

令和 2 年 6 月 5 日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K07315

研究課題名(和文) 既知フォールドの再配線による新規フォールド予測法の開発

研究課題名(英文) Development of a new fold protein structure prediction method using rewiring known folds

研究代表者

千見寺 浄慈 (Chikenji, George)

名古屋大学・工学研究科・助教

研究者番号：10420366

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：タンパク質のアミノ酸配列とその天然構造の関係を理解することは、遺伝暗号解読問題として興味深いばかりでなく、医学的な応用という観点でも重要である。本研究では新規フォールドを持つタンパク質に対しても立体構造予測が可能な手法の一つである既知構造の再配線という手法に着目した。再配線を行う際に最も重要なことは、物理的に許される再配線と許されないものを見分けることであるが、これまではこれを見分ける基準がなかった。本研究で、局所構造のルールのフラストレーションがこの基準になる事がわかった。この結果は立体構造予測ばかりでなく、合理的デザインにおける目標構造をする際にも役に立つ。

研究成果の学術的意義や社会的意義

タンパク質の天然構造をそのアミノ酸配列のみから予測したり(立体構造予測)、逆に望みの立体構造を持つアミノ酸配列を予測すること(デザイン)は、遺伝暗号解読問題として興味深いばかりでなく、医学的な応用という観点でも重要である。立体構造予測とデザインの両方を行う際に、どのような構造がタンパク質の天然構造になりうるのかを把握しておくことは問題を解く上で本質的に重要である。本研究では立体構造予測の研究を通して、これまで知られていなかった天然構造の条件の一つを明らかにする事が出来た。

研究成果の概要(英文)：Understanding the relationship between the amino acid sequence of a protein and its native structure is not only interesting as a genetic code deciphering problem, but also important in terms of medical applications. In this study, we focused on the rewiring technique of known structures, which is one of the methods that can predict the 3D structure of proteins with novel folds. The most important thing of the rewiring technique is to distinguish between physically permissible folds and impossible ones, but until now there has been no standard for distinguishing this. In the present study, we find that the frustration of the local structure rule is the criterion for this. The results are useful not only for three-dimensional structure prediction, but also for target structure selection in rational protein design.

研究分野：生物物理

キーワード：タンパク質

## 1. 研究開始当初の背景

1960年代に Anfinsen が「タンパク質の天然構造は自由エネルギー最小の状態である」ということを実験的に示して以来、アミノ酸配列の情報のみからタンパク質の立体構造を理論的に予測するという、いわゆる「タンパク質の折り畳み問題」は多くの研究者を惹きつけてきた。進化情報に頼らないタンパク質立体構造予測(非ホモロジーモデリング)が可能になるということは、アミノ酸配列という遺伝子型の情報から立体構造という表現型への情報への変換過程を理解するという観点で興味深いばかりでなく、この情報変換過程が翻訳時のリボソームのような大掛かりな装置を必要とせず、それ単独で自発的に行われるという物理的観点からも興味深い課題である。また、このような非ホモロジーモデリングによるタンパク質立体構造予測は学問的な興味だけではなく、実用的な観点からも重要である。医学・薬学・生物学・工学的に重要なタンパク質の中には、実験的な構造解析が困難でかつ、現時点ではホモロジーモデリングが不可能なものが依然として多く存在する。Levitt らの報告によると、配列データベースに存在する全タンパク質ドメインのうち、ホモロジーモデリング可能なファミリーは、全体の約 25% に過ぎなかった(PNAS 2009)。そのようなものの中で、工学的に注目を浴びている実験的構造解析が困難なタンパク質の例として、石油を作るタンパク質であるアルカン合成酵素(Schirmer et al. Science 2010)など、多数の重要なタンパク質が挙げられる。このようなタンパク質の機能の理解・およびそれに基づく合理的改良には構造情報が重要である。実際、乳ガンの増殖メカニズムの解明に立体構造予測が重要な役割を果たした例が報告されており(Yoshimaru et al. Nature Comm. 2013)、構造情報の重要性がますます高まっている。以上のことから、非ホモロジーモデリングによるタンパク質立体構造は実用的な観点からも極めて重要な課題であると言える。

