

平成 22 年 6 月 10 日現在

研究種目：若手研究(B)

研究期間：2008～2009

課題番号：20700274

研究課題名（和文） 空間統計学を用いた生体情報システムを解析する手法の開発

研究課題名（英文） Application of spatial statistics to investigate two biological networks.

研究代表者

根本 航 (NEMOTO WATARU)

独立行政法人産業技術総合研究所・生命情報工学研究センター・産総研特別研究員

研究者番号：10455438

研究成果の概要（和文）：本研究では、二つの異なるネットワークデータ(タンパク質分子内の残基間ネットワーク・タンパク質間相互作用ネットワーク)に、空間統計学を適用し、タンパク質の生化学的機能・生物学的機能を解明する次の二つの手法を開発した。

手法 A：タンパク質の立体構造が一つ与えられた時に、その立体構造とフォールドが同じで、かつ、機能部位の位置が保存されている配列空間の範囲を予測する手法。

手法 B：PPI ネットワークデータが与えられた時に、アミノ酸配列情報を利用して、ネットワーク内でハブとなっているタンパク質が、デートハブなのかパーティーハブなのかを予測する手法。

研究成果の概要（英文）： In order to investigate two different biological networks, we developed two methods by applying spatial statistics. I. A novel method to recognize the sequences, which share the same functional regions on the structure as those of the query structure. II. A method to classify party hubs and date hubs in protein-protein interaction network.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,800,000	540,000	2,340,000
2009 年度	1,500,000	450,000	1,950,000
年度			
年度			
年度			
総計	3,300,000	990,000	4,290,000

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：タンパク質、機能部位予測、空間統計学、相同配列、立体構造

## 1. 研究開始当初の背景

本研究では、二つの手法開発を行った。

手法 A の背景：ホモロジーモデリングによ

る立体構造予測の精度は、選択される鋳型に依存する。良い鋳型の第一条件は予測標的タンパク質と同じフォールドをとることだが、それだけでは不十分である。なぜなら、二つのタンパク質のフォールドが類似して

も機能部位(触媒残基、タンパク質間相互作用のインターフェイス、核酸結合部位等)の位置まで同じであるとは限らないためである。実際、レクチンのように、類似フォールド同士でもタンパク質間相互作用(PPI)のためのインターフェイスの位置を変化させている例がある。ドッキングシミュレーションなど、モデリング後に解析を行うためには高精度の予測立体構造モデルが必要であるため、予測標的タンパク質とフォールドが同じというだけでなく、機能部位の位置も同じであるタンパク質立体構造を鋳型として用いる必要がある。しかし、研究開始当時のフォールド認識プログラム開発の方向性は、いかに遠縁の配列を検出するかであり、機能部位の位置の保存について考慮した取り組みはなされていなかった。

手法 B の背景：過去の PPI ネットワーク研究は、スケールフリー性・モジュール構造・ネットワークモチーフなど、専らネットワークの静的な性質が問題とされてきた。しかし、様々な生物種のゲノム配列が解読され、様々な細胞・状態で個々のタンパク質の発現量が計測されたことにより、ネットワークの時空間的变化などの動的挙動を議論できる段階になった。時空間的变化を考慮した研究の例としては、発現量を指標にしたハブタンパク質の分類がある。PPI ネットワークにおいて定義されるパーティーハブ(PH)は同じ時空間において同時に多くのタンパク質と相互作用するハブタンパク質である。一方、デートハブ(DH)は異なる時空間において異なるタンパク質と相互作用するタンパク質であり、異なる時空間に存在するネットワーク同士を結ぶ役割を担い、ネットワーク内で動的な変化を司る。このように両者を区別することは、PPI ネットワークの時空間的に変動する領域を知るのに重要であるが、両者の区別は与えられた PPI ネットワークを観察しただけでは分からないため、詳細な実験により明らかにするしかない。しかし、実験により明らかにされる DH と PH の区別に関する情報の網羅性は低く、配列情報から予測する手法の開発が必要であった。

