

平成 30 年 6 月 20 日現在

機関番号：12608
研究種目：若手研究(B)
研究期間：2015～2017
課題番号：15K16082
研究課題名(和文)大規模タンパク質構造予測による顧みられない熱帯病創薬標的タンパク質選定の効率化

研究課題名(英文)Improvement of target protein selection in NTDs based on large-scale protein structure prediction

研究代表者
石田 貴士(Ishida, Takashi)
東京工業大学・情報理工学院・准教授

研究者番号：40508355
交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では顧みられない熱帯病の病原体の一つであるトリパノソーマ寄生原虫 *Trypanosoma cruzi* の全タンパク質についてゲノムワイドにタンパク質立体構造予測の一つであるホモロジーモデリングを用いて3次元の予測構造モデルを構築し、それらを参照するためのWebインタフェースを開発することで、標的選定の際にタンパク質構造を参照しながらの検討を可能とした。また、情報量の増大によりシステムが使いづらくなる問題に対して、機械学習手法の一つであるランキング学習を用いることで標的選択を効率化する手法を開発した。

研究成果の概要(英文)：We constructed protein structure models for whole proteins encoded on genome of *trypanosoma cruzi*, which is a pathogenic agent of Chagas' disease, by using homology modeling technique. We also developed web-interface for accessing those structure models, and enabled drug discovery researchers to find more suitable target using protein structure information. In addition, we developed a new system to help the user find their demanded targets more efficiently based on relevance feedback and learning-to-rank.

研究分野：バイオインフォマティクス

キーワード：トリパノソーマ寄生原虫症 タンパク質立体構造予測 適合性フィードバック

1. 研究開始当初の背景

トリパノソーマ寄生原虫症などの「顧みられない熱帯病」向け創薬では、病原体についての基礎生物学研究が不十分なことから、その創薬標的タンパク質の選定が困難であり、構造ベース創薬を行う際の問題となっていた。

2. 研究の目的

本研究では、病原体であるトリパノソーマ寄生原虫のタンパク質についてゲノムワイドに立体構造予測を行うことで、標的選定の際に構造を参照しながらの検討を可能とし、さらにその予測モデル構造にもとづき結合ポケットの位置と質の推定を行い、それによって薬剤設計のし易いポケットを持つタンパク質を優先して選択することを可能とし、「顧みられない熱帯病」創薬標的タンパク質選定の更なる効率化を目指した。

3. 研究の方法

本研究ではトリパノソーマ寄生原虫のタンパク質についてゲノムワイドにタンパク質立体構造予測を行う。その際、構造テンプレートの探索を行い、構造テンプレートが得られた全てのタンパク質についてホモロジーモデリングを用いて3次元の予測構造モデルを生成する。その後、生成された予測構造モデルに対して、計算された全ての情報は全世界の研究者がアクセス可能なように、iNTRONDBの一部としてインターネット上で公開する。また、容易に検索、参照が可能とするためのwebインタフェースとアルゴリズムを開発する。

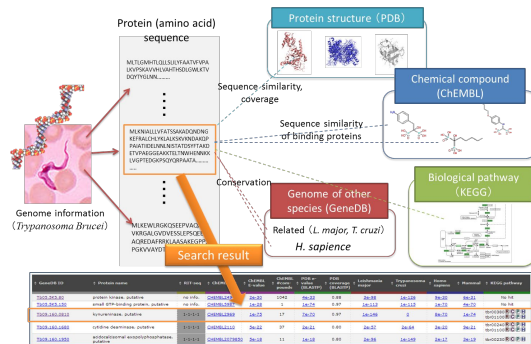


図1. 標的探索補助システム iNTRONDB の概念図

4. 研究成果

(1) スーパーコンピュータ TSUBAME2.5 上にタンパク質立体構造全自動予測パイプラインを構築し、代表的な寄生原虫症であるシャーガス病の病原体の *Trypanosoma cruzi* のゲノム上にコードされた全タンパク質である 10,339 タンパク質に対して、ホモロジー

モデリングソフトウェアである Modeller を用いて立体構造予測を行い、既知構造との間で配列の相同性が存在した 3,456 タンパク質について予測立体構造モデルを得た。また、予測構造モデルのクオリティについての評価スコアを計算し、信頼性についての情報も付加した。



図2. 予測構造例 (Spermidine synthase)

(2) 立体構造情報を参照しての選定が可能となった一方で、参照可能な情報が増えすぎ、利用者にとって逆にシステムが使いにくくなるという問題が発生した。そのため、機械学習手法の一つであるランキング学習を用いることで標的選択を効率化する手法を開発した。これは利用者に少数のタンパク質セットから、標的として好ましいタンパク質を選択してもらい、その情報を利用してランク予測モデルを構築し、10,339 タンパク質をランク付けするものである。学習アルゴリズムとして rankSVM を利用した手法では、nDCG₁ で 0.948 という性能を達成し、利用者にとって望ましいタンパク質を自動的に抽出可能であることが示された。

