

研究開発課題別事後評価結果

1. 研究開発課題名

タンパク質化合物相互作用の網羅的予測手法とデータベースの開発

2. 代表研究者名

榊原 康文（慶應義塾大学理工学部 教授）

3. 代表研究者による成果開発概要

3-1. 研究目的

創薬の初期ステップであるリード化合物の探索において、計算機によるタンパク質化合物間相互作用予測は有用な手法であり、本研究課題ではより汎用性が高く、入手が容易であるアミノ酸配列データと化合物構造データを用いたタンパク質化合物相互作用予測手法の開発を行う。継続課題では、予測の機能だけを取り出して結合化合物およびターゲットタンパク質の網羅的探索に特化した Chemical BLAST の開発、および学習モデルライブラリーをあらかじめシステム側に用意し、ユーザ定義予測モデルに加えて、これらのモデルを混合した統合予測モデルを利用できるシステムの構築、を行なう。また、統計的相互作用予測システムとウェット検証実験によるフィードバック戦略を応用したがん治療薬のリード化合物探索、およびウェブ上で公開している COPICAT システムの機能向上と改良、を行なう。さらに、タンパク質化合物相互作用予測システムの拡張として、統計的回帰手法である SVR を適用することにより 2 値識別では予測できなかった結合強度の予測や、新薬開発が見込まれない希少疾病に対するオフラベル薬探索への適用などを行なう。

3-2. 研究開発概要

本研究課題の実施項目は、次の 4 つから構成される：

タンパク質化合物相互作用の網羅的予測手法の開発

[Chemical BLAST の開発]

相互作用予測システムの完成に基づき、予測の機能だけを取り出して、結合化合物とターゲットタンパク質の網羅的探索に特化した Chemical BLAST の開発を行なう。この Chemical BLAST の開発は、統合データベースに本システムを組み込むときに、化合物検索のための BLAST として位置づけをしている。

[学習モデルライブラリーの構築]

統計的学習手法を利用する際の問題点の一つである、与えられた学習サンプルが少量の場合、適切な予測モデルの構築が困難という問題を克服するために、複数の特化型モデルをあらかじめシステム側が用意し、ユーザ定義予測モデルに加えて、これらのモデルを混合した統合予測モデルを利用するシステムを開発する。

ケミカルバイオロジーの手法を用いた結合予測の検証とがん治療薬への応用

がん治療薬リード化合物網羅的探索への応用として、今まで実験対象としてきたアンドロゲンレセプターに加えて、継続課題では新たながん関連タンパク質に対するリード化合物の探索と、薬剤候補化合物に対するターゲットタンパク質の同定とその機能解析を行なう。また、このようにして得られたデータと知見を、相互作用予測システムにフィードバック

する。

相互作用予測システム COPICAT とデータベースの実装

継続課題においては、利用者が構築する統合予測モデルの自由度を高め、また、その構築プロセスを支援する機能を拡充する。具体的には、最適な統合予測モデル構築の支援機能、モデル流通機能、学習時のモデル性能評価指標の表示機能、を追加する。さらに、タンパク質を入力すると PubChem データベースから網羅的に結合化合物を探索する機能と、化合物を入力すると RefSeq および Ensembl データベースから網羅的にターゲットタンパク質を探索する機能を完成する。

テキストマイニングを用いた相互作用予測の検証とデータ構築

タンパク質化合物相互作用情報を文献から自動抽出するシステムを作成することを目的として、文中の未知単語を同定・単離する技術、ある未知単語がタンパク質名・化合物名・その他のいずれであるかを推定する技術、タンパク質名・化合物名・相互作用を表現する単語からなる句・文を抽出する技術、を開発する。