非ホモロジーモデリングによる立体構造予測の現状を客観的に把握するには、世界的規模で行われているタンパク質立体構造予測のブラインドコンテスト CASP の結果を眺めるのが最もよい方法である。最近の CASP の結果によると、1997年に発表された Baker らによるフラグメントアセンブリ(Fragment Assembly:FA)法の登場により、非ホモロジーモデリングによる立体構造予測の成功例が報告され始め、現在に至るまで FA 法が最も強力な非ホモロジーモデリング手法であるとされている。しかし、FA 法の成功確率は極めて低く、最も優秀なチームの成功確率でさえ 10%に満たないと報告されている(Tai et al. Proteins 2014)。FA 法の成功率はなぜ低いのだろうか？我々は FA 法の長所とその限界に関して丁寧に調べてきた。その結果、FA 法で行われているような配列類似性検索を用いた局所構造予測では、正解構造に近い局所構造を予測することが極めて難しいこと、そして局所構造予測に失敗した場合、安定な二次構造パッキング構造をシミュレーション中に生成することが不可能であることが明らかとなった(Chikenji et al. PNAS 2006)。また、Kim らは FA 法による複雑なトポロジーをもったタンパク質の立体構造予測が失敗する理由は、FA 法では複雑なトポロジー構造のサンプリングが難しいからである、と報告している(JMB 2009)。以上から、非ホモロジーモデリングによるタンパク質立体構造予測の精度を向上させるための最重要課題の一つとして、複雑なトポロジーであっても効率的に構造生成できる手法を開発することが挙げられる。

どのようにすれば複雑なトポロジー構造を効率的に生成できるだろうか？本研究では、「配列順序をあえて無視した構造類似性」に着目する。配列順序を無視した構造類似性とは、トポロジー(二次構造のつながり方)は異なるが、二次構造のパッキングの仕方は類似している性質のことである。以前から、CASP の新規フォールド部門の問題として出題されたタンパク質の天然構造の多くは、既知構造と二次構造のパッキングの仕方が似ている事が指摘されていた。つまり、複雑なトポロジーをもつ新規フォールド構造であっても、既知のフォールドの二次構造のつながり方を変える(再配線する)だけで極めて高い精度で構造生成できる、という事である。このように既知のフォールドのつながりかえでどれくらいの新規フォールドの構造を生成できるのだろうか？我々はこの疑問に答えるために、高性能な配列順序を無視した構造比較方法を開発し(Minami et al. BMC bioinformatics 2013)、それを用いて異なるフォールド間で同じ二次構造パッキングを共有しているかを調べた(Minami et al. PloS ONE 2014)。その結果、約 80% のフォールドは他のフォールドのつながりかえでよく表現できることが明らかとなった。これは新規フォールド構造であっても、80% の確率で既知構造を改変することにより構造生成できることを示唆している。

しかしながら、再配線する際に理解しておかなければならない深刻な問題がある。それは、現時点では、再配線したフォールドが物理的にタンパク質の天然構造として存在できるものなのか否かを判定する基準がない、という事である。また、この問題を解明することは、なぜタンパク質のフォールドは約 1000 種類程しかないのか、という問題に対する回答を与える事ができるという意味でも重要である。

## 2. 研究の目的

本研究の目標は、再配線を用いて新規フォールド構造を持つタンパク質の立体構造を予測できるようにするために、既知のタンパク質立体構造の二次構造の順番を入れ替えて(再配線して)新規フォールド構造を生成して新規フォールド構造を作成した際に、その新規フォールドが物理的にタンパク質の天然構造として存在できるものなのか否かを判定する基準