## 2 . 研究の目的

次の二つの手法開発に取り組んだ。

手法 A：タンパク質の立体構造が一つ与えられた時に、その立体構造とフォールドが同じで、かつ、機能部位の位置が保存されている配列空間の範囲を予測する手法。

手法 B：PPI ネットワークデータが与えられた時に、アミノ酸配列情報を利用して、ネッ

トワーク内でハブとなっているタンパク質が、PHなのかDHなのかを予測する手法。

手法 A を用いて、配列空間と立体構造空間の対応付けを行えば、PDB にすでに登録されている立体構造と、フォールドだけでなく機能をも共有する配列空間の範囲が分かる。新たに立体構造が解かれねばならないタンパク質数を見積もることができるため、今後の Structural Genomics におけるターゲットタンパク質の選択指針として利用できる。

手法 B により、タンパク質のアミノ酸配列情報と PPI ネットワーク構造のみから DH と PH の区別を予測することにより、PPI ネットワークの動的な描像・時空間的挙動の解析に利用できる。

## 3 . 研究の方法

### 手法 A

任意の立体構造既知タンパク質 X と「フォールドが同じで、かつ、機能部位が同じ」配列空間の範囲を予測することを考える。我々のこれまでの研究により、タンパク質の保存度に関しては、空間的自己相関が観測されることが見出されていた。系統樹上、タンパク質 X とは遠縁な配列をマルチプルアライメントに含むほど空間的自己相関は強くなり、ある時点で最大値をとり、減少に転じる。この性質を利用し、任意の立体構造とフォールドが同じで、かつ、機能部位の位置が同じ配列空間の範囲の予測に取り組んだ。

### 手法 B

PH と DH について報告されている以下の性質 I-III を利用することで、DHらしさ、逆に PHらしさを示す指標をデザインした。

- I. DH には、PH に比べ長いディスオーダー領域(DR)が多い。
  - II. PH には、ドメインとして定義される進化的に保存されている領域が多い。
  - III. PH どうしは、たいていの場合、互いに相互作用しあう傾向にある。
- DR が進化的に保存される傾向にあるか否かについては議論が続いているが、ドメインとして定義される領域よりも保存度が高いとは考えにくい。

I-III を総合すると、PH はネットワーク内で保存度の高いハブタンパク質の局所的クラスタとして検出される可能性が高い。逆に、長い DR を含むハブがハブタンパク質どうし

で相互作用していない場合、DH である可能性が高い。

#### 4. 研究成果

手法 A: アミノ酸配列の保存度の、立体構造上における空間的自己相関の強度を利用して、タンパク質の立体構造が一つ与えられた時に、その立体構造とフォールドが同じで、かつ、機能部位の位置が保存されている配列空間の範囲を予測する手法を開発した。

この研究成果を、本研究に先立ち開発した相同配列と立体構造を利用する機能部位予測手法と統合し、Functional REgion Prediction of a protein by using Spatial statistics (FREPS)として提供した。ウェブサービスの名称は、我々自信の先行研究であるタンパク質の機能部位予測手法に沿った名前になっている。しかし、特定の立体構造の PDB ID を指定するか、もしくは、立体構造データを PDB フォーマットでアップロードすることにより、入力した立体構造とフォールドだけでなく機能部位をも共有すると予測される配列データセットを入手することができる。



図 1. FREPS : <http://freps.cbrc.jp>

例えば、PDB ID と CHAIN 名を入力することで特定の立体構造を指定すると、その立体構造とフォールドだけでなく機能部位の位置をも共有すると予測される配列データセットを取得することができる。入力された立体構造との機能部位の共有度を示す Degree of Spatial Autocorrelation (DSPAC) というスコアのプロットを図 2 に示した。



図 2. 指定した立体構造 (PDB : 1EL1 chain A) に対応する配列との配列一致度と DSPAC。

図 2 では、C-type Lysozyme の立体構造を入力した。Lactalbumin は C-type Lysozyme の構造ホモログであるが、非酵素化したタンパク質である。図 2 において、X 軸は、指定された立体構造に対応する配列に対する配列一致度についての閾値を表している。例えば、Percent Sequence Identity = 60% の点では、Query の配列との配列一致度が 60% 以上の配列のみが含まれる配列データセットが用いられて、DSPAC のプロットが打たれている。

グラフ左端より DSPAC が極大値をとる配列一致度までは、対応する配列データセット内に、Lactalbumin の配列は含まれない。しかし、極大値に対応する配列一致度より右側に位置する配列一致度における配列データセット内には Lactalbumin の配列が含まれている。このように、DSPAC を指標として利用することにより、酵素である Lysozyme の配列と、Lysozyme が非酵素化した Lysozyme の構造ホモログである Lactalbumin とを区別することができた。

