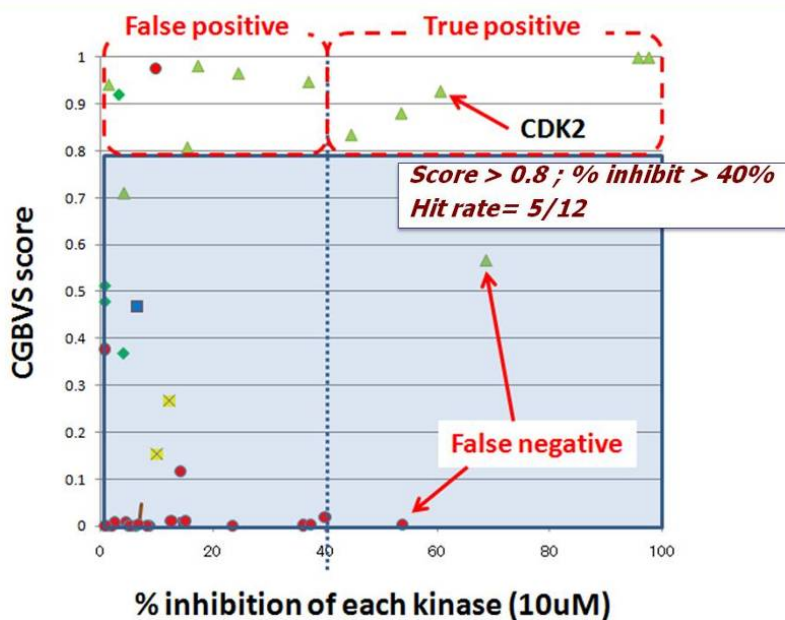
合理的投薬・副作用軽減を目指したときに、目的の標的タンパク質に作用する化合物を設計すること、すなわち「選択性」を制御することが、創薬プロセス上で重要な鍵となる。特に、キナーゼに関しては、タンパク質同士での結合部位の差別化が難しく、特定のキナーゼにのみ作用する阻害剤を設計することが、はなはだ難しいとされている。

そこで、相互作用マシンラーニング法がこの「選択性」予測にも適用可能か否かを明らかにするため、*in vitro*での検証評価を行った。まず、4-2-4で発見したCDK2 ヒット化合物「7N-773」について、同じ学習モデルを用いて、アッセイ可能な46種類のキナーゼに対する相互作用予測スコアを算出した。そして、実際にこれらのキナーゼに対して、7N-773による酵素活性阻害実験を行い、7N-773 10 μ Mにおける阻害率を測定し、予測結果との照合を行った。

得られた結果を図4-15に示す。まず、最も高い予測スコア(0.99超)を持つキナーゼ2種は、いずれもその阻害率が95%を超えており、また逆に、他のすべてのキナーゼはこの阻害率を大きく下回った。したがって、当モデルは高活性の標的キナーゼを精度良く予測できるといえる。一方、予測スコアの低いキナーゼは、おしなべて阻害活性が低いため、阻害効果の無いキナーゼに対しても、あらかじめ正しく予測できたといえる。たとえば、図4-15のようにスコア0.8/阻害率40%を閾値にしてクロス集計表を作成すると、予測スコアの高低と活性の有無の間に有意な関連性(Fisher 正確確率検定、 $P=0.0089$)が確認された。これらにより、我々の相互作用予測モデルは、一般的に困難とされるキナーゼ選択性の予測にも有効であることが示唆された。

Interaction prediction

between a CDK2 hit compound and 46 kinases



Family	Gene
AGC	AKT1
AGC	AKT2
AGC	AKT3
AGC	KPCB
AGC	KPCB
AGC	ROCK1
CAMK	CHIK1
CAMK	MAPK2
CKI	KC1E
CMGC	CDK2
CMGC	CDK2
CMGC	CDK4
CMGC	MKB3
CMGC	MKB1
CMGC	GSK3A
CMGC	GSK3B
CMGC	MKB8
CMGC	MKB9
CMGC	MK11
CMGC	MK12
Other	STK6
STE	MP2K1
STE	PAK2
TK	ABL1
TK	BTX
TK	EGFR
TK	FGFR1
TK	VGFR1
TK	FLT3
TK	VGFR3
TK	CSF1R
TK	ERBB4
TK	IGF1R
TK	INSR
TK	VGFR2
TK	KIT
TK	LCK
TK	LYN
TK	PGFRA
TK	PGFRB
TK	SRC
TK	KSYK
TK	TIE2
TK	ZAP70
TKL	RET
TG	RAF1

図右の表は46キナーゼをファミリーごとに分類し、色分けをしている。グラフにプロットされている▲や●などの色付きのポイントはキナーゼの各ファミリーに対応している。

図4-15. CDK2 ヒット化合物の46キナーゼに対する選択性予測の実験検証

4-2-6 まとめ

(1) イオンチャネル予測モデルの評価検討

電位依存性とリガンド依存性のイオンチャネルについて、データベースを整備の上、相互作用情報を用いて学習モデルの構築・評価を行った。予測モデル構築方法の評価検討から、化合物側のベクトル化手法はMACCS、または、CATSについて性能評価が高いことを見出し、タンパク側についてはmism30を使えば、MACCS、CATSのどちらにも共通して良い成績であることを見い出している。

MACCSとCATSはそれぞれ異なった視点から化合物をベクトル化する手法なので、場合によって使い分けて実際の化合物予測に用いることで良い化合物が見出せる可能性がある。

さらには、イオンチャネルについて、大規模データを用いた予測モデルの構築方法、具体的事例に対する活性予測と実証実験の結果、ヒット率として26% ($5\mu\text{M}$)の結果を得ており、開発目標であるCGBVS単独で10%以上のヒット率を達成している。

(2) キナーゼ予測モデルの評価検討

イオンチャネル同様、収集したおよそ330,000の相互作用情報を用いて行ったモデル構築・評価検証を行った。3つのターゲット (CDK2、EGFR、FGFR1) に対して、相互作用するかどうか全く未知の化合物に対する予測を行い、本手法との比較のため、Ligand-basedのモデル (主成分分析: PCA) およびドッキングシミュレーションによる予測も行い、予測スコアの高かった上位20化合物 に対して、アッセイ実験を実施。結果、SVMによる予測より導かれた候補化合物のうち、各々2~5個の化合物に強い阻害活性が見られた。結果として、目標である μM オーダーで10%以上の予測率を達成することが出来ており、さらには、キナーゼでは重要視される選択性に関しても、予測スコアの高低と阻害活性の有無との間の有意な関連性を示すことが出来ている。

4-3 (A) 連結モジュール・入力変換部の作成

4-3-1 詳細仕様設計 (20年度実施分)

連結モジュール・入力変換部は、以下の機能を設計した。

1. ユーザーインターフェース

ユーザーの使用しやすいグラフィカルなインターフェースを設け、学習モデルを作成するために必要なデータ・パラメータを受け取る (図 4-16)。

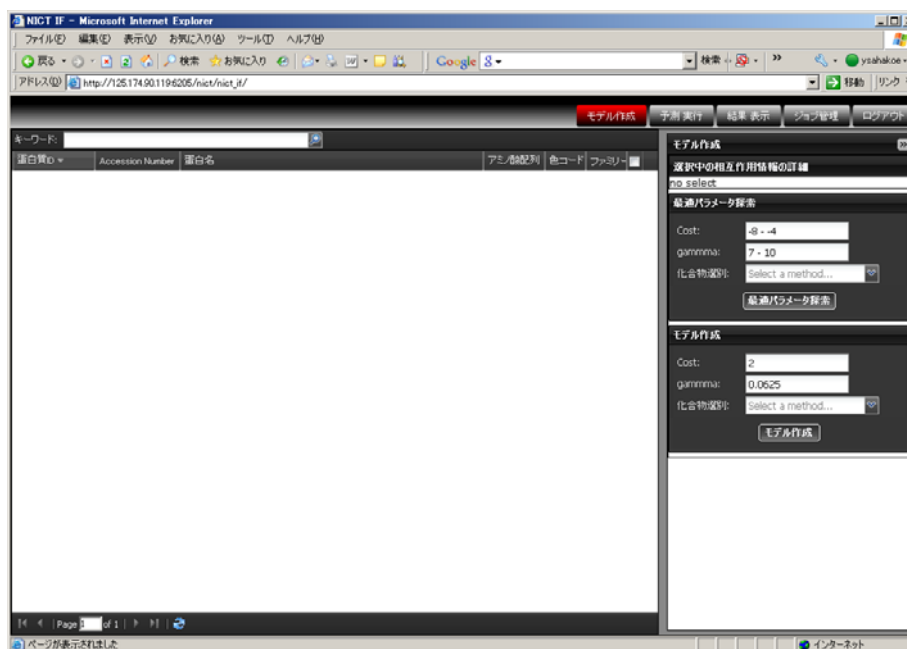


図 4-16. ユーザーインターフェース例 (学習モデル作成)