を作る事である。そしてそれをもとに、なぜタンパク質のフォールドの種類が極めて少ないのかという問題に対する回答を与えることを目指す。

### 3. 研究の方法

#### 1. 高精度なシーケンシャル構造アラインメント手法の開発

再配線の恩恵を正しく定量的に評価するためには、配列順序を無視した構造アラインメントと配列順序を保った(シーケンシャルな)構造アラインメントの両方が必要である。我々はこれまで、世界最高レベルの高精度な配列順序を無視した構造アラインメント手法を開発してきたが、これと同等なシーケンシャル構造アラインメント手法が必要である。本研究ではまず準備段階として、高精度なシーケンシャル構造アラインメント手法を開発する。そしてそれを用いて再配線の恩恵を正しく評価する。

#### 2. 再配線によって作られたフォールドが物理的に天然構造として存在できるか否かを判定する基準の開発

一つの既知構造を用いて再配線を行った場合、膨大な数のフォールドを生成する事ができる。しかし、それらが全て物理的に天然構造として存在できるものかどうかは自明でない。実際、現在タンパク質の立体構造データベースに登録されているフォールドの数を考えると、再配線を行ったフォールドの数と比べると極めて少ない事がわかる。これらのデータベース中に存在しないものは物理的に不可能だから存在しないのだろうか？あるいは、物理的には可能だが進化がたまたまサンプリングしていないだけなのだろうか？本研究では、ループの繋がり方によって物理的に天然構造として存在できるものとそうでないものを見分ける手法を開発し、理論的に考えられる膨大なフォールドのうち、物理的に可能なものはどれなのかを明らかにする。

### 4. 研究成果

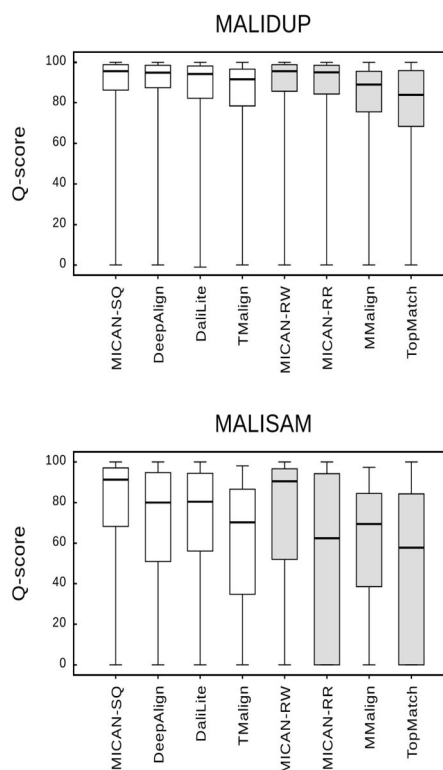
#### 1. 高精度なシーケンシャル構造アラインメント手法の開発

再配線の恩恵を正しく定量的に評価するために、シーケンシャルな構造アラインメント(MICAN-SQ)を開発した。MICAN-SQは、我々がこれまでに開発してきた世界最高レベルの高精度な配列順序を無視した構造アラインメント手法(MICAN)に、配列順序を保つような束縛条件を加える改良を施すことによって開発した。それゆえ、MICANプログラムはアラインメント時に配列順序を保ったもの(SQ)、二次構造の順序は無視するが鎖の方向は保ったもの(RW)、二次構造の順序も鎖の方向も考慮しないもの(RR)の3つのアラインメント方法を一つのプログラムで実行する事ができる。さらに、これら3つのプログラムは束縛条件のみが異なるので、鎖の繋がり方の違いの影響のみを計算する事ができる世界で唯一のプログラムである。