同様にして、ホモログであるが酵素と非酵素のタンパク質同士、同様にホモログだがタンパク質間相互作用のためのインターフェイスが異なるタンパク質同士のデータセットを作成し、現在、機能部位の違いをも識別するフォールド認識プログラムとしてのさらなる性能評価を行っている。

ある立体構造に対してフォールドだけでなく機能部位の位置をも共有する配列を予測することで、機能部位に関しより精緻な立体構造モデリングの実現が期待される。

手法 B:

(1) *S.cerevisiae* のタンパク質間相互作用ネットワークでは、ハブタンパク質に着目すると、各タンパク質の保存度に関する空間的自己相関が観察されることを示した(図 3)。



図 3. *S.cerevisiae* のタンパク質間相互作用ネットワークにおけるハブタンパク質の保存度に関する空間的自己相関。横軸は、各ハ

ブタンパク質の保存度を示す。縦軸は、各ハブタンパク質と相互作用するタンパク質の保存度の平均値を表す。

(2)また、PH のネットワーク上近傍に位置するタンパク質の進化的な保存度は、DH の場合と比べ高い傾向にあることも示された (図 3)。すなわち、PH は進化的に保存度の高いタンパク質と相互作用する傾向があることが示された。

(3)さらに、この PH と DH の性質を利用することで、タンパク質間相互作用ネットワークデータが与えられた時に、アミノ酸配列情報と複合体立体構造情報を利用して、ネットワーク内でハブとなっているタンパク質の PH らしさ、DH らしさを表現する指標をデザインした。

DH と PH の区別は、タンパク質間相互作用ネットワークの動的・時空間的挙動と強い関連性を持つため、本研究の成果はタンパク質間相互作用ネットワークを解析するための新たな指標として期待される。

#### 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)

根本航, “Title Bioinformatics for GPCR Oligomerization” 生物物理, 49, 308-9 2009. (査読無)

[学会発表](計 8 件)

#### 招待講演

Nemoto W. PepCon-2010, After a Solution for the Machines of Life. 21 Mar 2010. Beijing International Convention Center, Beijing, China. Session 6-2 GPCRs in Drug Discovery. “Bioinformatics Approaches toward Controlling Protein-Protein Interaction between GPCRs”

根本航. 第五回バイオインフォマティクス学会九州支部講演会、「バイオインフォマティクスの最近のトピックス」2009年2月27日.九州大学システム生命科学府講義棟. “相同配列と立体構造を利用したタンパク質機能部位予測”

#### 口頭発表

根本航, 福井一彦, 藤博幸. 日本生化学会. 2009年10月24日.神戸. “Gタンパク質共役型受容体オリゴマー研究へのバイオインフォマティクスからのアプローチ -G-Protein Interaction Partners (GRIP)-”.

#### ポスター発表

Nemoto W. Toh H. 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) & 8th European Conference on Computational Biology (ECCB). [June 29th, 2009. Stockholm, Sweden] “A novel index to evaluate the appropriateness of a set of homologous sequences for functional region prediction”. Poster session.

Nemoto W. Fukui K. Toh H. 3DSIG 2009: The 5th Structural Bioinformatics and Computational Biophysics. [June 27th, 2009. Stockholm, Sweden] “Bioinformatics for GPCRs Oligomerization”. Poster session.

Nemoto W and Toh H. European Conference on Computational Biology. [September 22nd - September 26th. 2008. Cagliari, Sardinia, Italy] “A novel index to select appropriate sequences for multiple sequence alignment to predict functional regions of a protein”. Poster session.

Nemoto W and Toh H. 16th Annual International Conference Intelligent Systems for Molecular Biology, Toronto Convention Centre (South Building), Toronto, Canada. July 19th - July 23th. 2008. A novel statistical method to predict functional regions of a protein. Poster session.

根本航, 藤博幸 第8回日本蛋白質科学会年会[タワーホール船堀, 東京, 10 June - 12 July, 2008] 「配列情報と立体構造情報を利用するタンパク質機能部位予測手法」Poster Session

[その他]  
ホームページ等

<http://freps.cbrc.jp/>

6 . 研究組織

(1)研究代表者

根本 航 (NEMOTO WATARU)

独立行政法人産業技術総合研究所・生命情報工学研究センター・産総研特別研究員

研究者番号：10455438

(2)研究分担者

( )

研究者番号：

(3)連携研究者

( )

研究者番号：