

超大规模 蛋白质折叠稳定性分析

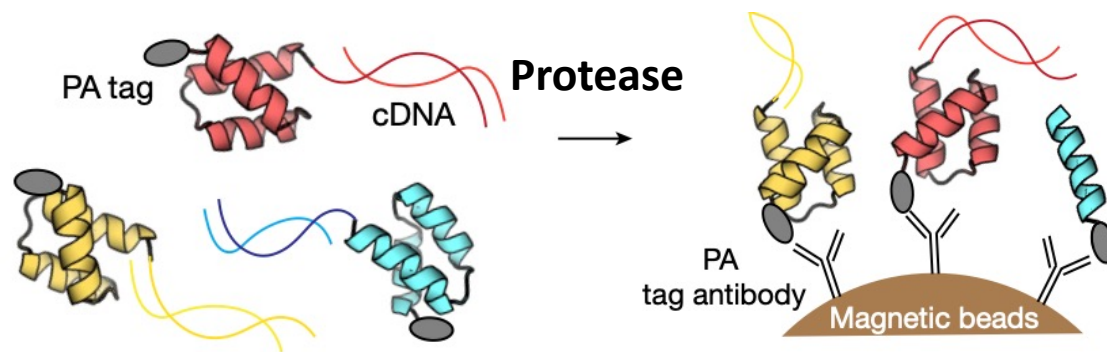
bioinformaticlayer
Papers of the week

Article | [Open Access](#) | [Published: 19 July 2023](#)

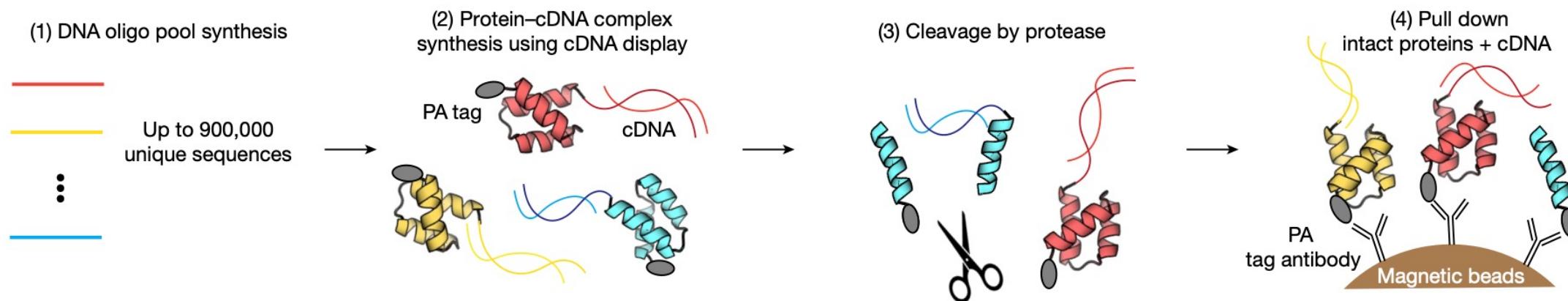
Mega-scale experimental analysis of protein folding stability in biology and design

[Kotaro Tsuboyama](#), [Justas Dauparas](#), [Jonathan Chen](#), [Elodie Laine](#), [Yasser Mohseni Behbahani](#),
[Jonathan J. Weinstein](#), [Niall M. Mangan](#), [Sergey Ovchinnikov](#) & [Gabriel J. Rocklin](#) ✉

[Nature](#) (2023) | [Cite this article](#)