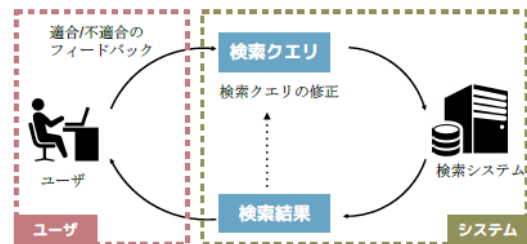


図3. 適合性フィードバックを用いた手法の概念図

また、この手法ではユーザに最初に非常に多くの情報を入力させることが必要となるという問題があったため、情報検索の技術の一つである適合性フィードバックを利用することでユーザの負担のより少ないシステムを開発した。様々なアルゴリズムを検討した結果、Random Forest ベースのアルゴリズムを用いた場合、4 回程度のイテレーションで上位のアイテムについては 0.9 を越える適合率が得られる事を確認した。

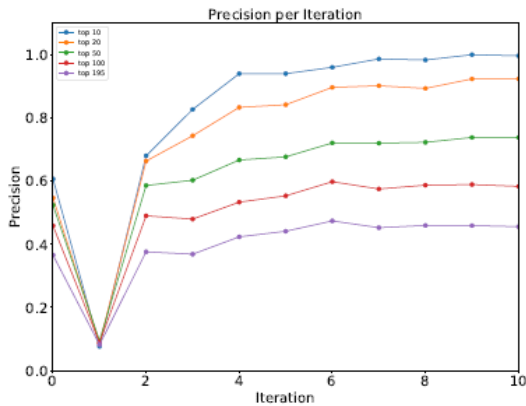


図 4. フィードバックによる予測適合率の変化

これは情報検索でよく用いられる Rocchio アルゴリズムの適合率約 0.8 を大きく上回る値となっている。さらに Random Forest を採用したことで、変数重要度の計算により、ユーザが特に重視している要素を明示的に得ることが可能となっており、単に重要と思われるタンパク質を推薦するにとどまらず、ユーザの嗜好の推定までがある程度可能であることを示した。

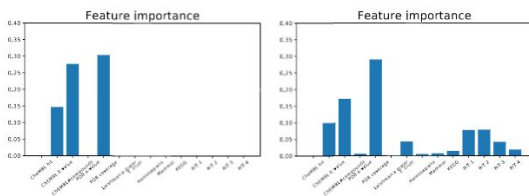


図 5. フィードバックによる変数重要度の変化

(3) また、創薬研究者が寄生原虫の株間の変異情報を立体構造上で確認し、その影響を判断できるように、薬剤標的タンパク質選定を支援するシステムの開発を行った。本研究では、感染症の中で顧みられない熱帯病の一つであるシャーガス病を対象とし、その原因となる Trypanosoma cruzi 寄生原虫の既に全ゲノム配列が決定された 7 つの株間での変異を調べ、それらの情報をタンパク質の立体構造にマップした。また、その情報を簡単に確認できるよう、JavaScript による分子構造ビューワである JSmol を使用して表示された立体構造上で変異の存在する位置を表示するための web ツールを実装した。

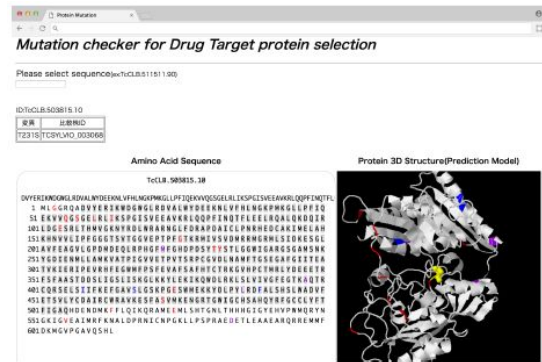


図 6. 変異情報を構造上で確認可能な Web システム

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 1 件)

Yoshino, R., Yasuo, N., Hagiwara, Y., Ishida, T., Inaoka, D. K., Amano, Y., ... & Orita, M. (2017). In silico, in vitro, X-ray crystallography, and integrated strategies for discovering spermidine synthase inhibitors for Chagas disease. Scientific reports, 査読有, 7(1), 6666.

[学会発表](計 3 件)

首我光瑛, 石田貴土. ランク学習を用いた NTDs 薬剤標的タンパク質候補選択の改良, 情報処理学会 第 46 回バイオ情報学研究会, 2016.

大石智博, 石田貴土, 薬剤標的タンパク質選定支援に向けた立体構造上での株間変異情報表示システムの開発, 情報処理学会 第 50 回バイオ情報学研究会, 2017.

首我光瑛, 石田貴土. 適合性フィードバックを用いたトリパノソーマ原虫症薬剤標的タンパク質選定の改良, 第 53 回バイオ情報学研究会, 2017.

[図書](計 0 件)

[産業財産権]

出願状況(計 0 件)

取得状況(計 0 件)

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

石田 貴士 (Ishida, Takashi)
東京工業大学・情報理工学院・准教授
研究者番号：40508355

(2)研究分担者
()

研究者番号：

(3)連携研究者
()

研究者番号：

(4)研究協力者
()