様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ８ 月 １８ 日現在

機関番号：８２４０１
研究種目：基盤研究(B)（一般）
研究期間：2018～2020
課題番号：１８Ｈ０２３９５

研究課題名（和文）タンパク質の構造設計の計算方法の研究開発および実験検証

研究課題名（英文）Development of Protein Design Methods and Their Experimental Verification

研究代表者
ＺＨＡＮＧ　ＫＡＭ（Zhang, Kam）

国立研究開発法人理化学研究所・生命機能科学研究センター・チームリーダー

研究者番号：６０５５８９０６

交付決定額（研究期間全体）：（直接経費）　　13,300,000円

研究成果の概要（和文）：私たちは、バックボーンの柔軟性を備えた計算タンパク質設計法を開発しました。これは、既知のタンパク質構造のローカル構造環境をエネルギーとともに利用して配列設計をガイドするものです。 これは、非連続接触アミノ酸残基モチーフのカスタマイズされたライブラリに基づいています。 この新しい方法を、RNA ポリメラーゼの 4 回対称の 8 ブレード ベータ プロペラと 2 回対称のダブル psi ベータ バレル コアの設計に適用しました。 これらの設計は、実験的に決定された構造がコンピュータによる設計モデルと厳密に一致することが X 線結晶構造解析によって検証されています。

研究成果の学術的意義や社会的意義
私たちが開発した柔軟なバックボーン タンパク質設計法は、現在の最先端のタンパク質設計の可能性を拡張しました。 それは、新しいバイオ センサー、治療法、および診断法の開発に大きな可能性を示しています。 RNA ポリメラーゼの 2 重対称ダブル psi ベータ バレル コアの成功した設計は、現代の複雑なタンパク質は、祖先のより単純なタンパク質からの遺伝子重複、融合、および多様化イベントから進化した可能性があるという考えを支持しています。

研究成果の概要（英文）：We have developed a computational protein design method with backbone flexibility, which exploits local structural environments in known protein structures together with energy to guide sequence design. It is based on customized libraries of non-contiguous in-contact amino acid residue motifs. We have applied this new method in the design of a four-fold symmetric eight-bladed beta-propeller and a two-fold symmetric double-psi beta-barrel core of RNA polymerase. These designs have been validated by X-ray crystallography that the experimentally determined structures closely matches the computationally design models.

研究分野：計算タンパク質設計

キーワード：計算タンパク質設計

１．研究開始当初の背景

Proteins are one of the most important component of living organisms. The ability to design proteins with a specified structure and thereby conferring it with a desired function would have tremendous impact on our ability to develop new therapeutics, diagnostics and biosensors. Researchers have been trying to computationally design proteins for several decades. However, there exist only a very limited number of experimentally verified computationally designed proteins, half of which have been obtained with Rosetta Design. The Rosetta Design approach starts from a rigid protein backbone, then searches for optimal sequences that are compatible with the given structure evaluated by an empirically derived energy function. The imposition of rigid backbone has greatly reduced the conformational space that needs to be searched, which increased the computational efficiency. However, it also significantly limited the power of this approach to achieve design goals.

２．研究の目的

The objective of our proposal is to develop a novel computational method for the *de novo* design of proteins that uses discontinuous fragments and flexible backbones to extend the range of protein design and to validate this method through the computational design and experimental verification of several novel proteins.

３．研究の方法

We propose to use a library of naturally occurring sequence segments that are known to fold into a given structural fragments to dramatically reduce the sequence space that has to be searched. Specifically, we plan to use customized libraries of non-contiguous in-contact amino acid residue motifs. We then use an evolutionary approach such as the estimation of distributions algorithm to efficiently search the sequence space by learning from previous populations.

４．研究成果

We have developed FRAGGER, which is a protein fragment picker that allows protein fragment databases to be created and queried. All fragment lengths are supported and any set of PDB files can be used to create a database. FRAGGER can efficiently search a fragment database with a query fragment and a distance threshold. Matching fragments are ranked by distance to the query. The query fragment can have structural gaps and the allowed amino acid sequences matching a query can be constrained via a regular expression of one-letter amino acid codes. FRAGGER also incorporates a tool to compute the backbone RMSD of one versus many fragments in high-throughput. FRAGGER should be useful for protein design, loop grafting and related structural bioinformatics tasks.

We have developed a computational protein design method with backbone flexibility called SHADES, which is a data-driven method that exploits local structural environments in known protein structures together with energy to guide sequence design. It is based on customized libraries of non-contiguous in-contact amino acid residue motifs. We have tested SHADES on a public benchmark of 40 proteins selected from different protein families and showed excellent results. WD40 proteins are a subfamily of propeller proteins, with a pseudo-symmetrical fold made up of subdomains called blades. By computationally reverse-engineering the duplication, fusion and diversification events in the evolutionary history of a WD40 protein, a perfectly symmetrical homolog called Tako8 was made. We have used SHADES to redesign Tako8 to create Ika8, a four-fold symmetrical protein in which neighbouring blades carry compensating charges. These artificial eight-bladed rings may find applications in bionanotechnology and as models to study the folding and evolution of WD40 proteins.