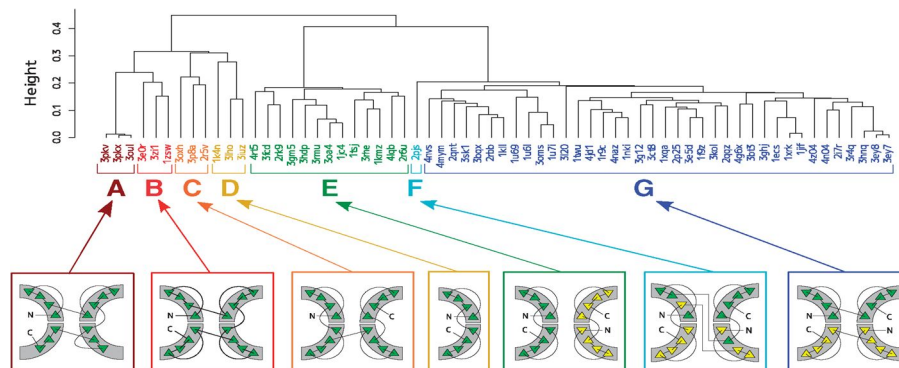
MICAN-SQの性能を評価するために、構造アラインメントの性能評価でよく使われているベンチマークセット(MALIDUPとMALISAM)を用いてベンチマークテストを行い、既存手法(DeepAlign、Dali-Lite、TMalign、MICAN-RW、MICAN-RR、MMalign、TopMatch)と性能を比較した(右図)。性能評価の結果、既存のどのプログラムよりも統計的優位差を持ってMICAN-SQは優れている事がわかった。

MICAN-SQを用いて、新規フォールドを再配線を用いて予測する際の恩恵がどれくらいあるか見積もったところ、約80%のフォールドに対して再配線によって類似した構造が構築できる事がわかった。これはTMalignを用いた評価である約40%と比べて非常に大きく、TMalignは構造類似性を過大評価している事が示唆された。

MICAN-SQのもう一つの大きな特徴は、他のどのシーケンシャルな構造アラインメントプログラムとも異なり、モノマー同士、モノマーと多量体、あるいは(鎖の本数が異なる)多量体と多量体のペアに対しても、鎖の対応関係を含めた構造アラインメントを行う事ができる、ということである。これを応用して、これまで手動で行われてきた



glyoxalase/bleomycin resistance protein (BRP)ファミリーに対して構造アラインメントを行い、系統樹を作成したところ、正しく系統樹が構築できたばかりでなく、新しい構造変化の様式を特定する事が出来た（下図）。



## 2. 4本のストランドと2本のヘリックスからなるフォールドの解析

ループの繋がり方によって物理的に天然構造として存在できるものとそうでないものを見分ける方法を開発するために、4本のストランドと2本のヘリックスからなるフォールドを例題として解析を行った。

まず、理論上可能な4本のストランドと2本のヘリックスからなるフォールド全てを列挙し、その中のどのフォールドがタンパク質立体構造データベースに登録されているかを計算した（下図）。

Homology	42	7	-	7	4	1	1	1	1	1	-
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
ILC loop	4	4	4	3	3	3	3	3	3	3	3
NLB pair	2	1	0	1	1	1	1	1	0	0	0
Homology	3	-	-	1	-	1	1	-	1	-	-

Homology	1	1	-	-	-	4	1	-	-	1	-	-
	(l)	(m)	(n)	(o)	(p)	(q)	(r)	(s)	(t)	(u)	(v)	(w)
ILC loop	2	2	2	2	2	2	2	2	2	1	1	1
NLB pair	1	1	1	1	1	0	0	0	0	0	0	0
Homology	-	-	-	-	-	1	-	-	-	-	-	-