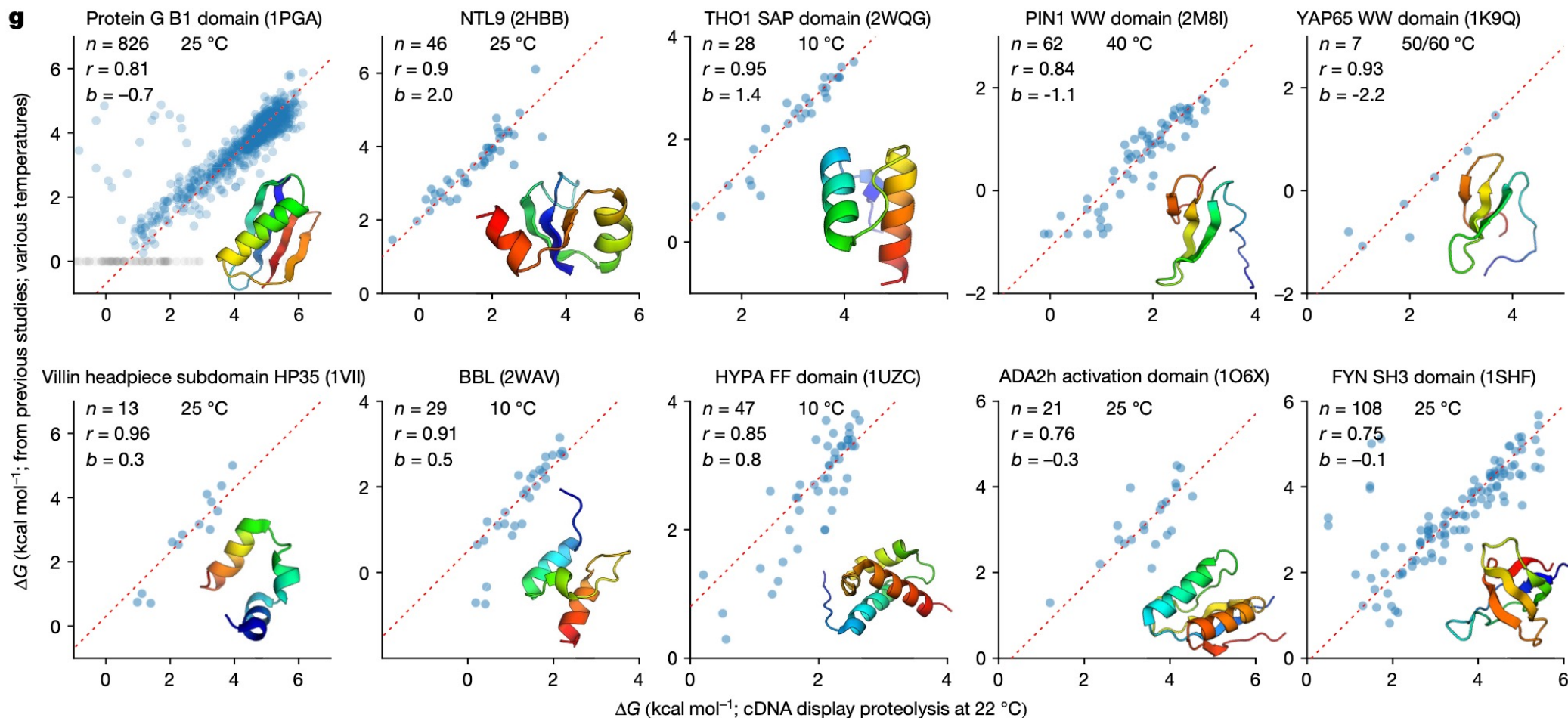


1. 什么是cDNA显示蛋白水解法 (cDNA display proteolysis assay)