3 - 3 . 研究成果概要

継続課題開発においては、(i) インターネット上で公開をしているタンパク質化合物相互作用予測システム COPICAT (<http://copicat.dna.bio.keio.ac.jp/>) に対して、さまざまな機能拡張を実施した、(ii) PubChem などの膨大なデータベースから結合化合物を網羅的に探索する、化合物探索に特化した Chemical BLAST を開発した、(iii) 学習モデルライブラリーをあらかじめシステム側に用意し、ユーザ定義予測モデルに加えて、これらのモデルを混合した統合予測モデルを利用できる機能を構築した、(iv) 統計的回帰手法を適用してタンパク質化合物間相互作用の結合強度が予測できるように拡張を行い、アンドロゲンレセプターの新規阻害剤の探索に応用した、(v) 網羅的インシリコ予測、ウエット実験検証、ウエット検証実験結果のインシリコ予測へのフィードバックというプロセスを繰り返すフィードバック戦略を適用して、がん薬剤候補のリード化合物に対する網羅的な結合ターゲットタンパク質の予測を行った。具体的な成果として、(i) タンパク質化合物相互作用予測システム COPICAT において、タンパク質を入力すると PubChem データベースから網羅的に結合化合物を探索する機能と、化合物を入力すると RefSeq および Ensembl データベースから網羅的にターゲットタンパク質を探索する機能が、利用可能となった、(ii) とくに PubChem データベースの探索機能では、従来の約 120 ~ 140 倍の高速化を実現することにより、2,000 万の化合物群から現実的な時間内で探索が可能となった、(iii) ユーザが与える学習サンプルに対して高い識別性能 (AUC 値) を示す特化型モデルを、モデルライブラリーから加えると予測精度が向上することを示した、(iv) 統計的回帰手法である SVR を適用することにより 2 値識別では予測できなかった結合強度を予測することができるようになり、アンドロゲンレセプターの新規阻害剤の探索において高い精度を達成した、(v) アポトーシス抑制タンパク質 Bcl-xL の機能を阻害する化合物 incednine が相互作用する新しいタンパク質を同定して、機能的にターゲットタンパク質である可能性を示した。また、COPICAT の公開時からの全閲覧数は 67.6 万回を数える。

4 . 事後評価結果

4 - 1 . 当初計画の達成度

当初の研究開発計画から見た進捗状況や達成度等

主要な課題については計画どおり進展した。テキストマイニングについては多少の遅れがあったが最終的にはシステムに組み込むことができ、全体として計画通りに進んだと評価できる。

当初計画では想定されていなかった新たな展開

網羅的インシリコ予測を組み合わせることで、規模や効率の点で従来の方法を大きく上回る新たなリード化合物探索法を提案・開発できたことは、望ましい展開となったと思われる。

4 - 2 . 知的財産権、外部発表（論文等）等研究開発成果の状況

外部発表（論文等）

主要な研究成果における論文が投稿中や準備中とやや遅れているが、論文、学会発表ともに多数行なわれている。

データベースの公開状況

開発した予測システム COPICAT は公開され多くのアクセスがあり、研究成果は十分に公開されていると判断できる。

知的財産権

開発したシステムで用いられている予測方法で、特許出願(1 件)の審査請求を行った。

4 - 3 . 研究開発成果の公開による波及効果

データベース、プログラムや論文、学会発表など、これまでの成果公開によるバイオインフォマティクス領域およびライフサイエンス分野全体への影響や効果について

創薬分野におけるバイオインフォマティクスの有用性を実証しうるツールであり、今後の検証に期待できる。また、開発したシステムには多くのアクセスがあり、海外からも関心を持たれるなど、ある程度の影響や効果を与えたと判断できる。

国内外のデータベース高度化・標準化の動向と比較して本課題の意義

タンパク質化合物間相互作用予測において新しいアプローチを示したが、創薬のためのデータベースの標準化への影響は今のところまだ目立っていないと思われる。今後の展開次第では期待できると考えられる。

4 - 4 . 成果の実用化の可能性及び成果から予想される波及効果

成果を今後発展展開することによりバイオインフォマティクス領域およびライフサイエンス分野全体に影響や効果を及ぼすと期待できるか。

化合物とタンパク質の相互作用予測に関して網羅的な検索を可能とする方法論、大規模データへ対応できる手法から、バイオインフォマティクスによる本分野での研究に影響を及ぼすと考えられる。今後のさらなる発展や展開が期待できる。

今後、成果の生命科学、研究基盤、産業への貢献度

創薬分野におけるバイオインフォマティクスの有用性を実証しうるツールであり、今後

の生命科学的研究や産業の発展に貢献すると考えられる。

4 - 5 . 総合評価

研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。

大規模な化合物データが取り扱える網羅的な予測方法で、しかも、実験的な検証に耐える相互作用予測方法の開発が行われたことは高く評価される。しかし、予測における拘束条件について今一度検討し、データの取りこぼしおよびノイズの軽減に注力すれば、より高度な知識発見の可能性があると考えられ、今後のさらなる発展が期待される。