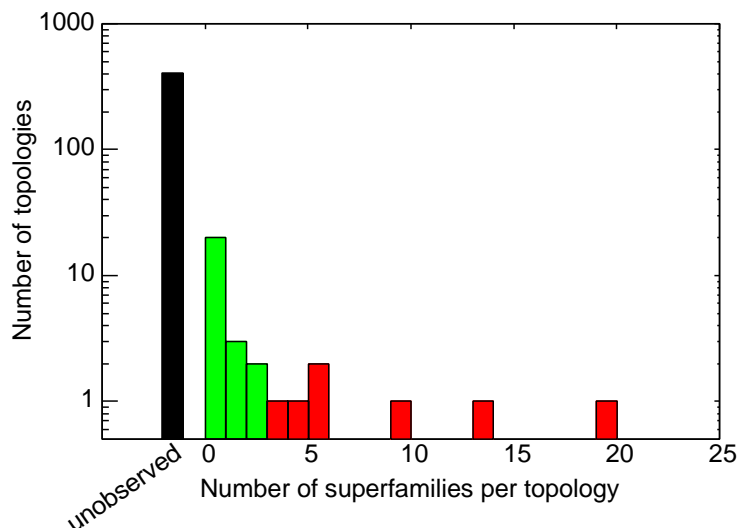
理論上可能なものとデータベースに登録されているものを比べると、データベース上に多く存在しているものは、二つのレイヤー間を交差する回数（ILC loop）が多く、シートの水素結合が配列に沿って非局所的なもの（NLB pair）が多い事がわかった。このことは、デザインしやすい構造は協同的な相互作用を持っている必要があること、逆に、協同的な相互作用がないフォールドは物理的に天然構造として存在しにくい、という事が示唆された。この事を利用すると、立体構造予測をする際に、再配線して生成した構造の非局所的な接触の割合を調べる事で、生成した構造の中から天然構造になりにくいものをフィルタリングする事ができるようになった。課題として、鎖のN末端からC末端の方向だけを入れ替えた“リバーシの関係”にあるフォールドは、NLB pairの数も ILC loop の数も同じであるが、データベース中での存在数は大きく異なっているのは、物理的な要因の帰結なのか進化のサンプリングバイアスなのか明らかにする必要がある、というものが残された。

## 3. 平行 シートの解析

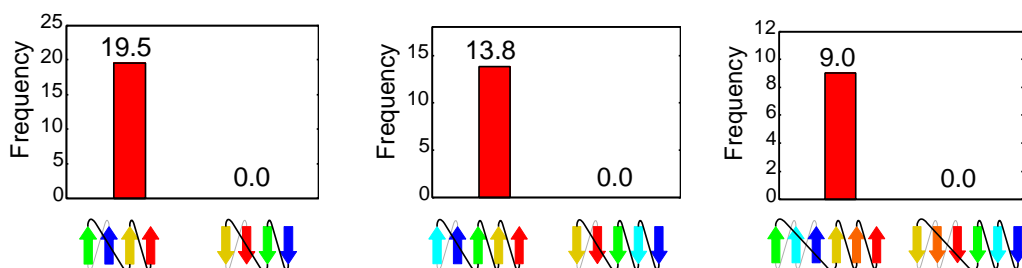
4本のストランドと2本のヘリックスからなるフォールドの解析で課題として残

された「N 末端から C 末端への鎖の方向の逆転」の効果を理解するために、問題を単純化して、平行 シートのみを対象として研究を行った。

具体的には3本から6本からなる平行 シートフォールドを考えた。これは理論的には435個の繋がり方があるが、そのほとんどはデータベース中に存在せず、ほんの数種類しか存在しないことがわかった(下図)。



データベース中に存在しているフォールドの中で、特に頻出するものはスーパーフォールドと呼ばれるが、スーパーフォールドのリバース構造はデータベース中に存在しないか、存在するとしても極めて少なかった。また、スーパーフォールドは必ず C 末端のストランドが シートの端に位置しているのに対し、スーパーフォールドのリバース構造は N 末端の ストランドが シートの端に位置していた(下図)。



この非対称性を説明する物理的要因を探ったところ、以下の三つの物理的なルールを満たす事ができるフォールドがスーパーフォールドで、ストランドの配置や長さをどうやっても三つのルールを同時に満たす事が出来ない(フラストレーションがある)フォールドがデータベース中に存在しない事がわかった。

ここで、三つのルールとは、  
 - ユニットは右巻きでなければならない (J. Richardson 1977)、  
 - ユニットは反平行でなければならない (Koga et al. 2012) 平行 シートのレジスタシフトは0以上でなければならない(本研究で発見)である。なお、レジスタシフトのルールは本研究で立体構造データベース解析により発見し、次いでそれが物理で説明できることも証明した。