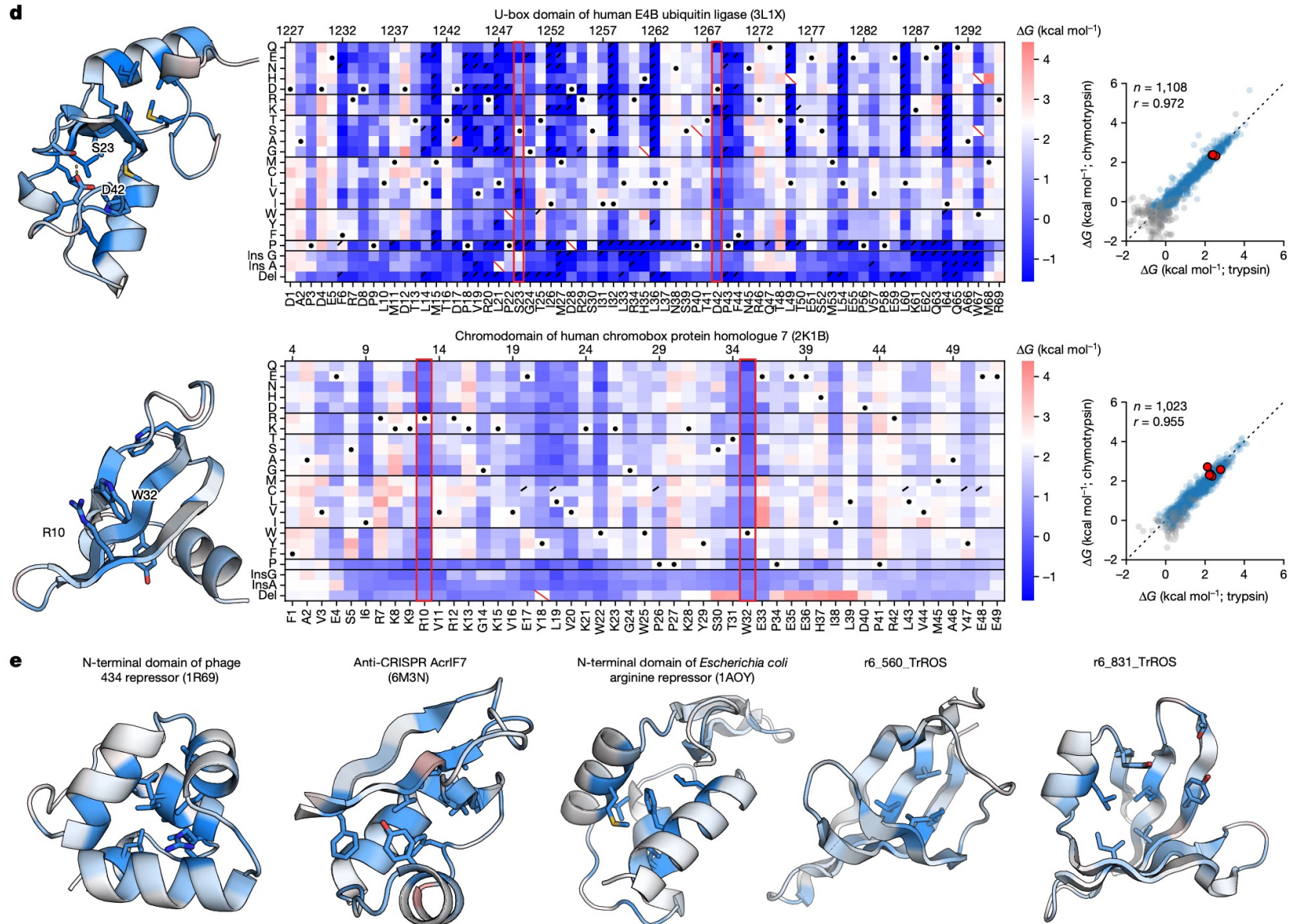
cDNA显示蛋白水解法可以实现一周的实验中测量高达90万个蛋白质结构域热力学折叠稳定性。从总共180万次测量中，该研究筛选出了大约77.6万个高质量的折叠稳定性数据，涵盖了所有单个氨基酸变异体和选定的长度为40-72个氨基酸的331个天然蛋白质结构域和148个新设计蛋白质结构域的双突变体。

