蛋白质折叠多样性的进化约束

Harutyun Sahakyan[a]、Eugene V. Koonin[a, *]

a - NIH 美国国家医学图书馆院内研究部计算生物学分部，

贝塞斯达，马里兰州 20894

* - 通讯作者 (koonin@ncbi.nlm.nih.gov)

摘要

蛋白质中二级结构的排列方式产生了多种多样的可能拓扑结构，但自然界中观察到的蛋白质折叠多样性有限。蛋白质折叠的数量是受生物物理约束限制，还是受进化偶然性限制？近年来，

关键词：蛋白质折叠进化；新型折叠；结构进化

1. 引言

尽管可折叠的蛋白质序列数量极其庞大，但自然界中观察到的不同蛋白质架构——即折叠——数量相对较少。如果将蛋白质折叠定义为其核心结构及结构元件的拓扑排列，而不考虑序列，

有人认为，新型蛋白质折叠的发现速度放缓，是因为自然进化已经耗尽了小型蛋白质结构域可能具有的架构，而且大多数折叠早已得到表征（Zhang *et al*., 2006）。然而，另一种观点

随着新型蛋白质设计方法的发展，我们得以更高效地探索蛋白质折叠空间，并从定量层面探讨此类问题（Kuhlman *et al*., 2003）。Minami 等人通过计算设计了所有可能的 αβ 蛋白质

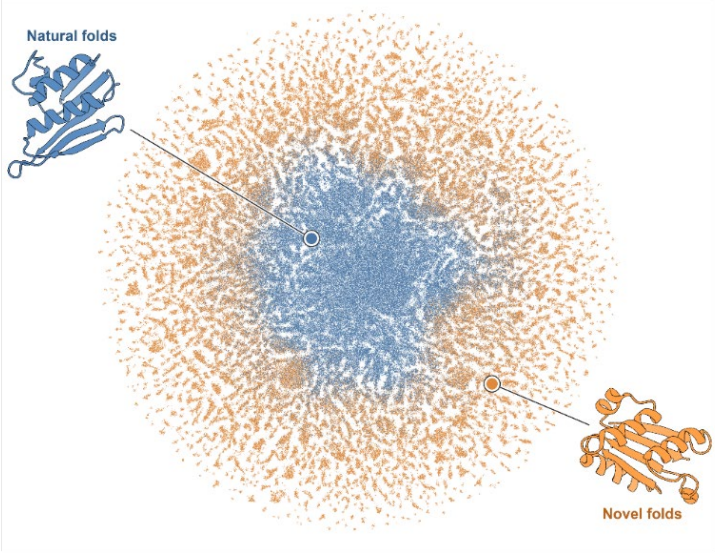

图 1. 蛋白质折叠空间。

在示意图中，每个点代表一种蛋白质。自然界中发现的蛋白质折叠仅占据了这一空间的一部分（蓝点），而许多其他稳定的折叠也有可能存在（橙点）。

2. 《蛋白质》从随机序列进化而来，通常会形成新型折叠

从第一性原理模拟随机序列中的球状蛋白质进化时，通过使用 Foldseek 搜索 PDB 和 AlphaFold数据库（AFDB），发现约有一半进化出的蛋白质与任何已知或建模的结构都不匹配

另一个更根本的问题是，目前没有通用度量标准能够准确判定一种蛋白质是否具有新型折叠。此外，与意味着具有共同起源和特定功能的超家族不同，蛋白质折叠在某种程度上是一个主观

相同折叠蛋白质的相似度阈值（Xu & Zhang, 2010），而不同折叠的蛋白质的 TM-score 可能高于 0.5（Edgar & Sahakyan, 2025）。这是一个本质上很难解决的问题，因为尽管蛋白质折

尽管存在这些注意事项，研究发现，计算机模拟演化出的新型折叠蛋白质，其可折叠性和稳定性与具有自然界中已发现折叠的蛋白质相当（Sahakyan 等，2025）。此外，在最初随机

3. 从头产生的蛋白质带来了新型折叠

理解蛋白质折叠进化的途径，对于解释为何某些折叠拓扑结构在所有三个生命域中都很常见、另一些仅见于范围狭窄的分类谱系，还有一些则从未在自然界中观察到，至关重要。现有

系统发育分析追踪；在其他情况下，变化则更为剧烈且突然，只能通过结构比较方法追溯（Grishin，2001）。还有些情况下，序列变化很小，却能触发剧烈的结构转变，导致折叠切换

