

令和 3 年 6 月 10 日現在

機関番号：14301

研究種目：若手研究

研究期間：2019～2020

課題番号：19K20398

研究課題名（和文）タンパク質の構造ダイナミクスに基づくゲノムバリエントと分子機能の関連解析

研究課題名（英文）Functional analysis of genomic variants based on structural dynamics of proteins

研究代表者

鎌田 真由美（Kamada, Mayumi）

京都大学・医学研究科・准教授

研究者番号：70749077

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：ゲノム解析技術の発展により、数多くのゲノムバリエントが報告されるようになったが、その多くが未だ機序不明である。タンパク質立体構造の動態（ダイナミクス）変化が薬剤感受性などに影響を与えることから、バリエントの機序解明には、ダイナミクスへの影響を捉えることが重要となる。タンパク質は膨大な数の原子で構成されており、構造ダイナミクスは超高次元データとして表現される。そこで本研究では、分子動力学計算で得られる時系列データから、情報学的手法により、構造ダイナミクスの変化を表す特徴の抽出と可視化表現の検討を行った。

研究成果の学術的意義や社会的意義

精密医療実現を担う「ゲノム医療」では、ゲノム解析により得られるゲノムの違い（バリエント）とそれらに対する既知情報に基づき、診断および治療法の選択を行う。しかし実際、ゲノム解析では機序および臨床的な意義が不明なバリエントが数多く検出され、診断・治療方針決定に結びつかないことが、ゲノム医療におけるボトルネックとなっている。本研究では、機序不明なバリエントの分子への影響を情報学的に評価することを試みた。今後様々な系に対して検証が必要であるが、低次元での特徴表現及び客観的な評価が可能であることから、効率的な機序に関わる因子探索の一助となることが期待できる。

研究成果の概要（英文）：With the development of genome analysis technology, many genomic variants have been reported, but the mechanism of many of them is still unknown. Since changes in the dynamics of protein structures affect drug sensitivity, it is crucial to understand the effects of variants on dynamics to elucidate their mechanisms. Proteins are composed of a large number of atoms, and the structural dynamics are represented as ultra-high dimensional data. In this study, we extracted features representing changes in structural dynamics from time-series data obtained by molecular dynamics simulation using informatics methods and investigated visualization representations.

研究分野：バイオインフォマティクス

キーワード：タンパク質構造ダイナミクス 時系列解析 低次元特徴量表現

1. 研究開始当初の背景

近年臨床でも応用が進められている「ゲノム医療」は、次世代シーケンサーなどを用いた遺伝子解析等で得られるゲノムバリエーションに基づき、診断や治療方針決定および疾患の予防を行う医療である。特にがんでは、バリエーションが検出された特定の分子を標的とした薬剤(分子標的薬)が多く開発されており、体細胞系列変異(後天的に細胞のがん化の過程等で生じる変異)の有無による最適な薬剤選択が可能であると期待されている。しかし一方で、機序及び疾患への関連が不明なバリエーション(Variants of uncertain significance, VUS)が検出されることが多く、薬剤選択や治療方針決定を困難にする要因となっている。バリエーションによる機能や薬剤感受性への影響を知るには、バリエーションを有する遺伝子にコードされるタンパク質への影響を知る必要がある。特に、タンパク質立体構造の動的挙動(ダイナミクス)が、分子機能や薬剤との親和性に重要な役割を示すことが知られていることから、ダイナミクスを考慮した解析が重要である。タンパク質立体構造のダイナミクスを検証するには、分子動力学(Molecular Dynamics, MD)シミュレーションが有効である。これまでにMDシミュレーションを用い、バリエーションによるタンパク質への影響(機能にとって有害かどうか)を検証する研究は多数報告されている。しかし、分子の持つ複雑な時空間階層性から、時系列的要素の扱いがされていないものや、平均的な構造安定性に関する議論に終始しているものがほとんどであること、さらに、構造的変化の解釈は各系における専門的な知識が必要であることから、主観的評価に偏る課題があった。

2. 研究の目的

本研究では、バリエーションのダイナミクスへの影響を客観的かつ効率的に検討することを目的に、MDシミュレーションで得られる時系列データから、タンパク質立体構造ダイナミクスの動的な変化を表す特徴抽出およびその可視化手法の検討を行うことを目的とする。通常、MDシミュレーションの結果解釈は、計算の結果得られるエネルギープロファイルやトポロジー変化量等の比較検証と、専用のソフトウェアを用いた3次元構造の重ね合わせ等の確認により行われることが多い。しかし目視による検証は主観的になりがちであり、また時間のかかる作業である。そこで本研究では、ウェーブレット変換と情報学的手法で定量化されたダイナミクスの特徴量を作成し、さらに可視化することで客観的な解釈と解析の効率化を図ることを目指す。

3. 研究の方法

タンパク質立体構造のダイナミクスは、二次構造単位の大きな動きから、一原子における微小な動きまでが複雑に重なり合った“時空間階層性”を持っている。そこで本研究では、時間軸を残したまま経時的な動きの特徴を捉えることが可能である、時間-周波数解析の一手法である「ウェーブレット変換」を用いる。ウェーブレット変換は、大きな動きに相当する低周波帯の情報から、微小な動きに相当する高周波帯の情報までを、同時に取得することが可能である。今回対象とする変異体と正常体との動的挙動の変化は、ダイナミクスを表現する時系列データにおける異常(正常と異なるパターン)検出と捉えることができる。そこで、ウェーブレット変換を用いた時系列の異常検出フレームワークを拡張し、動的挙動の変化の検出を行う。

