

SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, Iain H Moal
MOAL@EBI.AC.UK

Editor: A detailed contributor list can be found in the appendix of this paper.

Abstract

动机: 了解蛋白质-蛋白质相互作用的序列、结构、结合能、结合动力学和结合热力学之间的关系对于理解细胞信号传导、分子复合物的组装和调节、突变导致疾病的机制以及蛋白质工程至关重要。

结果: 我们推出了 SKEMPI 2.0, 这是对我们数据库的重大更新, 该数据库包含结构解析的蛋白质-蛋白质相互作用突变后结合自由能的变化。此版本现在包含 7085 个突变的手动整理的结合数据, 增加了 133%, 包括 1844 个突变的动力学变化、443 个突变的焓和熵变化以及 440 个突变, 这些突变消除了可检测的结合。

Availability: The database is available at: <https://life.bsc.es/pid/skempi2/>

Date: Oct 30, 2024

Copyright of Justina Jankauskaitė and Brian Jiménez-García

1 Introduction

蛋白质-蛋白质相互作用是几乎所有生物过程的核心，从细胞信号转导和肌丝等中观结构的组装，到病毒粘附和免疫反应。因此，蛋白质序列变化对蛋白质-蛋白质相互作用的结构、热力学和动力学的影响对于限制进化过程中允许的替换以及了解疾病的分子病因具有广泛的意义。测量、预测或优化这些变化的方法可用于设计从头相互作用 [20](#)，增强生物治疗的特异性和亲和力（例如 [3](#)），设计组合蛋白质库（例如 [24](#)），揭示病理突变的影响（例如 [??](#)），定位可用药结合位点（例如 [??](#)）和药物设计的结合热点 [25](#)，改变结合动力学 [??](#)，蛋白质-蛋白质对接（例如 [19](#)）以及表征过渡态（例如 [??](#)）、结合途径 [61](#) 和序列亲和力景观 [2](#)。

SKEMPI 是一个手动整理的数据库，其中包含结构表征的蛋白质-蛋白质相互作用突变以及这些突变对结合亲和力和其他参数的影响 [47](#)。第一个版本已被用作许多进一步研究的基础，包括开发能量函数 [??](#)，随后在 CCharPPI 网络服务器中实现，用于表征蛋白质 - 蛋白质相互作用 [49](#)，以及用于对接姿势进行排名 [??](#)。SKEMPI 还被用于研究人类疾病 [??](#)、评估动力学对结合的作用 [??](#)、探索结合区域的保守性 [28](#)、评估实验亲和力测量方法 [22](#)，以及作为预测突变 [1](#)、病理突变 [23](#)、热点残基（例如 [??](#)）和结合能变化（例如 [??](#)）后解离速率变化的模型的数据源。

这里，我们根据数据库中突变的数量和所包含的不同系统的数量对基准进行了重大更新（表 [??](#)）。我们现在还根据 [22](#) 的类别，为所有条目提供了实验方法的详细信息，以及消除了可检测结合或只能确定野生型或突变体亲和力上限或下限的突变。

表 1: Comparison with Previous Version		
	SKEMPI 1.1	SKEMPI 2.0
Entries	3047	7085
Unique entries	2792	6187
k_{on} and k_{off}	713	1844
ΔH and ΔS	127	443
Inequalities / no binding	0	440
Number of interactions	87	237
Number of PDB entries	158	345
Number of papers cited	66	295

2 方法

2.1 数据来源

超过五分之二的数据来自 SKEMPI 的前一个版本 [47](#)，主要由文献来源中的数据构成，这些数据在某些情况下在进行结构亲和性基准测试的数据收集时引起了作者的注意 [33](#)，以及相关参考文献中。一些 SKEMPI 1.1 中的条目是通过检查 ASEdb [??](#) 和 PINT [36](#) 数据库中的参考文献发现的，尽管并非所有数据都通过了结构亲和性基准测试所要求的检查 [??](#)。在数据收集期间，发布了三个其他相关数据库：ABbind [??](#)、PROXiMATE [31](#) 和 dbMPIKT [41](#)。如果它们尚未包含在内，则会检查其参考文献。这些来源的数据分别占 SKEMPI 2.0 的 4%、3% 和 6%。与 ASEdb 和 PINT 一样，所有数据均未直接复制到 SKEMPI 中。此外，引用的论文经过阅读，数据的输入遵循与其他条目相同的检查和程序。

2.2 数据收集

每个条目都在文献中找到并手动审查。为了确保数据质量，采用了一系列严格的检查。首先，我们确保结构和报告亲和力的论文引用的是相同物种中的相同蛋白质，并且结构和亲和力数据在辅因子、附属链和翻译后修饰方面相匹配。例如，我们区分 Ras GTP 和 RasGppNHp，因为核苷酸调节了 Ras 与其效应器的亲和力。在未使用全长蛋白的情况下，我们检查了晶体结构中的片段是否与报告亲和力的片段相匹配。