2. 学習モデル作成

予測に用いる学習モデルを作成する。

その際、1で選択されたデータ・パラメータを適切な形に変換する。

また、この処理は、一週間程度の長い処理時間を要する場合がある。よって進捗状況を確認できるインターフェースを設ける。

2-1. 学習モデルの登録

- ① システムのメインデータベースはタンパク質、相互作用情報、化合物の各情報を持ち、タンパク質検索することで相互作用情報を通じて化合物の情報を自動的にひき、学習用のデータセットを作成する。
- ② 検索インターフェースは通常の利用想定としてタンパク質を対象にした検索を行えるものとする。
- ③ 検索結果を複数選択できるインターフェースを用意し、選択毎に応じて対応する相互作用数等を確認できるようにする。
- ④ Cost や Gamma、ディスクリプタ等のパラメータを入力し、選択されたデータと共に SVM の学習データを作成(もしくは最適パラメータ探索)を開始し、ジョブとして登録する。

2-2. 学習モデルの編集

- ① 学習モデル登録により生成された学習モデルを編集できる。
- ② 主には、データセットの編集(相互作用データの追加/削除)、パラメータの変更等を

Web インターフェース上で行える。

- ③ 最適パラメータ探索を行ったジョブに関してはその結果を参照できる。
- ④ 予測実行にて学習モデルの検証を行ったモデルに関してはその結果を参照できる。

2-3. ジョブ登録

① 処理終了予測の表示

カスタム SVM モジュールより、処理終了予測ログが出力された際、それを取得し、表示できる。

② ジョブの一時停止及びレジューム機能

非常に長い処理が発生するものに関しては、処理を停止する機能をもたせる。

カスタム SVM モジュールは規定のイテレーション毎に処理終了予測や途中経過をデータとしてアウトプットする仕様で設計されるので、それを利用した処理再開機能を提供する。

③ イベントログの表示

カスタム SVM モジュールから出力されるログや、インプットパラメータの情報等を表示する。

2-4. カスタム SVM

① SVM のカスタマイズ

現状使っている SVM を、専門研究者の研究・調査結果（4-1-1、4-1-2 参照）を取り入れながら、パラメータや処理速度の最適化を行う。

libSVM、SVMlight を元とする。

② 処理終了予測

イテレーションの回数と現状までの処理時間を元に、処理終了予測を出力する。

③ ジョブの一時停止及びレジューム機能

処理の進捗状況をデータとして出力し、ジョブの一時停止及びレジュームを可能とする。

④ イベントログ

処理結果等を出力する。

4-3-2 詳細仕様設計（21 年度実施分）

連結モジュール・入力変換部は、以下の機能を持つ。

1. ユーザーインターフェース

ユーザーの使用しやすいグラフィカルなインターフェースを設け、学習モデルを作成するために必要なデータ・パラメータを受け取る。



図 4-17. ユーザーインターフェース例 (学習モデル作成)

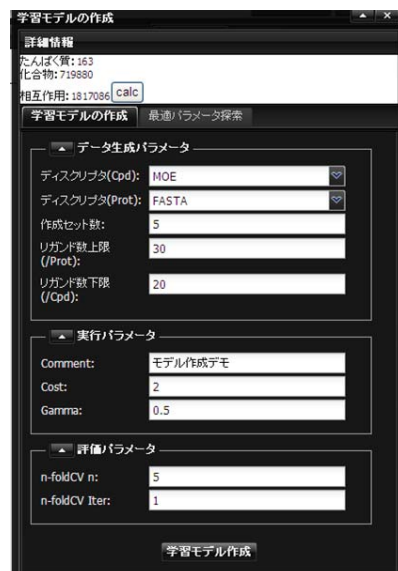


図 4-18. パラメータ入力インターフェース (学習モデル作成)

2. 学習モデル作成

予測に用いる学習モデルを作成する。

その際、1で選択されたデータ・パラメータを適切な形に変換する。

また、この処理は、一週間程度の長い処理時間を要する場合がある。よって進捗状況を確認できるインターフェースを設ける。

2-1. 学習モデルの登録

- ① システムのメインデータベースはタンパク質、相互作用情報、化合物の各情報を持ち、タンパク質検索することで相互作用情報を通じて化合物の情報を自動的にひき、学習用のデータセットを作成する。
- ② 検索インターフェースは通常の利用想定としてタンパク質を対象にした検索を行えるものとする。
- ③ 選択条件に応じて対応する化合物数、相互作用数等を確認できるようにする。
- ④ Cost や Gamma、ディスクリプタ等のパラメータを入力し、選択されたデータと共に SVM の学習データを作成し、ジョブとして登録する。

パラメーター一覧

項目名	初期値	内容
データ生成パラメータ		
ディスクリプタ (Cpd)	MOE	化合物の SVM 用ベクトル作成方法を選択。 MOE、Dragon、MPE で選択一つ選択。
ディスクリプタ (Prot)	FASTA	タンパク質の SVM 用ベクトル作成方法を選択。 現状、選択肢は FASTA の一つ。
作成セット数	5	負例を自動作成するセット数。
リガンド数上限 (/Prot)	30	1 タンパク質あたりの相互作用数の上限。
タンパク質数 (/Cpd)	20	1 化合物あたりの相互作用数の下限。
実行パラメータ		
Comment	モデル作成	学習モデルを認識する名前をつける。
Cost	2	SVM のパラメータ Cost 値。
Gamma	0.5	SVM のパラメータ Gamma 値。
評価パラメータ		
n-foldCV n	5	CrossValidation (モデル性能検証) の検証方法のパラメータ。
n-foldCV Iter	1	CrossValidation の繰り返し回数。

2-2. ジョブ管理

プロセスID	処理名	登録日	ステータス
2630	予測	2010/03/30 18:21:59	完了
21263	予測	2010/03/30 17:08:37	完了
21263	モデル作成	2010/03/30 13:15:47	完了
21266	予測	2010/03/29 23:01:09	完了
21267	モデル作成	2010/03/29 21:50:07	完了
2633	予測	2010/03/29 21:48:01	完了
21268	モデル作成	2010/03/29 21:41:39	完了
21266	モデル作成	2010/03/29 16:37:57	完了
21267	モデル作成	2010/03/29 16:25:51	完了
21262	予測	2010/03/29 11:07:00	完了
21265	予測	2010/03/29 11:05:33	エラー
2901	モデル作成	2010/03/29 10:56:07	完了
6309	モデル作成	2010/03/25 14:13:55	エラー
16708	予測	2010/03/14 13:40:31	完了
3175	モデル作成	2010/03/12 20:36:17	エラー
3175	最適パラメータ探索	2010/03/12 20:36:16	計算中
3181	モデル作成	2010/03/12 20:36:03	エラー
3181	モデル作成	2010/03/12 20:36:02	計算中
3177	予測	2010/03/12 16:14:08	完了
3162	予測	2010/03/11 16:48:41	完了
3179	予測	2010/03/11 16:38:00	計算中

図 4-19. ジョブ管理インターフェース

① ジョブ管理

処理名、更新日、ステータス (完了・エラー・計算中) が表示される。

② イベントログの表示

ジョブを選択すると、処理のログが表示される。

4-3-3 詳細仕様設計 (22年度実施分)

連結モジュール・入力変換部は、以下の機能を持つ。

1. ユーザーインターフェース

プロトタイプ版の評価を踏まえ、より使い勝手のよいインターフェースを設計した。また、大規模データを処理するための仕組みを走らせるために、必要なデータ・パラメータの受け渡しを追加した (図 4-20)。