以上のように、ミクロなルールを組み合わせることによって物理的に禁止されるフォールドとそうでないものがある事がわかった。物理的に禁止されているフォールドは全て立体構造データベース中にない事からこの理論の妥当性が支持される。この理論を用いることによってタンパク質立体構造予測の際に再配線によって生成した構造が物理的に天然構造として存在できるのか否か判断できるようになった。また、なぜタンパク質のフォールドの種類が有限であることを説明する理論としても有効なものであると考える事ができる。

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Shintaro Minami, Kengo Sawada, Motonori Ota, and George Chikenji	4. 巻 34
2. 論文標題 MICAN-SQ: A sequential protein structure alignment program that is applicable to monomers and all types of oligomers	5. 発行年 2018年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 3324,3331
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/bty369	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Shintaro Minami, George Chikenji, and Motonori Ota	4. 巻 26
2. 論文標題 Rules for connectivity of secondary structure elements in protein: two-layer sandwiches	5. 発行年 2017年
3. 雑誌名 Protein Science	6. 最初と最後の頁 2257-2267
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/pro.3285	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Shuntaro Chiba, 他30名 (George Chikenji は16番目の著者)	4. 巻 7
2. 論文標題 An iterative compound screening contest method for identifying target protein inhibitors using the tyrosine-protein kinase Yes	5. 発行年 2017年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 1-13
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-017-10275-4	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Masaki Sasai, George Chikenji, Tomoki P Terada	4. 巻 13
2. 論文標題 Cooperativity and modularity in protein folding	5. 発行年 2017年
3. 雑誌名 Biophysics and Physicobiology	6. 最初と最後の頁 281-293
掲載論文のDOI (デジタルオブジェクト識別子) biophysico.13.0_281	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tatsuya Okuno, Koya Kato, Shintaro Minami, Tomoki P Terada, Masaki Sasai, George Chikenji	4. 巻 13
2. 論文標題 Importance of consensus region of multiple-ligand templates in a virtual screening method	5. 発行年 2017年
3. 雑誌名 Biophysics and Physicobiology	6. 最初と最後の頁 49-156
掲載論文のDOI (デジタルオブジェクト識別子) biophysico.13.0_149	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計25件 (うち招待講演 4件 / うち国際学会 0件)

1. 発表者名 福田孝貴, 千見寺浄慈
2. 発表標題 立体構造データベース中での ループモチーフの出現頻度の偏りの起源
3. 学会等名 第18回日本蛋白質科学会年会
4. 発表年 2018年

1. 発表者名 今川駿, 古賀信康, 千見寺浄慈
2. 発表標題 シートタンパク質のデノボデザインにおける主鎖構造設計図の選択基準
3. 学会等名 第18回日本蛋白質科学会年会
4. 発表年 2018年

1. 発表者名 千見寺浄慈, 南慎太郎, 古賀信康
2. 発表標題 - - - モチーフにおけるレジスタシフトの非対称性
3. 学会等名 第18回日本蛋白質科学会年会
4. 発表年 2018年

1. 発表者名 千見寺浄慈
2. 発表標題 X線結晶構造解析から得られたタンパク質立体構造データの網羅的情報解析による知識抽出
3. 学会等名 第10回放射光学会若手研究会 (招待講演)
4. 発表年 2018年

1. 発表者名 犬飼耕平, 笹井理生, 千見寺浄慈
2. 発表標題 Diversification of P-loop protein structure simulated by imposing the functional requirement as a selection pressure
3. 学会等名 第56回生物物理学会年会
4. 発表年 2018年

1. 発表者名 福田孝貴, 千見寺浄慈
2. 発表標題 A skewed distribution of psi-loop motifs in the protein structure database
3. 学会等名 第56回生物物理学会年会
4. 発表年 2018年

1. 発表者名 今川駿, 古賀信康, 千見寺浄慈
2. 発表標題 Criteria for evaluating designability of pure parallel beta sheet structures
3. 学会等名 第56回生物物理学会年会
4. 発表年 2018年



1. 発表者名 千見寺浄慈
2. 発表標題 デザインしやすいタンパク質立体構造の条件
3. 学会等名 学際計算物理学研究会 (招待講演)
4. 発表年 2018年