5 . 主な論文発表等

- *H. Jo, K. Yugi, S. Ogawa, Y. Suzuki, Y. Sakakibara. Molecular basis of chemical chaperone effects of N-octyl-β-D-glucosamine on human α-glucosidase in low/neutral pH conditions. *J Proteomics Bioinform*, 3(4), 104-112, 2010. (査読有)
- *T. Kawamura, T. Fujimaki, N. Hamanaka, K. Torii, H. Kobayashi, Y. Takahashi, M. Igarashi, N. Kinoshita, Y. Nishimura, E. Tashiro and M. Imoto. Isolation and structure elucidation of a novel androgen antagonist, arabilin, produced by *Streptomyces* sp. MK756-CF1. *J Antibiot*, 63, 601-605, 2010. (査読有)
- *K. Sato, Y. Saito, Y. Sakakibara. Gradient-based optimization of hyperparameters for base-pairing profile local alignment kernels. *Genome Inform*, 23, 128-138, 2009. (査読有)
- Nishito, Y., Osana, Y., Hachiya, T., Pendorf, K., Toyoda, A., Fujiyama, A., Itaya, M., Sakakibara, Y. (2010). Whole genome assembly of a natto production strain *Bacillus subtilis* natto from very short read data. *BMC Genomics*, 11:243. (査読有)
- T. Ohtani, S. Tsukamoto, H. Kanda, K. Misawa, Y. Urakawa, T. Fujimaki, M. Imoto, Y. Takahashi, D. Takahashi, and K. Toshima. (2010). Total synthesis of incednam, the aglycon of incednine. *Org. Lett.*, 12, 5068-5071. (査読有)
- *T. Kawamura, T. Fujimaki, N. Hamanaka, K. Torii, H. Kobayashi, Y. Takahashi, M. Igarashi, N. Kinoshita, Y. Nishimura, E. Tashiro and M. Imoto (2010). Isolation and structure elucidation of a novel androgen antagonist, arabilin, produced by *Streptomyces* sp. MK756-CF1. *J. Antibiot.*, 63, 601-605. (査読有)
- M. Kitagawa, S. Ikeda, E. Tashiro, T. Soga and M. Imoto. (2010). Metabolomic identification of the target of the filopodia protrusion inhibitor glucopiericidin A. *Chemistry and Biology*, 17, 989-998. (査読有)
- Y. Sasazawa, Y. Futamura, E. Tashiro and M. Imoto : V-ATPase inhibitors overcome Bcl-xL-mediated chemoresistance through restoration of a caspase-independent apoptotic pathway, *Cancer Science*, 100, 1460-1467, 2009. (査読有)
- 笹澤有紀子, 井本正哉 : 抗がん剤の効かないがんに挑む - アポトーシス抑制タンパ

ク質 Bcl-xL 機能阻害剤の発見 - , 化学, 64 巻(11), 74-75, 2009.

- 河村達郎, 田代悦, 井本正哉 : ケミカルバイオロジーに基づく創薬研究、分子標的薬開発の新たな挑戦, 実験医学, 2009 増刊, 213-220, 2009.

* は当課題の成果として主要な論文

参考

論文発表	国内 2 件、海外 8 件
招待、基調講演	国内 0 件、海外 1 件
口頭発表	国内 1 件、海外 0 件
ポスター発表	国内 1 件、海外 0 件
特許	国内 1 件、国際 0 件

成果プログラムおよびデータベース等

タンパク質化合物相互作用予測システム COPICAT

COPICAT システムは、タンパク質アミノ酸配列の部分配列の頻度情報および化合物構造における部分パスの頻度情報を組み合わせて多次元ベクトル化し、SVM を用いてそれらの相互作用を網羅的に予測するシステムである。PubChem データベースから結合化合物を網羅的に探索する、化合物探索に特化した Chemical BLAST の機能をあわせ持つ。

[公開日] 2008 年 4 月 1 日

[公開 URL] <http://copicat.dna.bio.keio.ac.jp/>

[公開時からの全閲覧数] 677,294 回