一种截然不同的机制涉及从头产生的基因：它们源自此前的非编码序列，通常位于基因间区，并编码小蛋白质；在某些条件下，这些蛋白质能够获得功能并受到净化选择（Fesenko等，

由多种细菌和古菌病毒编码、能够使 CRISPR-Cas 适应性免疫系统失活的抗 CRISPR（Acr）蛋白（Davidson 等，2020；Pawluk 等，2018），通常与任何已知折叠都没有结构相似性（

效应物。尽管结构比较分析发现，部分抗CRISPR蛋白可能存在同源物，其中一些与毒素-抗毒素系统存在不同关联，但许多此类病毒蛋白仍未发现同源物，其起源也依然未知（Davidson

还有其他一些新型蛋白质折叠显然是从头产生的案例，尤其是通过重叠编码产生的案例，即利用预先存在的基因中的替代开放阅读框；这种现象在小型病毒中相对常见，例如微病毒（

4. 新型折叠出现的可能性会随着蛋白质大小的增加而提高

如果蛋白质折叠是从结构空间中随机抽取的，那么与较小的蛋白质相比，较大的蛋白质预计更常采用新型折叠，因为更多二级结构元件可进入更大的组合空间，因此可能的拓扑结构也

而成（Pereira & Lupas, 2022, Fournier等, 2013）。这些折叠的最小重复单元，例如β片层（叶片状结构）中的单个叶片，或α螺旋体中的一个α发夹结构，可以多次独立出现，并在一系

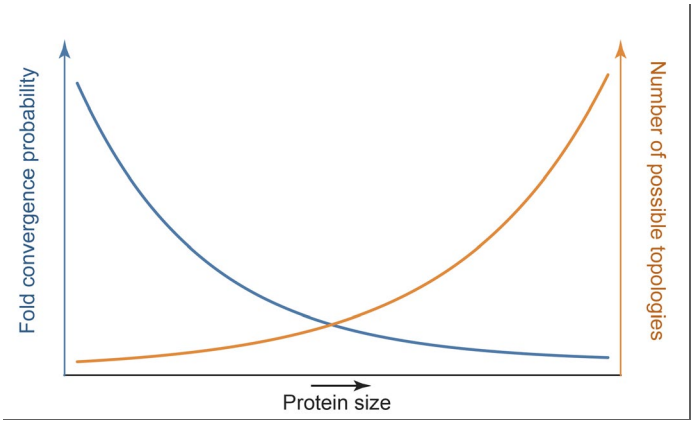

图 2. 蛋白质折叠进化中的趋同现象与蛋白质大小的关系。

趋同进化的概率随蛋白质大小变化而变化；示意图展示了相同蛋白质折叠发生趋同进化的概率。更多的二级结构元件意味着形成蛋白质折叠的可能组合更多。随着可能的折叠数量增加，趋同进化到相同折叠的概率会降低。

5. 新型折叠有用吗?

我们提出了相关论据，支持这样一种观点：所有可能的蛋白质折叠的空间远远——显然是极其——大于自然界中观察到的有限折叠集合，而进化尚未探索这一结构空间的很大一部分。一个引人

然而，这些新型折叠在生物学和/或生物技术上是否具有实用价值。进化给出的直接答案似乎是否定的：将新功能嫁接到现有折叠上似乎更容易，这也解释了为什么同一种折叠常常能够执行

蛋白质设计工具生成的许多蛋白质具有新型折叠，这一事实表明，这些模型学习的是蛋白质结构组织的一般原则，而不只是记忆现有的蛋白质结构。然而，即使是最先进的结构预测模

为训练出更具普适性的蛋白质结构预测和设计模型提供条件。因此，为了从总体上理解蛋白质序列-结构关系，超越自然进化所塑造的范围，我们需要探索那些在生物物理学上合理、但似

6. 结论

与数量 vast 的可能折叠相比，自然界中的蛋白质折叠为何如此之少？其中许多折叠按照现有的任何指标来看，似乎都不逊于生物体中发现的折叠。我们在此提出了多条证据，表明其原因