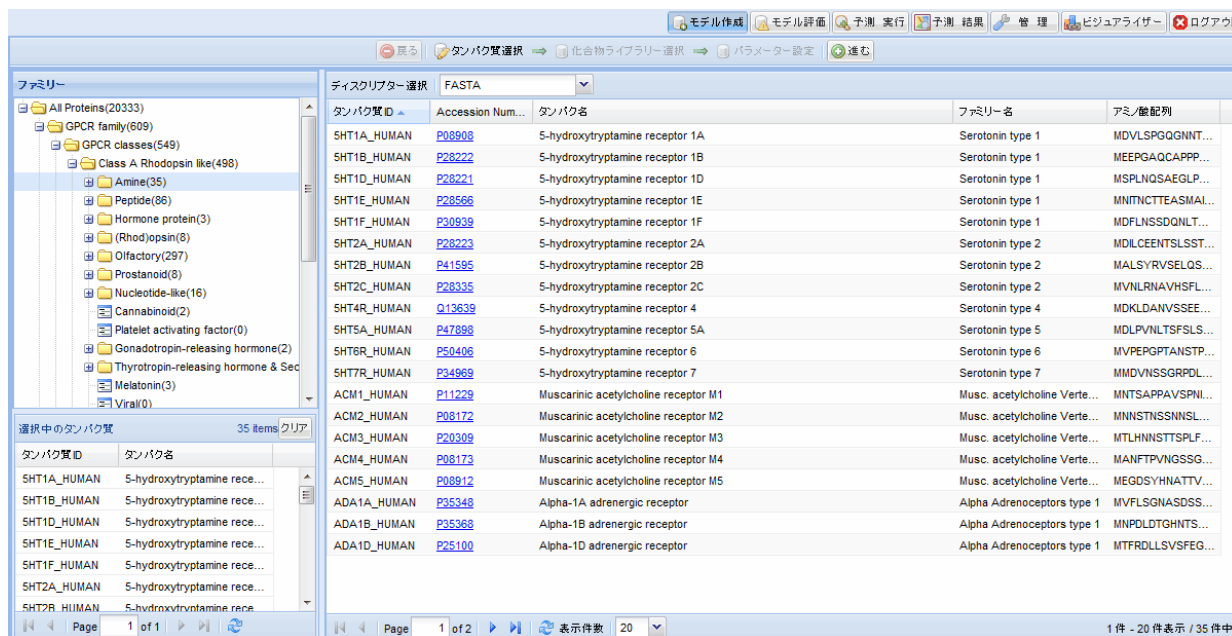


図 4-20. ユーザーインターフェース例 (学習モデル作成画面)

2. 学習モデル作成

予測に用いる学習モデルを作成する。

その際、1で選択されたデータ・パラメータを適切な形に変換する。

また、この処理は、一週間程度の長い処理時間を要する場合がある。よって進捗状況を確認できるインターフェースを設ける。

プロトタイプ版では、タンパク質-化合物の相互作用情報をデータベースとして持っているため、片方を選択することでもう一方を検索する仕組みにしていたが、大量データで使用すると速度が大変遅くなった。その使用感を改善するため、今回はタンパク質と化合物のインターフェースを別々に設けることにした。

2-1. 学習モデルの登録

- ① 学習モデルに使用するタンパク質を選択する。タンパク質は、ツリーから検索する方法と、一覧リストのフィルタリング機能で絞り込む方法がある。またこの際、どの記述子を使用するか選択する。
- ② 学習モデルに使用する化合物ライブラリーを選択する。また、記述子を選択し、その記述子のプロパティ値によって、ライブラリー内の化合物を絞り込むことができる。
- ③ パラメーターを入力し、ジョブとして登録する。

項目名	初期値	内容
データ生成パラメーター		
作成セット数	5	負例を自動作成するセット数。
1タンパク質あたりの相互作用数上限	30	1タンパク質あたりの相互作用数の上限。
1化合物あたりの相互作用数下限	20	1化合物あたりの相互作用数の下限。
実行パラメーター		
モデル名	モデル作成	学習モデルを認識する名前をつける。
Cost	10	SVMのパラメータ Cost 値。
Gamma	0.01	SVMのパラメータ Gamma 値。
評価パラメーター		
n-foldCV n	5	CrossValidation (モデル性能検証)の検証方法のパラメータ。
n-foldCV Iter	1	CrossValidationの繰り返し回数。
負例生成パラメーター		
負例を自動で生成する	チェックあり	負例を自動で生成しない(インポートした負例データを使用する)場合は、チェックを外す。

2-2. ジョブ管理

プロセスID	処理分類	コメント	実行日	ステータス
27735	化合物ライブラリー管理	化合物ライブラリー test2ファイル追加	2010/08/27 18:35:38	完了
26733	評価	モデル モデル作成 評価:PTHR1_HUMAN	2010/08/27 18:20:12	完了
15241	化合物ライブラリー管理	化合物ライブラリー namik2010ファイル追加	2010/08/26 18:47:48	停止
11531	予測	予測テスト	2010/08/26 11:55:55	完了
4640	評価	モデル モデル作成 評価:NK1R_HUMAN	2010/08/26 01:40:59	完了
4090	最適パラメーター探索	パラメーター探索	2010/08/26 01:16:13	完了
681	モデル作成	モデル作成	2010/08/25 19:29:13	完了
477	モデル作成	モデル作成	2010/08/25 19:17:21	エラー
31985	モデル作成	モデル作成	2010/08/25 18:54:59	完了
30889	モデル作成	モデル作成	2010/08/25 18:36:56	停止
10829	化合物ライブラリー管理	化合物ライブラリー yosokuファイル追加	2010/08/19 20:40:38	完了
10690	化合物ライブラリー管理	化合物ライブラリー test2ファイル追加	2010/08/19 20:38:04	完了
4459	化合物ライブラリー管理	化合物ライブラリー gpcrファイル追加	2010/08/19 12:15:09	停止
23662	予測	類似度テスト	2010/08/17 14:42:19	完了
20770	予測	test0817	2010/08/17 14:14:38	エラー
20523	化合物ライブラリー管理	化合物ライブラリー gpcrファイル追加	2010/08/17 13:50:09	停止
5668	化合物ライブラリー管理	化合物ライブラリー gpcrファイル追加	2010/08/16 14:40:56	完了
640	化合物ライブラリー管理	化合物ライブラリー gpcrファイル追加	2010/08/13 19:52:08	停止
23992	化合物ライブラリー管理	化合物ライブラリー kinaseファイル追加	2010/08/12 21:05:27	完了
23705	最適パラメーター探索	パラメーター探索	2010/08/12 20:29:03	停止

図 4-21. ジョブ管理画面

① ジョブ管理

プロセス ID、処理分類、処理コメント、実行日、ステータス（完了・エラー・計算中・停止など）が表示される。

② イベントログの表示

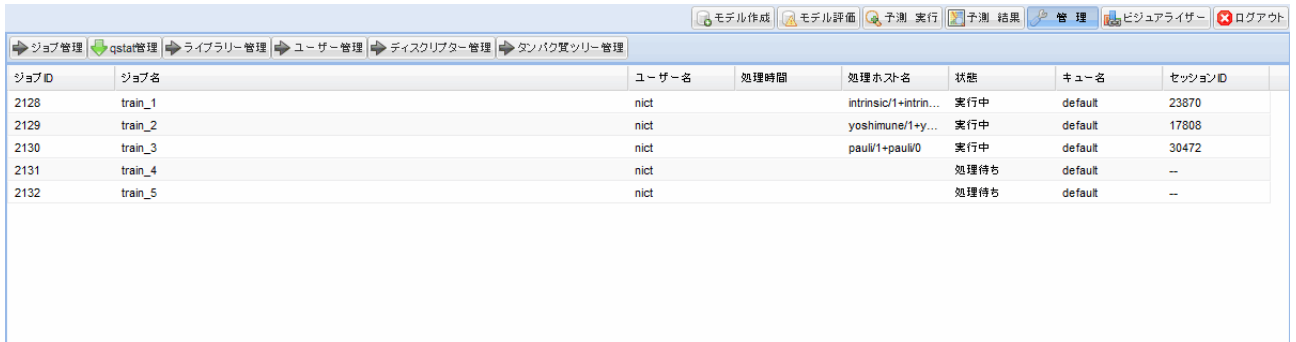
ジョブを右クリックし、「ログ閲覧」を選択すると、処理のログが表示される。処理はサーバーで走っていて、進行具合を直接確認できないので、このログから進捗を確認する。