2. 高通量数据准确性验证

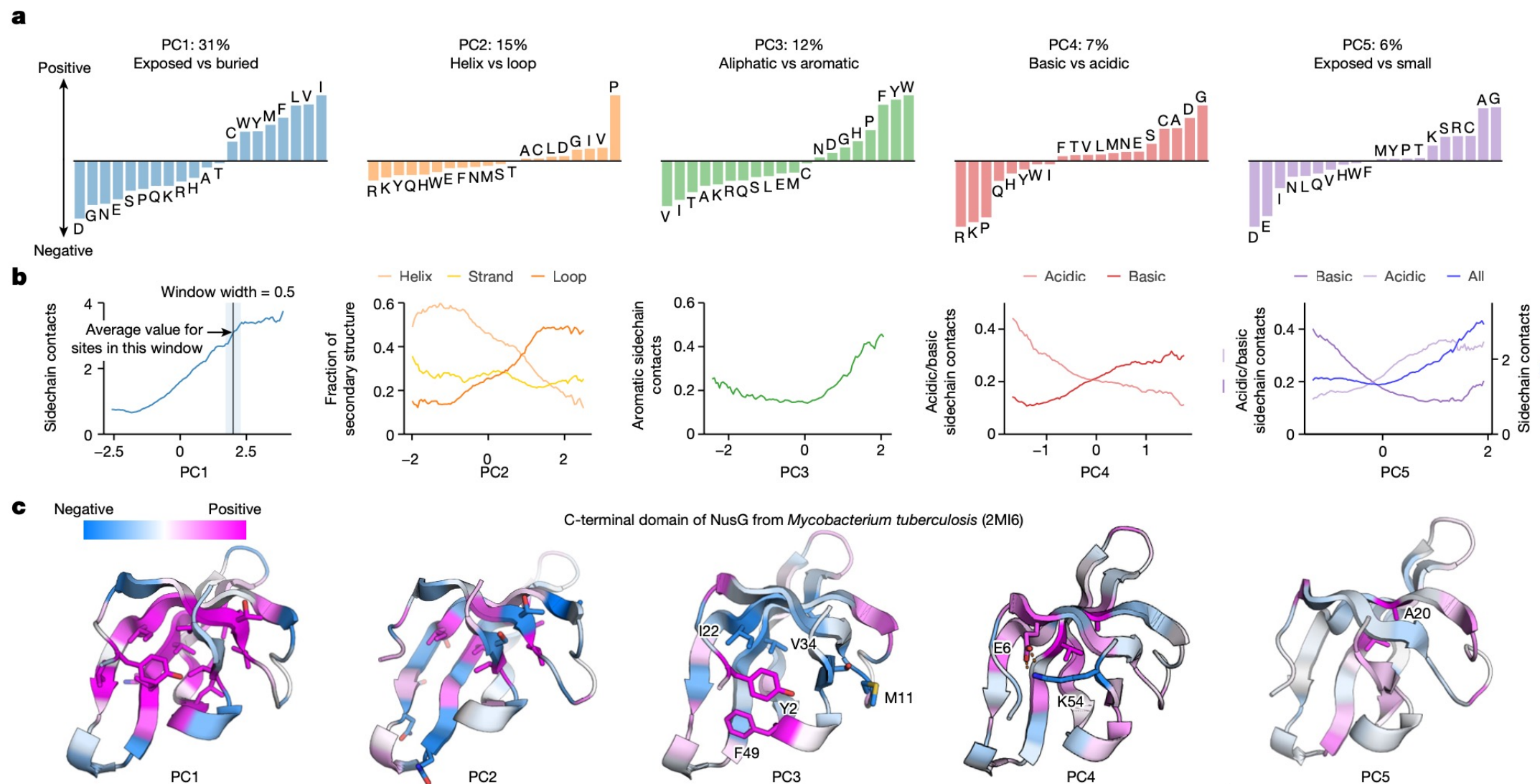


高通量的cDNA展示蛋白水解测量结果与使用纯化蛋白质样本的已发布研究高度一致，涉及10种蛋白质的1188种变体。所有的皮尔逊相关性都在0.75以上。

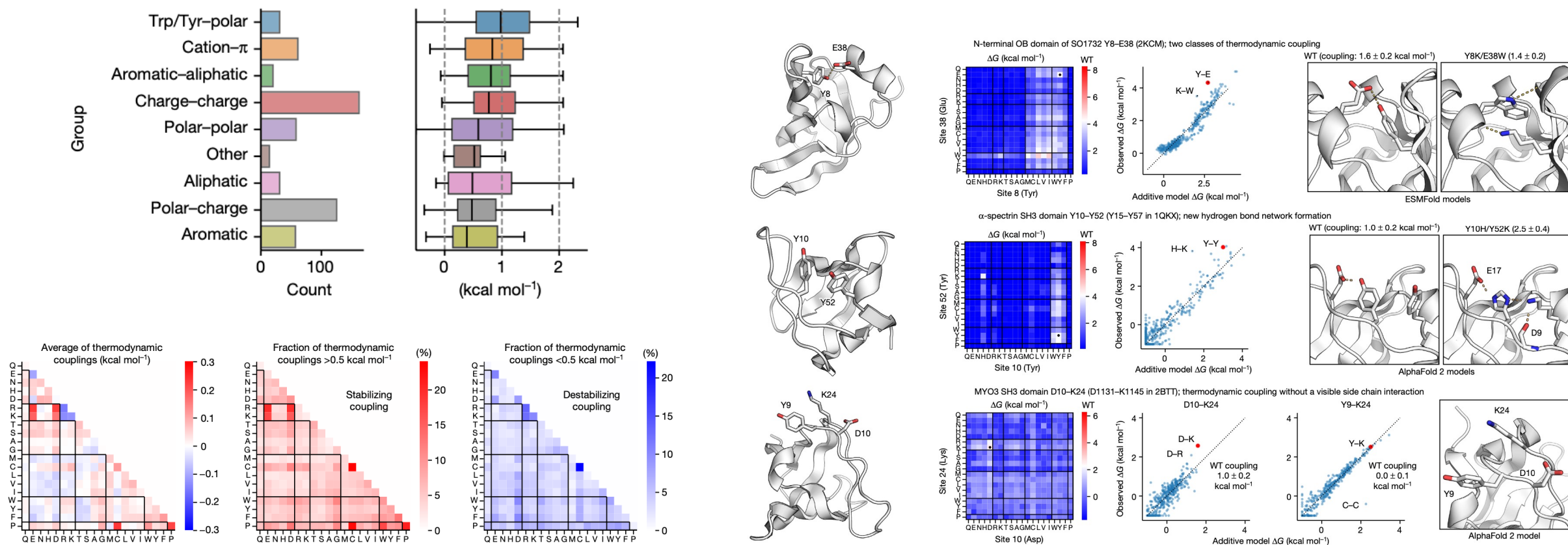
3. 不同蛋白结构域氨基酸突变扫描