资助信息

本研究得到美国 NIH（美国国家医学图书馆）内部研究计划的支持。

利益冲突

作者声明不存在利益冲突。

参考文献

Andreeva, A., Kulesha, E., Gough, J. & Murzin, A. G. (2020). 《核酸研究》 **48**, D376– D382. Aubel, M., Buchel, F., Heames, B., Jones, A., Honc, O., Bornberg-Bauer, E. & Hlouchova, K. (2024). 《基因组生物学与进化》 **16**. Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., Pinto, B. L., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G. A., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunic, I., Llinares-Lopez, F., Marchler- Bauer, A., Meng-Papaxanthos, L., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Piovesan, D., Rivoire, C., Sigrist, C. J. A., Thanki, N., Thibaud-Nissen, F., Thomas, P. D., Tosatto, S. C. E., Wu, C. H. & Bateman, A. (2025). 《核酸研究》 **53**, D444– D456. Buchel, F., Neuwirthova, T., Tureckiova, T., Fuertes, G., Benda, A., Panek, D., Fricek, M., AlQuraishi, M. & Hlouchova, K. (2026). *bioRxiv*. Castro, K. M., Watson, J. L., Wang, J., Southern, J., Ayardulabi, R., Georgeon, S., Rosset, S., Baker, D. & Correia, B. E. (2025). 《自然·化学生物学》*2025 22:4* **22**. Chakravarty, D., Porter, L. L., Chakravarty, D. & Porter, L. L. (2026). 《生物物理学年度评论》 **55**. Choi, H., Bauer, M. S., Coventry, B., Venkatesh, P., Chen, A., Kim, D., Bera, A. K., Kang, A., Nguyen, H., Sadre, S., Decarreau, J., Joyce, E., Shankaran, B., Thompson, T. R., Greenstein, G., Didi, K., Schaaf, L. L., Gershon, J. M., Shida, A. F., Lee, G. R., Hilvert, D., Pellock, S. J. & Baker, D. (2026). 《自然》 *2026*. Davidson, A. R., Lu, W. T., Stanley, S. Y., Wang, J., Mejdani, M., Trost, C. N., Hicks, B. T., Lee, J. & Sontheimer, E. J. (2020). 《生物化学年度评论》 **89**, 309–332. Dokholyan, N. V. & Shakhnovich, E. I. (2001). 《分子生物学杂志》 **312**, 289–307. Edgar, R. C. & Sahakyan, H. (2025). *bioRxiv*. Fesenko, I., Sahakyan, H., Dhyani, R., Shabalina, S. A., Storz, G. & Koonin, E. V. (2025). 《分子细胞》 **85**, 1024–1041 e1026. Fournier, D., Palidwor, G. A., Shcherbinin, S., Szengel, A., Schaefer, M. H., Perez-Iratxeta, C. & Andrade-Navarro, M. A. (2013). 《公共科学图书馆·综合》 **8**, e79894. Gabel, C., Li, Z., Zhang, H. & Chang, L. (2021). 《核酸研究》 **49**, 584–594. Grishin, N. (2001). 《结构生物学杂志》 **134**. Kim, I., Koo, J., An, S. Y., Hong, S., Ka, D., Kim, E. H., Bae, E. & Suh, J. Y. (2020). 《核酸研究》 **48**, 9959–9968. Kinch, L. N., Schaeffer, R. D., Kryshtafovych, A. & Grishin, N. V. (2021). 《蛋白质》 **89**, 1618– 1632. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). 《科学》 **302**, 1364–1368. Lau, A. M., Bordin, N., Kandathil, S. M., Sillitoe, I., Waman, V. P., Wells, J., Orengo, C. A. & Jones, D. T. (2024). 《科学》 **386**, eadq4946. Lauko, A., Pellock, S. J., Sumida, K. H., Anishchenko, I., Juergens, D., Ahern, W., Jeung, J., Shida, A. F., Hunt, A., Kalvet, I., Norn, C., Humphreys, I. R., Jamieson, C., Krishna, R., Kipnis, Y., Kang, A., Brackenbrough, E., Bera, A. K., Sankaran, B., Houk, K. N. & Baker, D. (2025). 《科学》 **388**. Makarov, M., Sanchez Rocha, A. C., Krystufek, R., Cherepashuk, I., Dzmitruk, V., Charnavets, T., Faustino, A. M., Lebl, M., Fujishima, K., Fried, S. D. & Hlouchova, K. (2023). 《美国化学会志》 **145**, 5320–5329.