③ ジョブ停止

ジョブを間違っって起動した場合や、他のジョブとの兼ね合いで停止したい場合は、ジョブを右クリックし、「ジョブ停止」を選択する。ステータスが「停止」に変わる。

2-3. qstat 管理

大規模データを扱うに際して、複数のクラスターを持ったサーバーのリソースを有効活用するため、Torque というジョブ管理システムを利用している。その進捗状況を確認するための機能を追加開発した。



ジョブID	ジョブ名	ユーザー名	処理時間	処理ホスト名	状態	キュー名	セッションID
2128	train_1	nict		intrinsic1+intrin...	実行中	default	23870
2129	train_2	nict		yoshimune1+y...	実行中	default	17808
2130	train_3	nict		pauli1+pauli0	実行中	default	30472
2131	train_4	nict			処理待ち	default	--
2132	train_5	nict			処理待ち	default	--

図 4-22. qstat 管理画面

4-3-4 まとめ

20, 21 年度にかけて基本的な機能の開発を経てプロト版を完成させ、実データによるテストを繰り返し、22 年度には大規模データへの対応、管理機能の追加等を行い、実践的なシステムの開発を完了することが出来ている。

4-4 (B) 相互作用マシンラーニング予測モジュールの作製

4-4-1 詳細仕様設計 (20 年度実施分)

相互作用マシンラーニング予測モジュールの作成は、以下の機能開発を設計した。

1. ユーザーインターフェース (図 4-23)

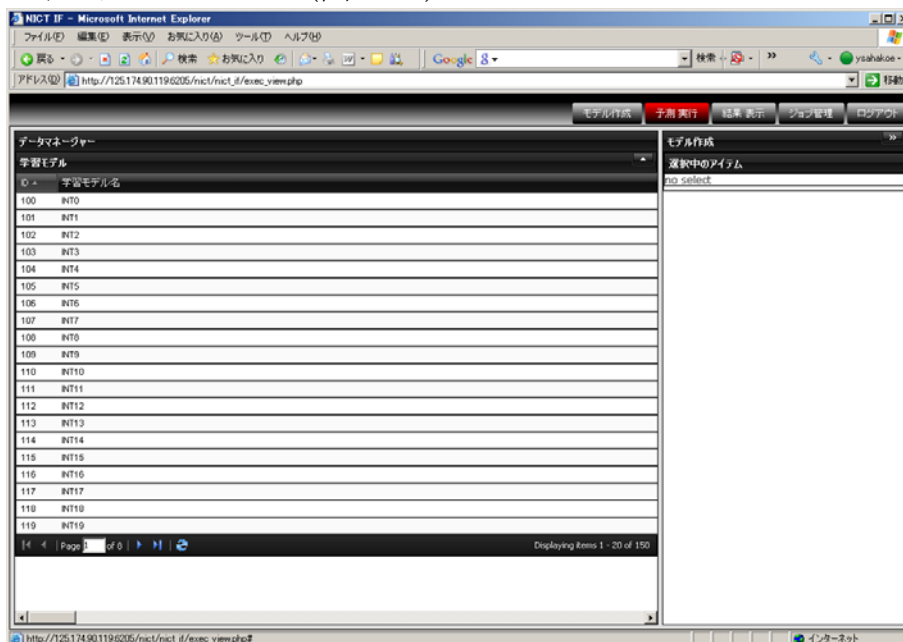


図 4-23. ユーザーインターフェース例 (予測実行)

2. 予測実行

1で選択されたデータ・パラメータを適切な形に変換する。

また、この処理は、一週間程度の長い処理時間を要する場合がある。よって進捗状況を確認できるインターフェースを設ける。

2-1. 予測

- ① 学習モデルを選択し、ユーザークエリとともに予測を行う。
- ② ユーザークエリはファイルアップロードによる化合物情報(SDF)か、化合物描画モジュールによる化合物情報を想定する。
- ③ あらかじめ評価用の化合物セットを用いることで、学習モデルの評価を行えるようにする。

2-2. ジョブ登録

- ① 処理終了予測の表示
カスタム SVM モジュールより、処理終了予測ログが出力された際、それを取得し、表示できる。
- ② ジョブの一時停止及びレジューム機能
非常に長い処理が発生するものに関しては、処理を停止する機能をもたせる。
カスタム SVM モジュールは規定のイテレーション毎に処理終了予測や途中経過をデータとしてアウトプットする仕様で設計されるので、それを利用した処理再開機能を提供する。
- ③ イベントログの表示
カスタム SVM モジュールから出力されるログや、インプットパラメータの情報等を表示する。

2-3. カスタム SVM

- ① SVM のカスタマイズ
現状使っている SVM を、専門研究者の研究・調査結果（4-1-1、4-1-2 参照）を取り入れながら、パラメータや処理速度の最適化を行う。
libSVM、SVMLight を元とする。
- ② 処理終了予測
イテレーションの回数と現状までの処理時間を元に、処理終了予測を出力する。
- ③ ジョブの一時停止及びレジューム機能
処理の進捗状況をデータとして出力し、ジョブの一時停止及びレジュームを可能とする。
- ④ イベントログ
処理結果等を出力する。

4-4-2 詳細仕様設計 (21年度実施分)

(A) で作成したモデルに対して、予測を実行するためのモジュールを作製する。相互作用マシンラーニング予測モジュールの作成は、以下の機能を持つ。

1. ユーザーインターフェース

ユニークID	学習モデル	実行ジョブ	スコア	ステータス	Cross Validation	Draw EnrichmentCurve
192	モデル作	モデル作成	53.14	完了	イテレータ番号	モデルセット番号
191	モデル作	モデル作成	51.84	完了	1	1
190	モデル作	モデル作成	0	完了	1	1
189	モデル作	モデル作成	0	完了	1	1
188	モデル作	モデル作成	0	完了	1	1
187	モデル作	モデル作成	0	完了	1	1
186	モデル作	モデル作成		エラー	1	1
182	モデル作	モデル作成		計算中	1	2
181	モデル作	モデル作成		計算中	1	2
180	パラメータ最適化探索		0	完了	1	2
171	ION CHA	モデル作成	89.15	完了	1	2
1	テストモ			エラー	1	2

図 4-24. ユーザーインターフェース例 (学習モデル評価)

The screenshot shows the '予測実行' (Prediction Execution) dialog box. It includes a 'Comment' field with the value '予測実行', a dropdown menu for '登録済みの化合物データを利用' (Use registered compound data), and a '選択' (Select) button. Below the dialog, there is a table of proteins with their IDs and accession numbers, and a '予測' (Prediction) button next to each row.

図 4-25. ユーザーインターフェース例 (予測実行)

2. 学習モデル評価

学習モデル作成で作成したモデルの性能を評価・確認する。

① 「Cross Validation」タブ

モデル作成計算の詳細を表示する。「イテレータ番号」、「モデルセット番号」、「グループ番号」、「正答率」を表示し、正答率の平均が、モデルの「スコア (性能)」となる。

