

現代生物學和醫學教育要求學生使用開放的計算機建模工具（在線工具）。使用在線生物信息學工具可以重建蛋白質-蛋白質和基因網絡，培養學生系統生物建模的技能。這需要數學知識、計算機方法以及生物和醫學數據。我們根據與特定疾病相關的基因列表，或基於公共在線生物信息學工具（STRING-DB、GeneMANIA、Metascape、Cytoscape 應用程序）的複雜疾病，考慮了基因網絡的計算機重建問題 - 相互作用的大分子復合物。計算機構建和腫瘤疾病基因網絡可視化的例子——神經膠質瘤、乳腺癌，以及複雜的精神障礙，如帕金森病、精神分裂症——已經在俄羅斯和國際期刊上發表 [1, 2]。

僅使用在線生物信息學工具在本質上是教育性的，主要針對數學、自然科學和醫學學科的學生，他們在計算機科學、編程和編寫自己的代碼方面沒有足夠的技能。使用開放數據庫（OMIM、GeneCards.org、MalaCard.org）自動構建與疾病相關的基因列表、基因網絡的計算機重建、基因本體類別的富集統計計算被用於生物信息學和信息技術教學方法中，發表在一系列論文和科學出版物中。考慮了學生在NSU、FEFU、PFUR、俄羅斯聯邦政府金融大學、謝切諾夫大學教學的基礎上掌握教材的問題。

醫學數字化的任務，信息技術的發展是俄羅斯的優先事項。近年來存在的流行病和向遠程學習的強製過渡加速了改變教育形式的措施的採用，新的學習平台的出現。特別令人感興趣的是與信息學相關的學科本身的計算機化和自動化教學，包括醫學 - 在遠程醫療、電子健康領域。讓我們注意到數字醫療領域的一些定性的新教育任務，例如使用區塊鏈技術、使用人工智能（AI）方法來支持醫療決策 [3]。

文學

1. Dokhoyan A.Yu., Glushchenko M.V., Orlov Yu.L. Reconstruction of the schizophrenia gene network to search for target genes. Ulyanovsk biomedical journal. 2022; 3:6–22. (in Russian)
2. Gubanova N.V., Orlova N.G., Dergilev A.I., et al. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools, Journal of Integrative Bioinformatics, 2021, vol.18, P. 20210031.
3. Koshechkin K.A., Lebedev G.S., Fartushnyi E.N., Orlov Y.L., Holistic Approach for Artificial Intelligence Implementation in Pharmaceutical Products Lifecycle: A Meta-Analysis. Applied Sciences, 2022, vol.12, no.16, P.8373.

DOI 10.18699/sblai2023-26

Использование нейронной сети AlphaFold2 для улучшения результатов белок-белкового докинга

Филиппов С.С.*, Казеннов А.М.

Московский физико-технический институт (национальный исследовательский университет),
Москва, Россия

* stanislav.filippov@phystech.edu

Белок-белковые взаимодействия являются основой многих биологических процессов. Большинство таких взаимодействий определяются трехмерной структурой и динамикой объектов. При этом взаимодействия между белками могут быть как постоянными, так и временными.

Для определения того, как происходят белок-белковые взаимодействия, используются алгоритмы молекулярного докинга. Однако во многих алгоритмах используется rigid-body

подход, который заключается в том, что белки рассматриваются как твердые тела. В этом случае происходит полный перебор всех возможных конфигураций вращений и трансляций объектов друг относительно друга. Недостатком такого алгоритма является то, что он не учитывает структурные изменения, которые могут происходить в комплексе после взаимодействий между белками.

В то же время алгоритмы, в которых белки рассматриваются как подвижные тела, имеют существенно большую вычислительную сложность, что ограничивает их применение на больших массивах данных.

Компромиссом является применение полугибких методов докинга, которые лишены перечисленных недостатков.

В работе описано использование идеи коэволюции, которая применяется в нейронной сети AlphaFold2 для прогнозирования взаимодействия в двухкомпонентных сигнальных белках. Для этого был взят алгоритм описанный в публикации [1], и совмещен с модифицированной моделью AlphaFold2, в которой было уменьшено число гиперпараметров.

Для тестирования итогового алгоритма использовался общедоступный набор данных из бенчмарка Dockground 4 [2]. В нем содержатся 219 пар белков, каждая цепочка которых содержит минимум 50 аминокислотных остатков. Среди рассмотренных комплексов 54 % относятся к эукариотическим белкам, 38 % – к бактериальным и 8 % – это комплексы из пар, в которых один бактериальный белок взаимодействует с эукариотическим.

Полученные результаты показывают, что использование идеи коэволюции в молекулярном докинге позволяет существенно улучшить результаты вычислений и реализовать полугибкий докинг.