1. 発表者名 千見寺 浄慈
2. 発表標題 デザイン可能なタンパク質構造の条件
3. 学会等名 日本計算数理工学会 第34回 計算数理工学フォーラム (招待講演)
4. 発表年 2018年

1. 発表者名 千見寺 浄慈、今川駿、南慎太郎
2. 発表標題 What are the structural features of superfolds? a case study of beta-sheet proteins
3. 学会等名 日本生物物理学会
4. 発表年 2017年

1. 発表者名 共田恭輔、増山陽汰、千見寺 浄慈
2. 発表標題 Revisiting a classical threading method with novel scoring function of sequence-structure compatibility
3. 学会等名 日本生物物理学会
4. 発表年 2017年

1. 発表者名 南慎太郎、古賀理恵、千見寺 浄慈、古賀信康
2. 発表標題 天然に存在しないフォールドを持つタンパク質の合理的デザイン
3. 学会等名 日本生物物理学会
4. 発表年 2017年

1. 発表者名 今川駿、古賀信康、千見寺 浄慈
2. 発表標題 Some factors that make a structure of ba beta-sheet protein more designable
3. 学会等名 日本生物物理学会
4. 発表年 2017年

1. 発表者名 小林大祐、千見寺 浄慈
2. 発表標題 Combining a docking software with a ligand-based virtual screening method, VS-APPLE
3. 学会等名 日本生物物理学会
4. 発表年 2017年

1. 発表者名 今川駿、古賀信康、千見寺 浄慈
2. 発表標題 シートタンパク質のデノボデザインでデザインしやすい ストランドの配置とはどんなものだろうか？
3. 学会等名 蛋白質科学会
4. 発表年 2017年

1. 発表者名 南慎太郎、千見寺 浄慈、古賀信康
2. 発表標題 新規フォールドタンパク質の合理デザインに向けて
3. 学会等名 蛋白質科学会
4. 発表年 2017年

1. 発表者名 千見寺 浄慈、南慎太郎
2. 発表標題 スーパーフォールドの決定因子：平行 シートタンパク質を例にして
3. 学会等名 蛋白質科学会
4. 発表年 2017年

1. 発表者名 西山 俊介、南 慎太郎、千見寺 浄慈
2. 発表標題 タンパク質立体構造におけるレア構造
3. 学会等名 蛋白質科学会
4. 発表年 2016年

1. 発表者名 千見寺 浄慈
2. 発表標題 タンパク質構造を利用したバーチャルスクリーニングとその周辺
3. 学会等名 第5回生命医薬情報学連合大会（招待講演）
4. 発表年 2016年

1. 発表者名 西山 俊介、南 慎太郎、千見寺 浄慈
2. 発表標題 New rules for beta sheet topology of protein structures
3. 学会等名 第5回生命医薬情報学連合大会
4. 発表年 2016年

1. 発表者名 西山 俊介、南 慎太郎、千見寺 浄慈
2. 発表標題 New rules of protein structures
3. 学会等名 日本生物物理学会
4. 発表年 2016年

1. 発表者名 共田 恭輔、増山 陽太、千見寺 浄慈
2. 発表標題 A new threading method based on the physical characteristics of sequence-structure compatibility
3. 学会等名 日本生物物理学会
4. 発表年 2016年

1. 発表者名 小林 大祐、千見寺 浄慈
2. 発表標題 Analysis of protein complexes structures towards rational design of inhibitors of Protein-protein interactions (PPIs)
3. 学会等名 日本生物物理学会
4. 発表年 2016年

1. 発表者名 南 慎太郎、千見寺 浄慈、太田 元規
2. 発表標題 二次構造順序の変化によって起こる蛋白質フォールドの多様化
3. 学会等名 日本生物物理学会
4. 発表年 2016年

1. 発表者名 増山 陽太、千見寺 浄慈
2. 発表標題 立体構造予測において疎水効果を評価するための新しい指標： 仮想原子の周りのコンタクト数
3. 学会等名 日本生物物理学会
4. 発表年 2016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
---------------------------	-----------------------	----