② 「Draw EnrichmentCurve」タブ

「タンパク質」、「相互作用数」が表示され、「Draw」ボタンを押下すると、エンリッチメント曲線という、モデルの性能の指標の一つが表示される。

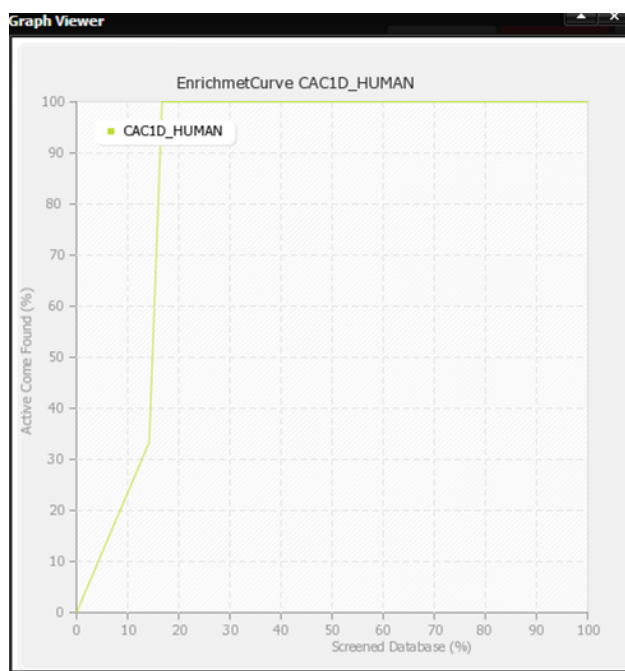


図 4-26. エンリッチメント曲線

3. 予測実行

1で選択されたデータ・パラメータを適切な形に変換する。

また、この処理は、一週間程度の長い処理時間を要する場合がある。よって進捗状況を確認できるインターフェースを設ける。

3-1. 予測

- ① 学習モデルを選択し、予測したいタンパク質を選択する。
- ② 予測に使用する化合物を、ライブラリー (SDF) か、化合物描画検索によって作成される化合物群か選択する。

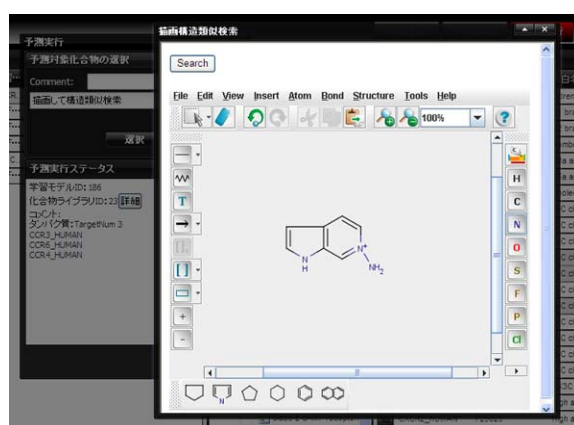


図 4-27. 化合物描画検索

3-2. ジョブ登録

① ジョブ管理

処理名、更新日、ステータス (完了・エラー・計算中) が表示される。

② イベントログの表示

ジョブを選択すると、処理のログが表示される。

4-4-3 詳細仕様設計 (22 年度実施分)

(A) で作成したモデルに対して、予測を実行するためのモジュールを作製する。相互作用マシンラーニング予測モジュールの作成は、以下の機能を持つ。

1. ユーザーインターフェース

プロトタイプ版の評価を踏まえ、より使い勝手のよいインターフェースを設計した。

大規模データを処理するために、必須の処理と任意の処理を分離し、また学習モデル作成のインターフェースと似た外観を採用することで、ユーザーが直観的に利用しやすい作りに改修した (図 4-28、図 4-29、図 4-30)。

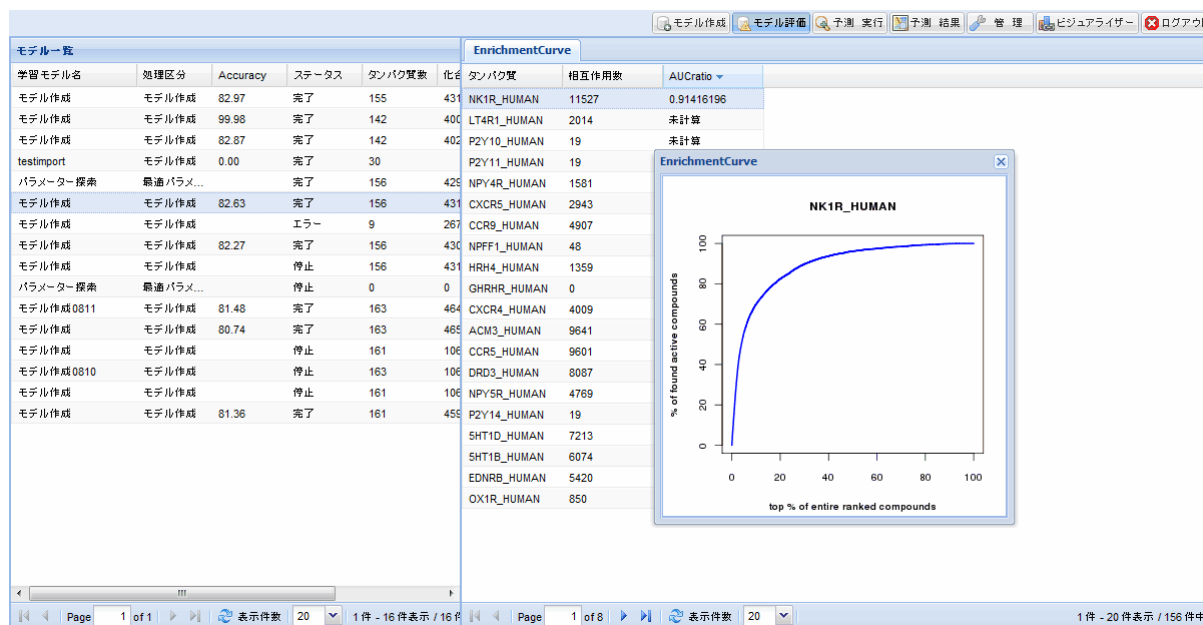


図 4-28. ユーザーインターフェース例 (学習モデル評価画面)

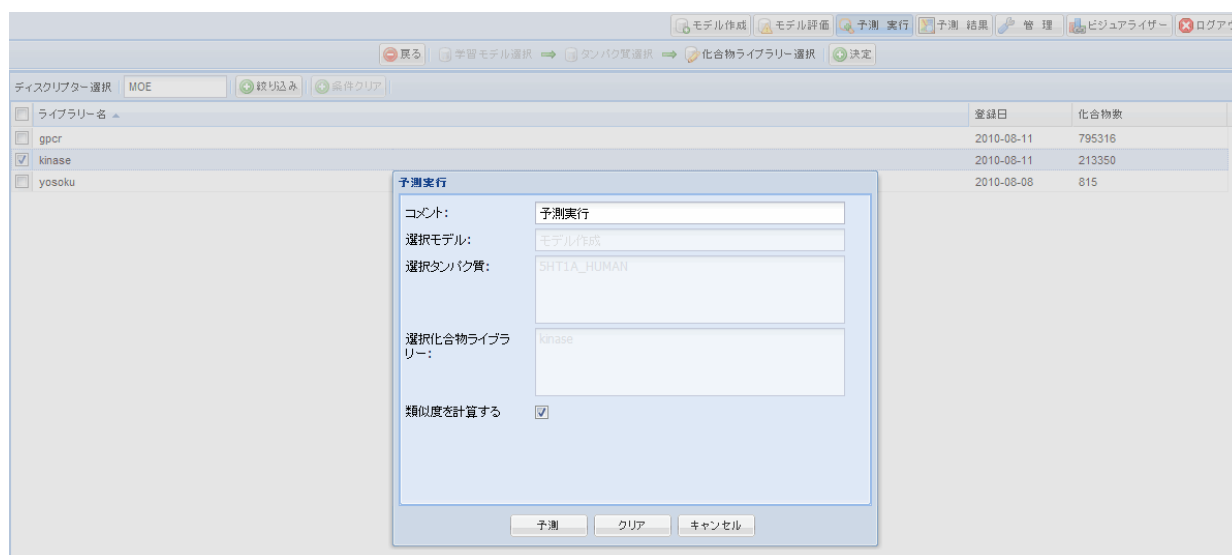


図 4-29. ユーザーインターフェース例 (予測実行画面)

コメント	登録日	モチ...	ライブ...	Pディ...	Cディ...	ステ...
予測テスト	2010-08-26	90	15	FASTA	MOE	完了
類似度テスト	2010-08-17	78	15	FASTA	MOE	完了
予測テスト0811_3	2010-08-11	63	15	FASTA	MOE	エラー
予測テスト0811_3	2010-08-11	63	15	FASTA	MOE	エラー
予測テスト0811_3	2010-08-11	63	15	FASTA	MOE	完了
予測テスト0811_2	2010-08-11	63	2	FASTA	MOE	停止
予測テスト0811	2010-08-11	63	2	FASTA	MOE	完了
予測テスト0810	2010-08-10	63	15	FASTA	MOE	完了