また、タンパク質立体構造は多自由度からなるため、ウェーブレット変換の結果に対して、特異値分解(Singular value decomposition)を用い、タンパク質を構成する原子の振動方向に関する特徴量を作成する。作成した特徴量により表現される時系列に対し、定量的な比較検討を行うため時系列間の類似度を評価するDynamic Time Warping (DTW)を適用する。本研究で用いる特徴抽出手法では、各原子に対する特徴ベクトルを得ることができる。しかし、アミノ酸置換を起こすバリエーションを有する場合、変異体間および野生型とでは、分子を構成する原子数が異なる。そこで、t-SNE (t-Distributed Stochastic Neighbor Embedding)を用いて同次元への埋め込みを行い可視化することで、分子間での違いについて検証する。

本研究では、薬剤耐性について生化学的実験でも多くの研究報告のあるEGFR(epidermal growth factor receptor)の野生型と2つの変異型を対象に、MDシミュレーションを実施し時系列データを取得した。本研究では、Gromacs 4.6.5及び国立遺伝学研究所のスーパーコンピュータを用いて全原子シミュレーションを実施した。シミュレーションに用いる立体構造データは、Protein Data Bankより取得し、構造データに欠損領域が存在する場合、SWISS-MODELを用いたホモロジーモデリングにより、補完を行った。

4. 研究成果

EGFR の野生型および2変異体(薬剤奏功変異 L858R, 耐性変異 G719S)を対象に、50ns の MD シミュレーションを実施した。シミュレーションで得られたそれぞれの分子に対する時系列データから 200ps ごとにトラジェクトリを抽出し、Morlet wavelet を用いた連続ウェーブレット変換を適用した。対象とする薬剤耐性変化に関連する動きのスケールと一致するよう周波数スケール範囲の策定を行い、200ps~20ns の動きに相当する周波数スケールを抽出した。さらに、特異値分解により、タンパク質を構成する原子の振動方向に関する特徴量を作成した。抽出した特徴量(左特異ベクトル)を用いてタンパク質と薬剤分子との協調した動きを評価するための指標を作成した。

まず、特徴量で表現される分子間の違いを検証するため、t-SNE を用いて同次元への埋め込みを行い、可視化を行った。検証の結果、結晶構造解析論文で報告されてる薬剤感受性に係るアミノ酸残基を、射影空間上での違いとして確認することができた(図1)。

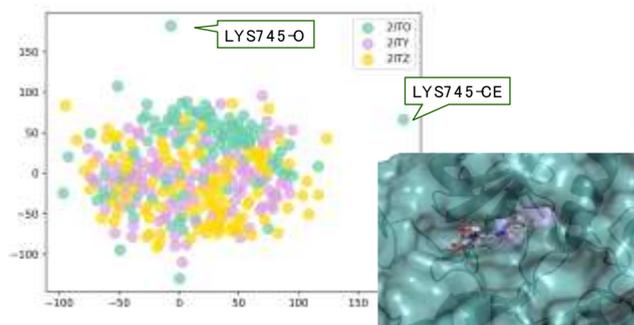


図 1. t-SNE による変異体間での違いの検証

また、3分子間で共通する原子を対象に、時系列間の類似度を評価する Dynamic Time Warping (DTW)を適用し、分子間の類似度を検証した(図2)。検証の結果、既知の変異体間の関係性を示すには、DTWへ与える時系列の特徴量作成方法に課題があることが明らかになった。今後、指標の作成方法の改良により、より有効な比較検証が可能であると考えられる。

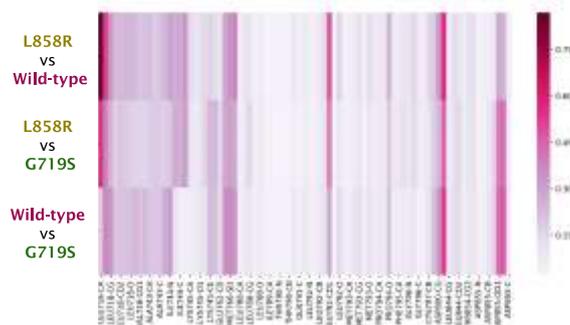


図 2. DTW による変異体間での類似性検証

5. 主な発表論文等

- [雑誌論文] (計 0 件)
- [学会発表] (計 2 件)
- [図書] (計 0 件)
- [産業財産権]
- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 1件 / うち国際学会 1件）

1. 発表者名 Mayumi Kamada, Mikito Toda
2. 発表標題 Evaluation of differences in protein dynamics caused by genomic variant using time-series data
3. 学会等名 The annual international conference on Intelligent Systems for Molecular Biology (ISMB) (国際学会)
4. 発表年 2019年～2020年

1. 発表者名 鎌田真由美
2. 発表標題 生体分子の構造ダイナミクスとゲノム医療への応用
3. 学会等名 「複雑系数理：物理・化学・生物・情報 と カオス」シンポジウム（招待講演）
4. 発表年 2020年～2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------