

機関番号：12601

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500269

研究課題名（和文）タンパク質の不規則領域のデータベースの構築と機能解析

研究課題名（英文）Database Development and Functional Analysis of Protein Disorder Regions

研究代表者

清水 謙多郎（SHIMIZU KENTARO）

東京大学・大学院農学生命科学研究科・教授

研究者番号：80178970

研究成果の概要（和文）：PDBに登録されている各タンパク質構造に対し、disorderの「程度」を評価する基準を求め、データベースとしてまとめた。また、disorder領域と機能部位との関係について解析を行った。

研究成果の概要（英文）：We developed a new method for defining the degree of disorder of protein structures. The results of this analysis were stored in our protein disorder database. We also analyzed the relation between the disorder regions and functional sites of proteins.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,500,000	450,000	1,950,000
2009年度	1,300,000	390,000	1,690,000
2010年度	700,000	210,000	910,000
総計	3,500,000	1,050,000	4,550,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：プロテオーム、生体生命情報学、タンパク質、不規則領域

## 1. 研究開始当初の背景

タンパク質の不規則領域（disorder領域）とは、タンパク質中で特定の構造をとっていない領域をいう。長年にわたって、こうした特定の構造をもたない領域は、タンパク質の機能発現に重要ではないと考えられてきたが、近年、disorder領域にリン酸化部位がよく見られることや、ある分子標的が近づくことによって特定の構造をとることなどが分かってきており、disorder領域が注目されるようになってきた。

disorderの定義は、「タンパク質のX線結晶構造解析において、座標を特定できなかったアミノ酸残基」とされている。X線結晶構造解析では、X線回折データから電子密度マップを作成した後、そのマップに合うように原子を置いていくが、結晶中でよく動く原子

はその電子密度が「疎」になってしまい、その原子の座標を特定できなくなるのである。しかし、座標が特定できているアミノ酸残基においても、電子密度が疎であるようなものが見られる。つまり、電子密度が疎かどうかは連続的な値であり、disorderの度合いを数値化することができる。このような情報をデータベース化した例はほとんどなく、また、タンパク質の機能とdisorder領域との関連も十分に探究されていないのが現状である。

## 2. 研究の目的

本研究では、まず、PDBの各構造に登録された構造因子を直接解析し電子密度マップを再現するとともに、温度因子、溶媒露出面積、二次構造も合わせて解析することで、実験事実からorderの程度（order degree）

を求める手法を開発し、PDBに登録された各構造に対して利用者がその基準を自由に指定し検索することができるデータベースを構築する。このデータベースをもとに、従来、異なる基準で抽出あるいは予測され解析の対象とされてきた disorder 領域を、実験事実に基づく統一した基準で広範に解析し、disorder の「程度」も考慮して、より精確かつ詳細にそれらの特性と機能との関係を明らかにする。さらに、disorder 領域と相互作用部位、細胞局在、細胞分化、成長因子など disorder 領域と関連が深いと言われている機能をもつタンパク質について詳細な解析を行い、最終的には disorder 領域の機能データベースを構築することを目指す。

### 3. 研究の方法

あるアミノ酸残基の周りの電子密度が疎であるということを定量化する方法として、そのアミノ酸残基に属する原子の座標中心からある一定の距離  $d$  内にある電子密度を積分して求まる値をその残基の「周囲にある電子数」と定義し、その電子数を「そのアミノ酸残基が本来持っている電子数」で割ることで求まる値を「order degree」とすることとした。order degree の値は、周りの電子密度が疎である程低くなるので、この値が低いほど disorder していると見なせることになる。

上記の  $d$  は、原子半径にそのタンパク質の平均温度因子から求まる変位を足し合わせた距離とした。平均温度因子から求まる変位を足す理由は、X線結晶構造解析時の温度が高いほど原子は動きやすくなり、周りの電子密度が疎になってしまうので、その効果を補償するために積分範囲を広げるためである。

### 4. 研究成果

上記の手法に基づき、disorder 領域の解析を行い、disorder 領域のデータベースのプロトタイプを構築した。

(1) PDBに登録されている、X線結晶構造解析によって構造決定された各タンパク質構造に対し、温度因子を抽出した。また、PDBに登録されているすべての構造に対し、DSSPを用いて、溶媒露出表面積、二次構造判定の結果を収集した。これらの結果を内部的なデータベース（「構造特徴データベース」と名付ける）に記録した。