Литература

1. Kozakov D. et al. PIPER: an FFT-based protein docking program with pairwise potentials //Proteins: Structure, Function, and Bioinformatics. – 2006. – V. 65. – No. 2. – P. 392-406.
2. Kundrotas P. J. et al. Dockground: a comprehensive data resource for modeling of protein complexes //Protein Science. – 2018. – V. 27. – No. 1. – P. 172-181.

Using the AlphaFold2 neural network to improve the results of protein-protein docking

Filippov S.*, Kazennov A.

Moscow Institute of Physics and Technology (National Research Institute), Moscow, Russia

* stanislav.filippov@phystech.edu

Protein-protein interactions are the basis of many biological processes. Most such interactions are determined by the three-dimensional structure and dynamics of the objects. Moreover, interactions between proteins can be either permanent or temporary.

Molecular docking algorithms are used to determine how protein-protein interactions occur. However, many algorithms use a rigid-body approach, which involves treating proteins as rigid bodies. In this case, all possible configurations of rotations and translations of objects relative to each other are fully explored. The disadvantage of such an algorithm is that it does not take into account structural changes that may occur in the complex after interactions between proteins.

At the same time, algorithms in which proteins are treated as flexible bodies have significantly greater computational complexity, which limits their application to large datasets. A compromise is the use of semi-flexible docking methods, which are free from the listed drawbacks.

This work describes the use of the co-evolution idea, which is applied in the AlphaFold2 neural network to predict interaction in two-component signaling proteins. To do this, the algorithm described in the publication [1] was taken and combined with a modified AlphaFold2 model in which the number of hyperparameters was reduced.

The publicly available dataset from the Dockground 4 benchmark [2] was used to test the final algorithm. The subset contains 219 protein pairs, each chain of which contains a minimum of

50 amino acid residues. Among the complexes considered, 54 % are eukaryotic proteins, 38 % are bacterial, and 8 % are complexes of pairs in which one bacterial protein interacts with eukaryotic.

The results show that the use of the co-evolution idea in molecular docking allows for a significant improvement in computational results and the implementation of semi-flexible docking.

References

1. Kozakov D. et al. PIPER: an FFT-based protein docking program with pairwise potentials //Proteins: Structure, Function, and Bioinformatics. – 2006. – V. 65. – No. 2. – P. 392-406.
2. Kundrotas P. J. et al. Dockground: a comprehensive data resource for modeling of protein complexes //Protein Science. – 2018. – V. 27. – No. 1. – P. 172-181.

使用 AlphaFold2 神經網絡改進蛋白質-蛋白質對接的結果
不銹鋼 菲利波夫

Filippov S.*, Kazennov A.

《莫斯科物理技術學院（國立研究型大學）》

* stanislav.filippov@phystech.edu

蛋白質-蛋白質相互作用是許多生物過程的基礎。

大多數這些相互作用是由物體的三維結構和動力學決定的。

同時，蛋白質之間的相互作用可以是永久的也可以是暫時的。

分子對接算法用於確定蛋白質-蛋白質相互作用如何發生。

然而，許多算法使用剛體方法，將蛋白質視為剛體。

在這種情況下，存在對象相對於彼此的所有可能的旋轉和平移配置的完整枚舉。

該算法的缺點是它沒有考慮蛋白質之間相互作用後復合物中可能發生的結構變化。

同時，將蛋白質視為運動體的算法具有明顯更高的計算複雜度，這限制了它們在大型數據集上的使用。

妥協是使用半柔性對接方法，沒有列出的缺點。

該論文描述了協同進化思想的使用，該思想在 AlphaFold2
神經網絡中用於預測雙組分信號蛋白中的相互作用。

為此，採用了出版物 [1] 中描述的算法並將其與修改後的 AlphaFold2
模型相結合，其中減少了超參數的數量。

為了測試最終算法，使用了來自 Dockground 4 基準 [2] 的公開數據集。
它含有219對蛋白質，每條蛋白質鏈至少含有50個氨基酸殘基。在所考慮的複合物中，54%
屬於真核蛋白，38% 屬於細菌蛋白，8%
是一種細菌蛋白與一種真核蛋白相互作用的成對複合物。

得到的結果表明，在分子對接中使用協同進化的思想，可以顯著提高計算結果，實現
半柔性對接。

文學

1. Kozakov D. et al. PIPER: an FFT-based protein docking program with pairwise potentials //Proteins: Structure, Function, and Bioinformatics. – 2006. – V. 65. – No. 2. – P. 392-406.
2. Kundrotas P. J. et al. Dockground: a comprehensive data resource for modeling of protein complexes //Protein Science. – 2018. – V. 27. – No. 1. – P. 172-181.