化合物	タンパク質	Score	Min	Max	Variance	Similarity	logP(...)	Weight	diam...
1	L00652	5HT1B_HUMAN	0.75493042	0.47229219	0.94343650	0.18906091	0.97468354	1.862	417.9...
2	L00075	5HT1D_HUMAN	0.65655713	0.29415171	0.95365133	0.28936257	0.62500000	4.007...	279.383
3	5572	5HT1A_HUMAN	0.45270236	0.15968904	0.75964547	0.21743465	0.86842105	5.008...	301.474
4	3206	5HT1D_HUMAN	0.14474742	0.01719136	0.21400879	0.08950967	0.73684211	2.24806	277.3...
5	154417	5HT1A_HUMAN	0.48435721	0.31771512	0.56842500	0.11197133	1.00000000	2.402	289.375
6	82148	5HT1B_HUMAN	0.14112421	0.05060385	0.22895411	0.08226196	0.68888889	2.632	243.306
7	1833	5HT1A_HUMAN	0.68699011	0.54906794	0.91729555	0.14887169	1.00000000	1.609	190.246
8	1238	5HT1D_HUMAN	0.90327888	0.73591029	0.98986370	0.09806916	1.00000000	4.380...	344.91
9	1970	5HT1B_HUMAN	0.06952685	0.03484419	0.14406393	0.04393755	0.72058824	-1.06...	399.4...
10	2689	5HT1B_HUMAN	0.89507439	0.79224999	0.97805195	0.07488491	1.00000000	2.977...	334.345
11	L00360	5HT1D_HUMAN	0.92001038	0.77573072	0.98168282	0.08347304	0.91891892	3.793	292.3...
12	L00335	5HT1A_HUMAN	0.26047892	0.14791065	0.45120011	0.11305050	0.69230769	6.068...	373.8...
13	L00228	5HT1B_HUMAN	0.28144418	0.11122672	0.69506906	0.23746569	0.78181818	6.344...	448.418
14	3623	5HT1A_HUMAN	0.38001844	0.29990708	0.46087040	0.06270095	0.92500000	2.578...	275.3...
15	15780	5HT1D_HUMAN	0.58096344	0.31286098	0.75737898	0.17757938	0.96153846	1.951	174.2...
16	33625	5HT1A_HUMAN	0.81947655	0.50103014	0.97943935	0.18410541	0.88636364	3.977	347.4...
17	23056	5HT1B_HUMAN	0.44881436	0.14694196	0.86033365	0.31853196	0.75000000	3.592...	337.4...
18	33630	5HT1B_HUMAN	0.15309285	0.02335922	0.42537553	0.15699348	0.93333333	8.252...	523.9...
19	163925	5HT1D_HUMAN	0.26259669	0.06283883	0.58194277	0.21277619	0.96000000	-1.054	419.5...
20	3066	5HT1A_HUMAN	0.34710729	0.20879175	0.44161640	0.08870986	1.00000000	2.872	583.6...
21	43103	5HT1A_HUMAN	0.75467219	0.35658136	0.89648072	0.22602361	0.83333333	3.805...	311.4...
22	43815	5HT1B_HUMAN	0.90042363	0.71022882	0.96279634	0.10720377	0.90384615	3.438...	329.371

図 4-30. ユーザーインターフェース例 (予測結果画面)

2. 学習モデル評価

学習モデル作成で作成したモデルの性能を評価・確認する。

プロトタイプ版では、モデル作成時に評価も実行してしまい、この機能では結果を閲覧するだけだったが、大量データを使用する場合、モデル評価に膨大な時間がかかってしまう。また、関係する全てのタンパク質に対する評価処理を一括して行うため、一部の評価結果だけを見たい場合は、無駄な計算時間がかかってしまう。

これらの評価結果を元に、今年度は、評価画面から、個別のタンパク質に対して評価処理を実行できるように変更した (図 4-31)。

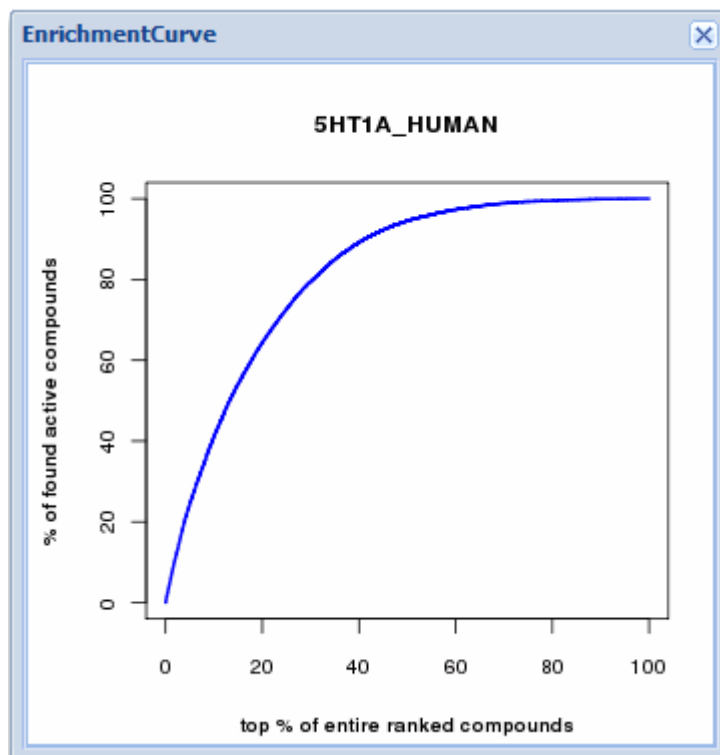


図 4-31. EnrichmentCurve

3. 予測実行

1で選択されたデータ・パラメーターを適切な形に変換する。

この処理も、学習モデル作成と同様、一週間程度の長い処理時間を要する場合がある。よってジョブとして登録し、ジョブ管理機能で進捗を確認できるようにする。

3-1. 予測

- ① 学習モデルを選択し、予測したいタンパク質を選択する。
- ② 予測に使用する化合物を、ライブラリーから選択する。

3-2. ジョブ登録

- ① ジョブ登録
処理名、更新日、ステータス（完了・エラー・計算中）が表示される。
- ② イベントログの表示
ジョブを選択すると、処理のログが表示される。

4. 予測結果

プロトタイプ版では、データ量が多い場合、一定条件で自動的にデータを切り捨て、予測結果を整形していたが、プロトタイプを評価した結果、ユーザー自らが自由な条件で予測結果を抽出したい、との要望が出たため、その機能を追加で開発した（図 4-32）。

	化合物	タンパク質	Score ▲	Min	Max	Variance	Similarity	logP(o/w)	Weight	radius
1	L00098	5HT1B_HUMAN	0.00189746	0.00065376	0.00364134	0.00156167	0.62500000			
2	160	5HT1B_HUMAN	0.00198063	0.00051462	0.00492141	0.00178717	0.65625000			
3	L00480	5HT1B_HUMAN	0.00198063	0.00051462	0.00492141	0.00178717	0.65625000			
4	L00477	5HT1B_HUMAN	0.00198776	0.00042576	0.00488057	0.00180765	0.65625000			
5	L00478	5HT1B_HUMAN	0.00214560	0.00037392	0.00437805	0.00173167	0.65625000			
6	L00479	5HT1B_HUMAN	0.00215757	0.00047072	0.00539178	0.00199778	0.65625000			
7	L00406	5HT1B_HUMAN	0.00251085	0.00064595	0.00479991	0.00193338	0.60465116			
8	L00194	5HT1B_HUMAN	0.00273371	0.00180314	0.00383112	0.00077628	0.74468085			
9	L00119	5HT1B_HUMAN	0.00277979	0.00055544	0.00452439	0.00171092	0.67741935			
10	5311223	5HT1B_HUMAN	0.00311810	0.00212686	0.00416861	0.00086743	0.67391304			
11	L00610	5HT1B_HUMAN	0.00312793	0.00150044	0.00509744	0.00141341	0.61224490			
12	L00346	5HT1B_HUMAN	0.00341195	0.00082009	0.00640370	0.00275764	0.59574468			
13	L00350	5HT1B_HUMAN	0.00351913	0.00108039	0.00551035	0.00175361	0.67796610			
14	L00621	5HT1B_HUMAN	0.00384227	0.00056731	0.01167654	0.00454575	0.66197183			
15	L00049	5HT1B_HUMAN	0.00400571	0.00170056	0.00736478	0.00249728	0.84000000			
16	L00053	5HT1B_HUMAN	0.00409259	0.00165270	0.01257857	0.00477062	0.58139535	2.142	386.48798	9
17	6434544	5HT1B_HUMAN	0.00433591	0.00142692	0.00813292	0.00328815	0.65306122	3.793	350.49899	9
18	L00305	5HT1B_HUMAN	0.00464439	0.00159109	0.01018217	0.00333388	0.64000000	4.6490002	402.87399	7
19	L00056	5HT1B_HUMAN	0.00517265	0.00200934	0.01025701	0.00326856	0.59375000	3.0050001	352.47098	11
20	60182	5HT1B_HUMAN	0.00521047	0.00127809	0.01096349	0.00367292	0.72222222	4.1030002	461.38599	7
21	L00185	5HT1B_HUMAN	0.00523821	0.00129949	0.01391979	0.00507502	0.58139535	2.339	378.509	9