We have developed a protein sequence fitness scoring function that implements sequence and corresponding secondary structural information at tripeptide levels to differentiate natural and non-natural proteins. The proposed fitness function is extensively validated on a dataset of about 210,000 natural and non-natural protein sequences and benchmarked with existing methods for differentiating natural and non-natural proteins. The high sensitivity, specificity and percentage

accuracy of the fitness function demonstrates its potential application for sampling the protein sequences with higher probability of mimicking natural proteins. The protein sequence characterization aided by the proposed fitness function could facilitate the exploration of new perspectives in the design of novel functional proteins.

Modern proteins with complex structures are thought to have evolved from small and simple ancient proteins with prototype folds. How such prototype proteins emerged on the primitive earth remains enigmatic. RNA polymerases are ancient proteins found in all kingdoms of life and have a large complex structure consisting of multiple domains. We hypothesize that the double-psi beta-barrel (DPBB) domain at the core of RNA polymerase is the ancestor of modern RNA polymerases. We have designed a two-fold symmetric protein of the DPBB fold, which consists of a duplicated half DPBB sequence. The design of completely symmetrical DPBB sequences were carried out using the "reverse engineering evolution" computational protein design approach. We aligned the orthologous sequences of DPBB to generate a phylogenetic tree. The aligned sequences and constructed phylogenetic tree were used together to generate putative ancestral consensus sequences. The predicted ancestral sequences were then mapped onto a symmetrical DPBB backbone structural model and their energies were calculated. The top-scored designs were selected for experimental validation based on the Rosetta scores, RMSD from design template, predicted solubility and visual inspection. These designs have been validated by X-ray crystallography that the experimentally determined structures closely matches the computationally design models. This design provided support for the notion that modern day complex proteins could have been evolved from the gene duplication, fusion and diversification events from ancestral simpler proteins.

Publications:

1. Berenger, F., Simoncini, D., Voet, A., Shrestha, R., Zhang, K. Y. J. (2018) Fragger: a protein fragment picker for structural queries. *F1000Research*, **6**, 1722. doi:10.12688/f1000research.12486.1.
2. Simoncini, D., Zhang, K. Y. J., Schiex, T., Barbe, S. (2019) A Structural Homology Approach for Computational Protein Design with Flexible Backbone. *Bioinformatics*, **35**, 2418-2426. doi:10.1093/bioinformatics/bty975.
3. Noguchi, H., Addy, C., Simoncini, D., Wouters, S., Mylemans, B., Van Meervelt, L., Schiex, T., Zhang, K. Y. J., Tame, J. R. H., Voet, A. R. D. (2019) Computational design of symmetrical 8-bladed ß-propeller proteins. *IUCrJ*, **6**, 46-55. doi:10.1107/S205225251801480X.
4. Kaushik, R., Zhang, K. Y. J. (2020) A Protein Sequence Fitness Function for Identifying Natural and Non-Natural Proteins. *Proteins: Struct., Funct., Bioinf.*, **88**, 1271–1284. doi:10.1002/prot.25900.
5. Berenger, F., Kumar, A., Zhang, K. Y. J., Yamanishi, Y. (2021) Lean-Docking: Exploiting Ligands' Predicted Docking Scores to Accelerate Molecular Docking. *J. Chem. Inf. Model.*, **61**, 2341−2352. https://doi.org/10.1021/acs.jcim.0c01452.
6. *Kumar, N., Kaushik, R., Tennakoon, C., Uversky, V. N., Longhi, S., Zhang, K. Y. J., Bhatia, S. (2021) Insights into the evolutionary forces that shape the codon usage in the viral genome segments encoding intrinsically disordered protein regions. *Brief. Bioinformatics*, https://doi.org/10.1093/bib/bbab145.
7. Yamashita, M., Kuehn, H. S., Okuyama, K., Okada, S., Inoue, Y., Mitsuiki, N., Imai, K., Takagi, M., Kanegane, H., Takeuchi, M., Shimojo, N., Tsumura, M., Padhi, A. K., Zhang, K. Y. J., Boisson, B., Casanova, J.-L., Ohara, O., Rosenzweig, S. D., Taniuchi, I., Morio, T. (2021) A Variant in Human AIOLOS Impairs Adaptive Immunity by Interfering with IKAROS. *Nat. Immunol*. **22**, 893–903. https://doi.org/10.1038/s41590-021-00951-z.
8. Kukimoto-Niino, M., Katsura, K., Kaushik, R., Ehara, H., Yokoyama, T., Uchikubo-Kamo, T., Nakagawa, R., Mishima-Tsumagari, C., Yonemochi, M., Ikeda, M., Hanada, K., Zhang, K. Y. J., Shirouzu, M. (2021) Cryo-EM structure of the human ELMO1-DOCK5-Rac1 complex ELMO1-DOCK5-Rac1 structure. *Sci. Adv.*, **7**, eabg3147. https://doi.org/10.1126/sciadv.abg3147.