(2) PDBに登録されている各タンパク質構造に対し、構造因子を用い、逆フーリエ変換・フーリエ変換を行って電子密度マップを再現した。各残基の周辺の電子密度を求め、disorder の「程度」を評価する基準を求めた。得られた結果を内部的なデータベース（「電子密度マップデータベース」と名付ける）に記録した。

order degree の計算を、図 1 に示す PDB

ID: 1qg8 のタンパク質構造に適用した結果を示す。この結果から、座標が特定できなかった残基付近のアミノ酸残基の order degree の値が低くなっていることが分かる。この結果は、先に可視化した電子密度マップ（図 1）と合致しており、この方法が disorder の程度を示す指標として使えることが確認できました。

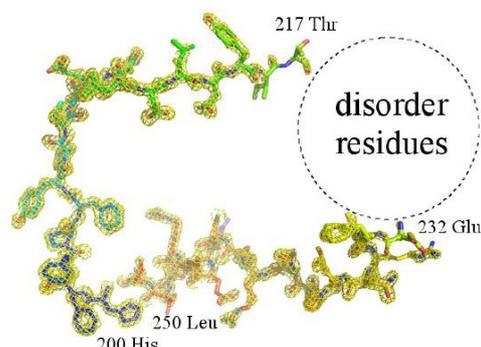


図 1 電子密度マップの例  
PDB ID: 1qg8 の例、残基位置 218-231 は disorder 領域のため不可視。

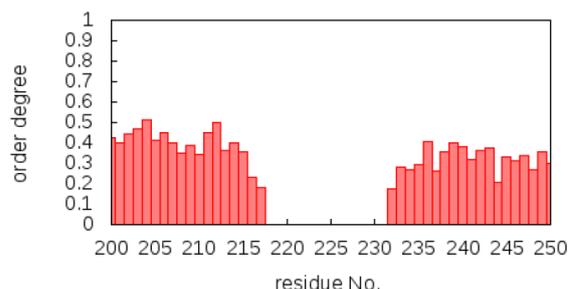


図 2 order degree の計算  
PDB ID: 1qg8 の残基位置 200-250 の order の程度を示す。残基位置 218-231 は 0 である。

上の結果で、周囲に密な電子密度をもっているアミノ酸残基の order degree の値が低くなっているが、これらのアミノ酸残基は、理想的には 1.0 に近い値をとるべきであり、さらに周囲の電子密度をより詳しく調べて、積分範囲などを再検討する予定である。

そのほか、本研究では、利用者が与える基準に応じて、構造特徴データベース、電子密度マップデータベースのそれぞれから、柔軟に disorder 領域を検索できるシステムの開発に着手した。

### 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 27 件）

- ① K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu: Protein-Protein Docking