図 4-32. 結果詳細画面

① フィルタリング機能

化合物・タンパク質の ID、スコア値、プロパティ値など、各種条件で絞り込みを実行できる。

② データエクスポート機能

1で絞り込んだデータをエクスポートして、統計を取ったり、グラフを描画したりすることができる。

4-4-4 まとめ

20, 21 年度にかけて基本的な機能の開発を経てプロト版を完成させ、実データによるテストを繰り返し、22 年度には大規模データへの対応、実使用を想定したインターフェイスの修正・結果抽出機能追加を行い、実践的なシステムの開発を完了することが出来ている。

4-5 (C) 連結モジュール・スコア統合部の作製

4-5-1 MOE 連結モジュール詳細設計 (21 年度実施分)

本モジュールは、相互作用マシンラーニングシステムと、タンパク質と化合物の結合部位を探索してドッキングシミュレーションを行う MOE システムの機能を連結させるモジュールである。

相互作用マシンラーニングシステムで予測スコアを計算するタンパク質と化合物に対して、MOE を用いたドッキングスコアも算出・加味することで、結合予測の精度向上に役立てる。

具体的には、タンパク質の ID・化合物の ID リスト・化合物の構造データ (標準の SDF 形式) を与えると、そのタンパク質と各化合物に対して一連のドッキングスコアを計算する。

相互作用マシンラーニングシステムは、その計算結果をインポートし、統合する。

1. 処理の流れ

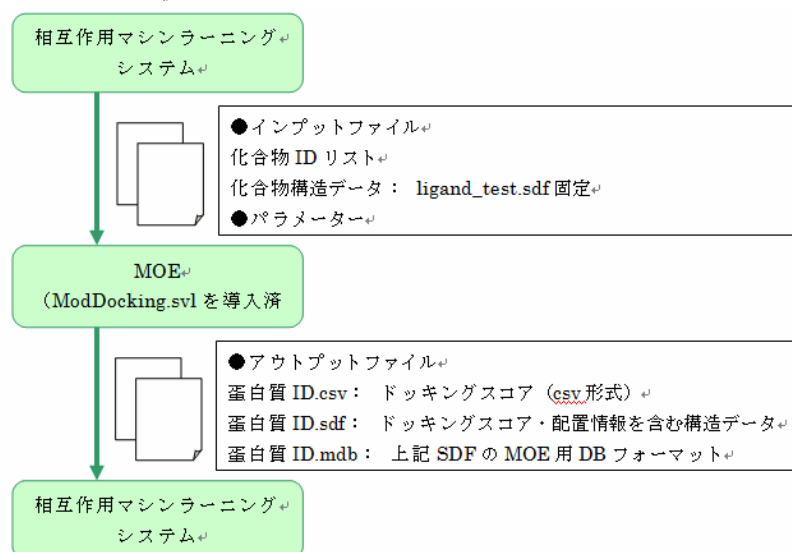


図 4-33. MOE 計算結果の処理の流れ

2. 事前準備

事前に MOE をインストールする。(インストール方法は MOE のマニュアルを参照)

モジュール本体 SVL ファイル (ModDocking.svl) を、MOE インストールルート /lib/svl/user.svl/ディレクトリにコピーして MOE を起動する。

また、デフォルトのタンパク質 DB を MOE システムに導入しておく。

3. 連結パラメータ

- ① ドッキングスコアを計算したいタンパク質の ID (PDBID) 一覧と、化合物 SDF ファイルを渡す。
- ② 配置計算メソッドを、「London dG」「ASE」「Affinity dG」「Alpha HB」から選択す

る。デフォルトは「London dG」となる。

4. ドッキング計算

- ① 化合物構造データ SDF を、MOE 用データベース形式 MDB に変換する
- ② ドッキングサイトの探索
タンパク質立体構造に化合物があてはまるサイトを指定する。
本来はマニュアル操作でサイトを探索するのが望ましいが、自動化する場合は、最大サイトのみを使用する。
- ③ ドッキング実行
3. 連結パラメータ②で選択した、配置計算メソッドを元に、ドッキングスコアを計算する。そのメソッドで最もスコアの高かった配置に決定し、SDF に構造情報を入力する。
- ④ スコアの再計算
3 で決まった配置に対し、全てのメソッドでスコアを再計算する。

4-5-2 大規模データへの対応 (22 年度実施分)

プロトタイプ版を評価した結果、今年度は、大量データの処理に適するよう、細部の調整を行った。

具体的には、大量データを一括で処理すると、MOE のシステム上、エラーが発生することが判明したため、複数に分割して計算を実行し、最終結果を統合する仕組みを追加開発した。

処理フローは、いくつかのパラメーターの調整を行い、ユーザーインターフェースに関しては、本モジュールは、予測処理の際に自動で実行するため、プロトタイプ版評価後の改修が必要なく、完了版とした。

4-5-3 まとめ

21 年度にプロト版を作成し、22 年度には実データを用いたテストを実施し、大規模データの処理が可能となるよう最終開発を行った。結果、実践的なシステムとして十分対応できるモジュールを完成させることが出来ている。

4-6 総合システムの評価

4-6-1 CGVS 及び SBVS による SRC キナーゼ阻害剤予測

CGBVS は、従来法とは異なる理論による新規手法である為、その予測結果もまた、新規たる傾向を示す。我々は、主要な創薬標的タンパク質ファミリー (GPCR、キナーゼ、イオンチャネル) のリガンドスクリーニングに於いて、CGBVS の優位性を示し、また実際に、上記のとおり、従来法 (LBVS、SBVS) では検出困難な活性化合物を発見してきた。ただし、その反面で、従来法でのみ検出可能な活性化合物が存在するのもまた事実である。したがって、現状の CGBVS は、従来法に取って代わる万能的手法ではなく、それらを別の観点から補完する第三のアプローチと位置づけるのが適切である。本研究開発の最終的な目標としては、相互作用マシンラーニング予測モジュールと、ドッキングシミュレーション法と組み合わせることにより、さらに予測精度を向上させる点にある (ヒット率 20%以上・ μ M オーダー)。以下に、本研究開発システムの総合的評価として、CGBVS と SBVS を組合せた予測計算を行うことにより、新規リガンドのスクリーニング効率を更に向上させる方法論を検討した。

CGBVS、SBVS のスコア結果から、その順位の和を総合スコア（以下、CGBVS+SBVS スコアと称す）として算出し、CGBVS 単独、SBVS 単独、CGBVS+SBVS スコアのそれぞれ上位 30 化合物を購入し、SRC キナーゼの活性阻害効果を *in vitro* アッセイで評価した。試験濃度 10 μ M で阻害活性が確認された化合物については、濃度を薄くしながら実験を繰り返し、50%阻害濃度（IC50）を算出した。

また、ヒット率の算出において、4 種類の IC50（10nM, 100nM, 1 μ M, 10 μ M）をヒットする基準として、それぞれ評価を行った（表 4-2）。