Maxwell, K. L., Garcia, B., Bondy-Denomy, J., Bona, D., Hidalgo-Reyes, Y., Davidson, A. R., Maxwell, K. L., Garcia, B., Bondy-Denomy, J., Bona, D., Hidalgo-Reyes, Y. & Davidson, A. R.（2016）。《自然·通讯》2016 7:1 7。Meeske, A. J., Jia, N., Cassel, A. K., Kozlova, A., Liao, J., Wiedmann, M., Patel, D. J. & Marraffini, L. A.（2020）。《科学》369，54–59。Minami, S., Kobayashi, N., Sugiki, T., Nagashima, T., Fujiwara, T., Tatsumi-Koga, R., Chikenji, G. & Koga, N.（2023）。《自然·结构与分子生物学》2023 30:8 30。Nomburg, J., Doherty, E. E., Price, N., Bellieny-Rabelo, D., Zhu, Y. K. & Doudna, J. A.（2024）。《自然》633，710–717。Orengo, C. A., Jones, D. T. & Thornton, J. M.（1994）。《自然》372，631–634。Pawluk, A., Davidson, A. R. & Maxwell, K. L.（2018）。《自然评论：微生物学》16，12–17。Paysan-Lafosse, T., Andreeva, A., Blum, M., Chuguransky, S. R., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Llinares-Lopez, F., Meng-Papaxanthos, L., Colwell, L. J., Grishin, N. V., Schaeffer, R. D., Clementel, D., Tosatto, S. C. E., Sonnhammer, E., Wood, V. & Bateman, A.（2025）。《核酸研究》53，D523–D534。Pereira, J. & Lupas, A. N.（2022）。《分子生物学前沿》9，895496。Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R. & Karlin, D.（2009）。《病毒学杂志》83，10719–10736。Ren, J., Wang, H., Yang, L., Li, F., Wu, Y., Luo, Z., Chen, Z., Zhang, Y. & Feng, Y.（2022）。《生物化学杂志》298。Sahakyan, H., Babajanyan, S. G., Wolf, Y. I. & Koonin, E. V.（2025）。《美国国家科学院院刊》122，e2509015122。Sahakyan, H., Makarova, K. S. & Koonin, E. V.（2023）。《CRISPR杂志》6。Sakuma, K., Kobayashi, N., Sugiki, T., Nagashima, T., Fujiwara, T., Suzuki, K., Kobayashi, N., Murata, T., Kosugi, T., Tatsumi-Koga, R. & Koga, N.（2024）。《自然·结构与分子生物学》2024 31:2 31。Sato, K. & Tomii, K.（2026）。bioRxiv，2026.2006.2024.732253。Schaeffer, R. D., Roy, B., Guo, R., Cong, Q., Medvedev, K. E. & Grishin, N. V.（2026）。《核酸研究》。Smock, R. G., Yadid, I., Dym, O., Clarke, J. & Tawfik, D. S.（2016）。《细胞》164，476–486。Sun, W., Zhao, X., Wang, J., Yang, X., Cheng, Z., Liu, S., Wang, J., Sheng, G. & Wang, Y.（2023）。《核酸研究》51，1984–1995。Sykes, J., Holland, B. R. & Charleston, M. A.（2023）。《剑桥哲学学会生物学评论》98，243–262。Tassios, E., Pirgelis, N., Rinker, D., Tzermpou, E. M., Hittinger, C. T., Rokas, A., Nikolaou, C. & Vakirlis, N.（2026）。bioRxiv，2026.2008.2029.748005。Waman, V. P., Bordin, N., Lau, A., Kandathil, S., Wells, J., Miller, D., Velankar, S., Jones, D. T., Sillitoe, I. & Orengo, C.（2025）。《核酸研究》53，D348–D355。Wang, H., Gao, T., Zhou, Y., Ren, J., Guo, J., Zeng, J., Xiao, Y., Zhang, Y. & Feng, Y.（2022）。《生物化学杂志》298。Wang, X., Li, X., Ma, Y., He, J., Liu, X., Yu, G., Yin, H. & Zhang, H.（2022）。《核酸研究》50，512–521。Xu, J. & Zhang, Y.（2010）。《生物信息学》26，889–895。Zeldovich, K. B., Chen, P., Shakhnovich, B. E. & Shakhnovich, E. I.（2007）。《公共科学图书馆·计算生物学》3，e139。Zhang, Y., Hubner, I. A., Arakaki, A. K., Shakhnovich, E. & Skolnick, J.（2006）。《美国国家科学院院刊》103，2605–2610。Zheng, I., Learn, B. & Bailey, S.（2026）。《生物化学杂志》483，289–300。