５．主な発表論文等

〔雑誌論文〕　計8件（うち査読付論文　8件／うち国際共著　8件／うちオープンアクセス　3件）

１．著者名 Berenger Francois、Simoncini David、Voet Arnout、Shrestha Rojan、Zhang Kam Y.J.	４．巻 6
２．論文標題 Fragger: a protein fragment picker for structural queries	５．発行年 2018年
３．雑誌名 F1000Research	６．最初と最後の頁 1722～1722
掲載論文のDOI（デジタルオブジェクト識別子） 10.12688/f1000research.12486.1	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 Simoncini David、Zhang Kam Y J、Schiex Thomas、Barbe Sophie	４．巻 35
２．論文標題 A structural homology approach for computational protein design with flexible backbone	５．発行年 2018年
３．雑誌名 Bioinformatics	６．最初と最後の頁 2418～2426
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/bty975	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 該当する

１．著者名 Noguchi Hiroki、Addy Christine、Simoncini David、Wouters Staf、Mylemans Bram、Van Meervelt Luc、Schiex Thomas、Zhang Kam Y. J.、Tame Jeremy R. H.、Voet Arnout R. D.	４．巻 6
２．論文標題 Computational design of symmetrical eight-bladed β-propeller proteins	５．発行年 2019年
３．雑誌名 IUCrJ	６．最初と最後の頁 46～55
掲載論文のDOI（デジタルオブジェクト識別子） 10.1107/S205225251801480X	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 Kaushik Rahul、Zhang Kam Y. J.	４．巻 88
２．論文標題 A protein sequence fitness function for identifying natural and nonnatural proteins	５．発行年 2020年
３．雑誌名 Proteins: Structure, Function, and Bioinformatics	６．最初と最後の頁 1271～1284
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/prot.25900	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 該当する

1．著者名 Berenger Francois、Kumar Ashutosh、Zhang Kam Y. J.、Yamanishi Yoshihiro	4．巻 61
2．論文標題 Lean-Docking: Exploiting Ligands’ Predicted Docking Scores to Accelerate Molecular Docking	5．発行年 2021年
3．雑誌名 Journal of Chemical Information and Modeling	6．最初と最後の頁 2341～2352
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.0c01452	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 該当する

1．著者名 Kumar Naveen、Kaushik Rahul、Tennakoon Chandana、Uversky Vladimir N、Longhi Sonia、Zhang Kam Y J、Bhatia Sandeep	4．巻 1
2．論文標題 Insights into the evolutionary forces that shape the codon usage in the viral genome segments encoding intrinsically disordered protein regions	5．発行年 2021年
3．雑誌名 Briefings in Bioinformatics	6．最初と最後の頁 1-12
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbab145	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 該当する

1．著者名 Yamashita, M., Kuehn, H. S., Okuyama, K., Okada, S., Inoue, Y., Mitsuiki, N., Imai, K., Takagi, M., Kanegane, H., Takeuchi, M., Shimojo, N., Tsumura, M., Padhi, A. K., Zhang, K. Y. J., Boisson, B., Casanova, J.-L., Ohara, O., Rosenzweig, S. D., Taniuchi, I., Morio, T.	4．巻 22
2．論文標題 A variant in human AIOLOS impairs adaptive immunity by interfering with IKAROS	5．発行年 2021年
3．雑誌名 Nature Immunology	6．最初と最後の頁 893～903
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41590-021-00951-z	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 該当する

1．著者名 Kukimoto-Niino Mutsuko、Katsura Kazushige、Kaushik Rahul、Ehara Haruhiko、Yokoyama Takeshi、Uchikubo-Kamo Tomomi、Nakagawa Reiko、Mishima-Tsumagari Chiemi、Yonemochi Mayumi、Ikeda Mariko、Hanada Kazuharu、Zhang Kam Y. J.、Shirouzu Mikako	4．巻 7
2．論文標題 Cryo-EM structure of the human ELMO1-DOCK5-Rac1 complex	5．発行年 2021年
3．雑誌名 Science Advances	6．最初と最後の頁 3147～3147
掲載論文のDOI（デジタルオブジェクト識別子） 10.1126/sciadv.abg3147	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

〔学会発表〕　計0件

〔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