表 4-2. SRC 阻害剤リガンドスクリーニング結果

	CGBVS+SBVS	CGBVS	SBVS
IC50 < 10nM	50%	60%	10%
IC50 < 100nM	87%	77%	23%
IC50 < 1 μ M	100%	83%	53%
IC50 < 10 μ M	100%	87%	90%

結果として、低濃度領域（IC50 < 10nM）では CGBVS 単独、SBVS 単独、CGBVS+SBVS スコアのうち、CGBVS 単独のヒット率が最も高く、中・高濃度領域（IC50<100n~10 μ M）では CGBVS+SBVS スコアが最高値を示した。

4-6-2 まとめ

これらの結果から、CGBVS 単独により高精度な結果を得ることが出来るだけでなく、CGBVS と SBVS との組み合わせにより、中高濃度領域において、より広範で予測精度の高い結果を得ることが出来ることが確認された。すなわち、本研究開発によるシステムを用いて、相互作用マシンラーニングモジュールにより計算されるスコア（システムとしては A 及び B のモジュールにより導出）と、代表的な計算ソフトである MOE のスコア（C の連結モジュール・スコア統合部により導出）とを合わせることで、より高精度な結果（ヒット率 87% 以上・ μM オーダー）を得ることが確認されている。

また、本件実証実験では、評価を行ったライブラリがキナーゼ指向性ライブラリであったため、CGBVS 単独でも 60% 以上、CGBVS+SBVS では 87% 以上と言う非常に高い正解率を示していた。昨年までの我々の検討から、通常の化合物ライブラリを用いた CGBVS 単独のヒット率は平均して 15% であったことから、CGBVS+SBVS のヒット率は 21.8% 程度（ $=15\% \times 87\% / 60\%$ ）と換算できる。

4-7 総括

研究開発終了後の目標であった「①創薬シード・リード化合物探索システム(相互作用マシンラーニングモジュール)のパッケージ化」においては、20~21 年度にかけてプロトタイプ版のシステム開発を行い、22 年度には実データによるテストを通じて大規模データへの対応、管理機能の追加、ユーザーインターフェイスの改善など製品としての開発を行い、各モジュール(A 連結モジュール・入力変換部、B 相互作用マシンラーニング予測モジュール) を完成させることが出来ている。

また、主要な創薬ターゲットである GPCR、イオンチャネル、キナーゼのそれぞれに対して、予測モデルの構築・評価検証を行うことにより、相互作用マシンラーニングモジュールによる計算により、ヒット率 10% 以上 (μM オーダー) を達成することが出来ている。

「②外部ソフトとの連携のための連結モジュールの付加による統合システムの開発」については、21 年度にプロトタイプ版の開発を行い、22 年度には実データを用いたテストを実施し、大規模データの処理が可能となるよう最終開発を行い、連結化を完成している。

また、統合システムの評価として、CGVS 及び SBVS による SRC キナーゼ阻害剤予測を実施した。結果、CGVS 及び SBVS を統合させることでより高い予測精度を得ることが出来ることを確認できており、目標ヒット率 20% 以上 (μM オーダー) を達成することが出来ている。

以上より、当初研究開発目標についてはすべて達成することが出来ている。開発終了後は当システムの社内運用を実施し、追加開発を経て平成 23 年度中を目途に製薬企業及び大学等研究開発機関に向けて製品リリースをする予定である。

5 参考資料

5-1 研究発表・講演等一覧

(H20 年度)

論文

1. van der Horst, E., Okuno, Y., Bender, A., Ijzerman, A.P. "Substructure mining of GPCR ligands reveals activity-class specific functional groups in an unbiased manner" *J. Chem. Inf. Model.* **49**, 348-60, 2009
2. Okuno, Y. "In silico drug discovery based on the integration of bioinformatics and chemoinformatics" *YAKUGAKU ZASSHI* **128 (11)**, 1645-51, 2008

学術解説等

新島 聡、奥野 恭史「ケミカルゲノミクスに基づくインシリコ創薬」
日本薬理学雑誌 2009年3月1日発行 133,173(2009)

一般口頭発表

1. 関西バイオネットワーク「創薬バリューチェーンの構築に向けて」発表交流会
(財団法人京都高度技術研究所 産学連携事業部 連携支援グループ 主催)
「非結晶性標的タンパク質に対する化合物探索」(2008.12.8)
講演者：奥野 恭史
2. CBI2008 学会にてポスター発表
 1. Yabuuchi, H., Nijima, S., Okuno, Y., Evaluation of a chemical genomics-based virtual screening, CBI2008 年大会, P5-14, 東京, 2008.10
和題：ケミカルゲノミクスに基づく、バーチャルスクリーニング法の評価
3. 日本薬学会第 129 年会にてポスター発表
 1. 新島聡, 藪内弘昭, 藤井信孝, 奥野恭史, 数理的手法に基づく化合物-タンパク質間相互作用様式に関わる知識発見, 日本薬学会第 129 年会, 27Q-am212, 京都, 2009.3
 2. 藤田淳人, 藪内弘昭, 小川哲平, 荒木通啓, 大石真也, 藤井信孝, 中山和久, 奥野恭史, ケミカルゲノミクス情報を用いた標的タンパク質の予測法の開発, 日本薬学会第 129 年会, 27Q-am213, 京都, 2009.3
 3. 多門啓子, 牧口大旭, 荒木通啓, 藪内弘昭, 奥野恭史, 金子周司, 統合型薬学情報ナビゲーションシステムの開発, 日本薬学会第 129 年会, 27Q-am214, 京都, 2009.3
 4. 荒木通啓, 馮春来, 多門啓子, 藪内弘昭, 奥野恭史, 医薬品-疾病作用情報データベース, 日本薬学会第 129 年会, 27Q-am216, 京都, 2009.3
 5. 藪内弘昭, 新島聡, 奥野恭史, 新規 in silico スクリーニング手法 CGBVS のための GPCR カネルの改良, 日本薬学会第 129 年会, 26Q-am203, 京都, 2009.3

(H21 年度)

著書

「次世代創薬テクノロジー 実践：インシリコ創薬の最前線」(株)メディカル・ドゥ, 奥野恭史共著

一般口頭発表

1. 日本薬物動態学会第 2 回ビジョン・シンポジウム 奥野恭史
「ケミカルゲノミクスに基づくインシリコ創薬」
2. 構造活性フォーラム 2009 奥野恭史「多重標的創薬のためのインフォマティクス」
3. 第 5 2 回日本神経化学学会大会 奥野恭史
「創薬シード化合物探索のための画期的な新規計算手法の紹介」

4. 独立行政法人 産業技術総合研究所 生命情報科学技術者養成コース「平成 21 年度 創薬インフォマティクスコース 特別講義」奥野恭史「ケミカルゲノミクスに基づく化合物探索」
5. Bio Japan2009 奥野恭史、村上竜太、金井千里、藪内弘昭
「創薬シード化合物探索の為の画期的な新規計算手法」
6. イノベーション・ジャパン 2009 奥野恭史
創薬を加速化する医薬候補化合物の革新的探索計算技術の開発
7. 医薬品産業情報研究会 (PI フォーラム) 奥野恭史
「ケミカルゲノミクス情報を用いたパターン認識による新規スクリーニング手法」
8. 第 82 回日本生化学会大会 奥野恭史
「ケミカルゲノミクス情報に基づく創薬インフォマティクス」
9. CAC フォーラム 奥野恭史「ケミカル情報とバイオ情報の統合に基づくインシリコ創薬」
10. IT が拓く未来－創薬を加速するインフォマティクス 奥野恭史
「Chemical Genomics knowledge に基づく画期的 In silico 創薬手法の開発とそのシステム化」
11. 第 3 回京都・大阪バイオクラスター連携プロジェクト 関西バイオネットワーク「創薬研究支援ツールの高度化推進」発表交流会 奥野恭史「インシリコ創薬支援ツールで加速するゲノム創薬」

その他資料

1. 第 8 回国際パイオエクスポ・国際バイオフォーラム
村上竜太、金井千里、藪内弘昭 ポスター展示
2. JETRO NY 展示会 奥野恭史
A novel computational approach for drug discovery: Chemical genomics-based in silico ligand screening

(H22 年度)

一般口頭発表

1. Bio Japan2010 村上竜太「相互作用マシンラーニング法によるヒット化合物の探索事例」

5-2 産業財産権

該当なし