- Using Multi-layered Spherical Basis Functions, Proceedings of the 2011 International Conference on Bioscience, Biochemistry and Bioinformatics, 査読有, 2011, 342-347.
- ② S. Someya, M. Kakuta, M. Morita, K. Sumikoshi, W. Cao, Z. Ge, O. Hirose, S. Nakamura, T. Terada, K. Shimizu: Prediction of carbohydrate-binding proteins from sequences using support vector machines. Advances in Bioinformatics, 査読有, 2010, 289301.
- ③ T. Furuta, K. Shimizu, T. Terada: Accurate prediction of native tertiary structure of protein using molecular dynamics simulation with the aid of the knowledge of secondary structures, Chemical Physics Letters, 査読有, 472, 2009, 134-139.
- ④ S. Nakamura, K. Shimizu: Comprehensive analysis of sequence-structure relationships in the loop regions of proteins, GIW 2009, 査読有, 23, 2009, 106-116.
- ⑤ W. Cao, K. Sumikoshi, T. Terada, S. Nakamura, K. Kitamoto, K. Shimizu: Computational Protocol for Screening GPI-anchored Proteins, Proceedings of the First International Conference on Bioinformatics and Computational Biology (BICoB), Springer Lecture Notes in Bioinformatics Series, 査読有, 5462, 2009, 164-175.
- ⑥ M. Kakuta, S. Nakamura, K. Shimizu: Prediction of protein-protein interaction sites using only sequence information and using both sequence and structural information, IPSJ Transactions on Bioinformatics, 査読有, 49, 2008, 25-35.
- ⑦ R. Ishitani, T. Terada, K. Shimizu: Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations, Molecular Simulation, 査読有, 34, 2008, 327-336.
- ⑧ M. Morita, S. Nakamura, K. Shimizu: Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures, PROTEINS: Structure, Function, and Bioinformatics, 査読有, 73, 2, 2008, 468-479.
- [学会発表] (計46件)
- ① W. Cao, S. Someya, M. Kakuta, K. Sumikoshi, S. Nakamura, T. Terada, K. Shimizu: Carbohydrate-binding proteins: A sequence-based prediction model and the database development, The 2010 Annual Conference of the Japanese Society for Bioinformatics (JSBi2010), 九州大学 (福岡) (2010.12).
- ② 角越和也, 中村周吾, 清水謙多郎: タンパク質-タンパク質ドッキングへの分子内フレキシビリティの導入, 日本生物物理学会第48回年会, 東北大学 (宮城) (2010.9).
- ③ 寺田透, 西村麻彦, 清水謙多郎: 二次構造情報を取り入れた分子動力学法によるタンパク質立体構造予測法の一般性の検証, 日本生物物理学会第48回年会, 東北大学 (宮城) (2010.9).
- ④ 池澤弘貴, 森田瑞樹, 清水謙多郎: バイオロジカルユニットを考慮したリガンド結合タンパク質データベースの構築, 日本生物物理学会第47回年会, 徳島文理大学 (徳島) (2009.10).
- ⑤ 西村麻彦, 寺田透, 清水謙多郎: all- $\beta$  タンパク質の二次構造拘束付きマルチカノニカル分子動力学シミュレーション, 日本生物物理学会第47回年会, 徳島文理大学 (徳島) (2009.10).
- ⑥ 角田将典, 角越和也, 中村周吾, 清水謙多郎: ドメイン配列を用いたタンパク質間相互作用予測, 日本生物物理学会第46回年会, 福岡国際会議場 (福岡) (2008.12).
- ⑦ 城野亮太, 渡邊佑輔, 清水謙多郎, 寺田透: 生体分子の溶媒効果: マルチカノニカル QM/MM-MD による研究, 日本生物物理学会第46回年会, 福岡国際会議場 (福岡) (2008.12).
- ⑧ 角越和也, 寺田透, 中村周吾, 清水謙多郎: 精密化プロセスを含めたタンパク質-タンパク質ドッキング手法の開発, 日本生物物理学会第46回年会, 福岡国際会議場 (福岡) (2008.12).
- ⑨ 清水謙多郎: タンパク質構造予測, JST ゲノムリテラシー講座, JST サイエンスプラザ (東京) (2008.7).
- ⑩ 中村周吾, 清水謙多郎: タンパク質のオーダー・ディスオーダーの配列構造相関の解析, 第8回日本蛋白質科学会年会, タワーホール船堀 (東京) (2008.6).
- ⑪ 角田将典, 角越和也, 中村周吾, 清水謙多郎: タンパク質ペアに対するカーネル関数を用いたタンパク質間相互作用予測, 第8回日本蛋白質科学会年会, タワーホール船堀 (東京) (2008.6).
- ⑫ 西村麻彦, 西謙一郎, 寺田透, 清水謙多郎: 分子動力学シミュレーションによる WW ドメインのフォールディング機構の解析, 第8回日本蛋白質科学会年会,

タワーホール船堀（東京）（2008.6）.

〔図書〕（計2件）

- ① 清水謙多郎: タンパク質の構造予測手法の開発, 科学技術振興機構中国総合研究センター, 中国科学技術月報 第44号 中国・日本科学最前線 - 研究の現場から - 2011年版, 2011, 137-140.
- ② 中村周吾, 清水謙多郎: タンパク質の ab initio 構造予測, メディカルドゥ, 遺伝子医学 MOOK 14 次世代創薬テクノロジー 実践: インシリコ創薬の最前線, 2009, 49-54.

〔その他〕

ホームページ等

<http://www.bi.a.u-tokyo.ac.jp/~shimizu/>

## 6. 研究組織

### (1) 研究代表者

清水 謙多郎 (SHIMIZU KENTARO)

東京大学・大学院農学生命科学研究科

・教授

研究者番号: 80178970