4. 氨基酸突变对折叠稳定性影响的总体趋势

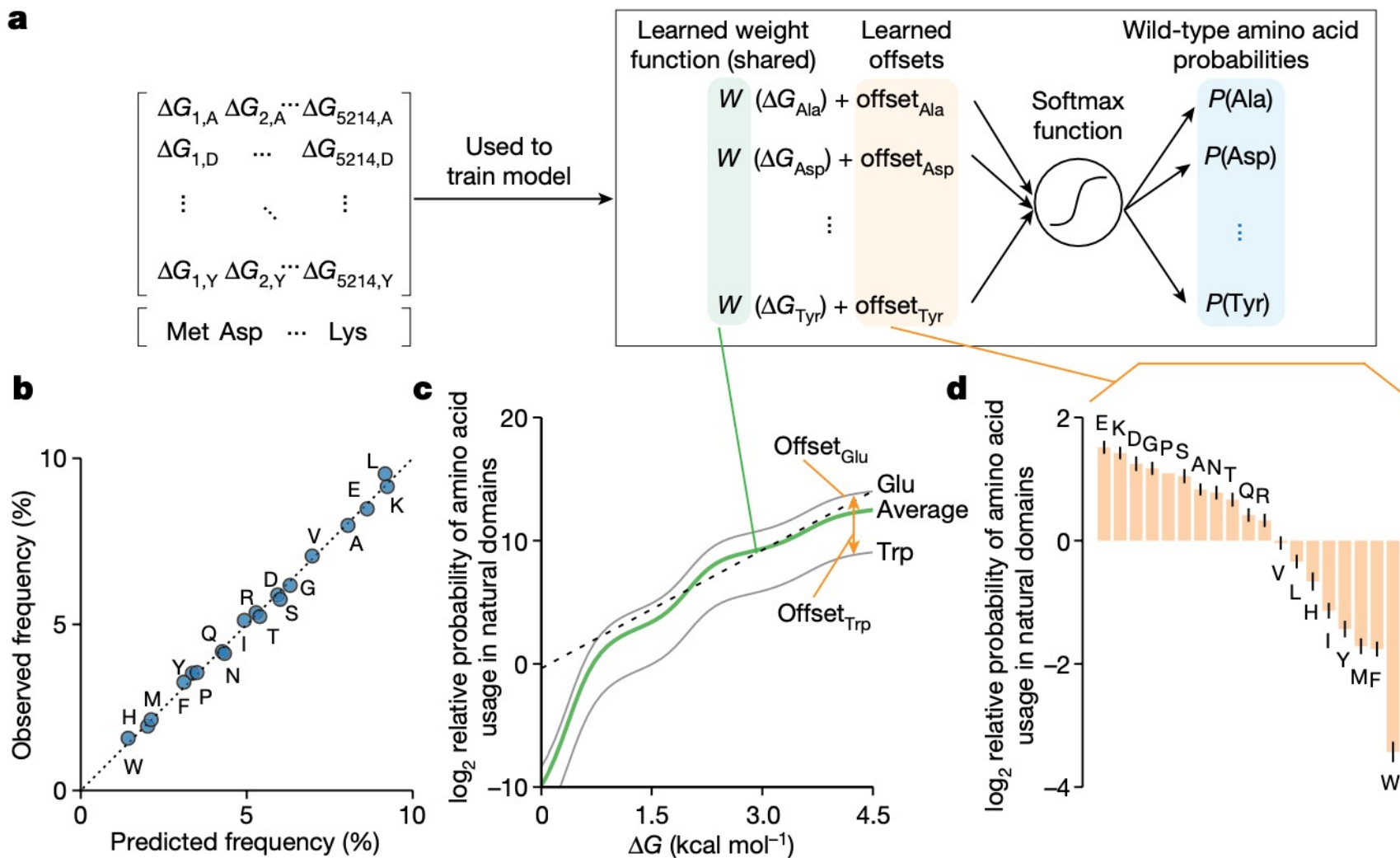


5. 大规模氨基酸耦合的热力学数据

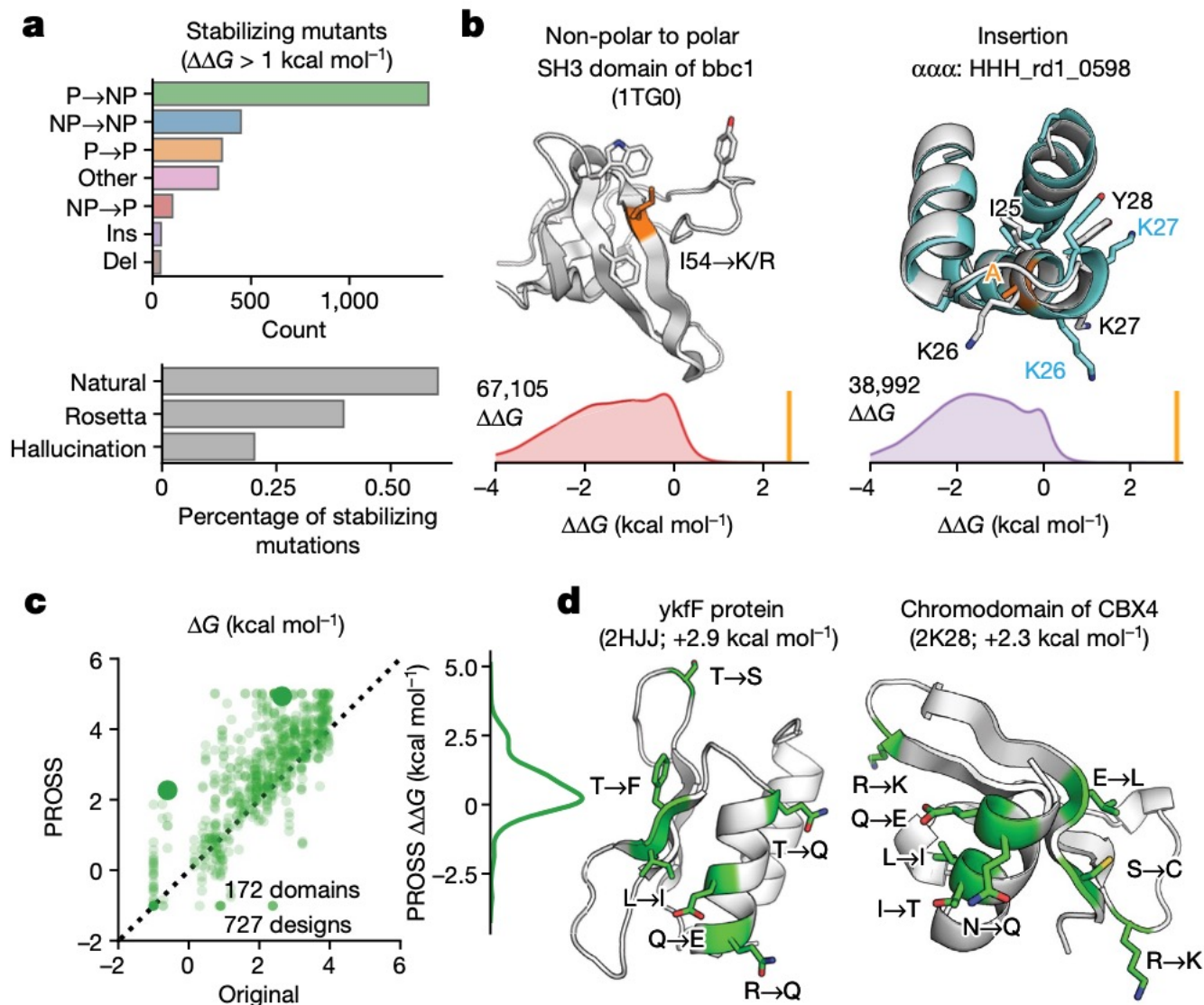


具有相反电荷的对和半胱氨酸的氨基酸对倾向于具有正（有利）的耦合，而具有相同电荷的对和酸性-芳香性-脂肪族对倾向于具有负耦合

6. 蛋白质折叠稳定性对进化的影响



7. 蛋白质折叠稳定性数据指导蛋白质设计



8. 大规模实验数据对生物信息研究的意义

1. 为建立新的数学模型或者机器学习算法提供训练和测试数据，提高预测准确度。

例如：预测单个或者多个氨基酸突变对蛋白质功能的影响 (SIFT, Polyphen, DDgun)

2. 为生物信息分析或者预测结果提供免费的实验验证。

例如：在大规模的分析中发现某个氨基酸突变很重要，可以在这个数据库中去找是不是该突变会对蛋白质折叠稳定性有较大影响

3. 整合分析多个源头的大规模的数据可能产生具有临床或者生物学意义的新发现。

例如：结合高通量基因编辑数据（细胞内实验）和该研究的高通量数据（细胞外实验）可以发现那些对细胞生长或者药物应答重要的突变，再结合临床数据 (Clinvar) 进一步发现那些跟疾病相关的突变