在通过检查后，数据被收集，包括 PDB 文件、相互作用亚基的链、突变、野生型和突变亲和力 (K_D , M)、参考文献、蛋白质名称、实验进行的温度 (T , K)、使用的实验方法 (22 的分类方案的扩展)、条目的备注，以及在可用时的结合速率 (k_{on} , $M^{-1}s^{-1}$)、解离速率 (k_{off} , s^{-1})、焓 (ΔH , $kcal \cdot mol^{-1}$) 和熵 (ΔS , $cal \cdot mol^{-1} \cdot K^{-1}$)。对于多个 PDB 条目可用的情况，选择分辨率较高的结构。当亲和力或动力学或热力学参数以不同单位报告时，将其转换为上述指定的单位。在某些情况下，当未直接报告时， K_D 、 k_{on} 、 k_{off} 、 ΔH 和 ΔS 是通过关系 $\Delta G = \Delta H - T\Delta S = RT \ln(K_D)$ 和 $K_D = 1/K_A = k_{off}/k_{on}$ 计算得出的。

为了确保一致性，SKEMPI 中的残基编号与 PDB 文件中报告的编号相同。因此，编号通常相对于引用的论文发生变化或调整，例如，如果抗体的晶体结构按照 Kabat 编号方案报告，但突变数据没有，则在输入到 SKEMPI 之前会转换突变数据。对于所有条目，“野生型”列中报告的亲和力对应于 PDB 文件中的亲和力，而“突变”列中的亲和力是应用了指定突变后的 PDB 文件中的蛋白质亲和力。因此，在 PDB 报告突变形式的情况下，如果该条目对应于返回野生型的反向突变，则前者的亲和力出现在“野生型”列中，而后的亲和力出现在“突变”列中。这些情况在数据库中有所记录。

除了检查新条目外，我们还重新评估了 SKEMPI 1.1 中引用的论文，以确保数据的完整性和准确性。

2.3 后处理和注释

除了上述数据，SKEMPI 2.0 还提供了突变残基的位置、数据集中相互作用的同源性以及可轻松解析的处理后 PDB 文件。

残基位置： 每个突变残基根据 38 提出的方案进行分类；界面上的残基被分类为支持（未结合时大部分被埋藏，结合后完全被埋藏）、核心（未结合时大部分暴露在溶剂中，但结合后被埋藏）和边缘（结合时部分被埋藏），而远离结合位点的残基则被分类为内部或表面。溶剂暴露的表面积使用 CCP4 4.7 计算。

处理后 PDB 文件： 从蛋白质数据银行 8 下载的相互作用 PDB 文件通常包含单位胞中相互作用蛋白的多个副本或与相互作用无关的其他链。在一种情况下，二聚体肌肉生长因子与类促卵泡激素样 3 的结合，必须通过对单位胞进行平铺来创建肌肉生长因子二聚体。此外，一些 PDB 文件包含某些软件无法轻易解析的特征，例如残基插入代码或负残基编号。为了帮助用户，我们提供了“清理过的”PDB 文件，这些文件仅包含相关链，从一开始重新编号，以及与任何相关链的非氢原子在 5 Å 范围内的水和其他分子。因此，每个突变都以 PDB 编号和重新编号的形式报告。

定义同源相互作用： 每个条目还指定了哪些其他条目是同源相互作用的突变。如果两个相互作用具有共享的结合伙伴或同源结合伙伴，并且至少 70% 的对应界面残基在两个相互作用中是共同的，则认为这两个相互作用是同源的。我们使用 GAP4 程序 29 确定蛋白质之间的同源性，并将同源蛋白定义为相似度评分大于 50 且序列相似度至少为 30% 的蛋白质。界面残基被定义为在与结合伙伴的非氢原子距离 10 Å 范围内的非氢原子。落在手动分配的同源相互作用聚类中的相互作用被指定为 pMHC/TCR、抗体/抗原或蛋白酶/抑制剂。虽然这些聚类的名称是为了反映其组成相互作用的主要功能而选择的，但它们反映了数据集中的同源性，而不是功能分配。因此，例如，一些抗体被分类为抗体，而其他与西妥昔单抗结合到同一位点，14.3.d 被归类为 TCR，尽管它只是一链，而其结合伙伴肠毒素 C3 被归类为 pMHC。

3 结果与讨论

3.1 SKEMPI 中的多样性、偏差和相互关系

总共收集了 7085 条条目，如表 I 和图 1A 所示。这些数据来源于文献，因此，尽管涵盖了广泛的残基、蛋白质、相互作用和系统，但偏向于研究社区的兴趣和能力。这些偏差在数据库的组成中显而易见，具体取决于图 1 中所示的参数。 $\Delta\Delta G$ 值范围广泛，但大多数落在 -3 到 7 kcal.mol^{-1} 之间（图 1B），这既有生物物理原因，也有技术原因（见第 III-C 节）。几乎四分之三的数据对应于单点突变，其中超过一半是对丙氨酸的突变（图 1C）。电荷交换突变和芳香族残基之间的突变也被过度表示。大多数单点突变位于结合位点，大多数突变位于界面的核心。同样，大多数双突变也位于结合位点，并且其中大多数位于核心。测量结合亲和力的最常用方法是表面等离子共振和荧光光谱方法。对于特定相互作用和相互作用类别也存在较大的偏差，例如早期对蛋白酶抑制和免疫相互作用（如抗体-抗原复合物、T 细胞受体识别呈现在细胞表面的肽、细胞因子信号传导和补体系统）的研究。实际上，几乎一半的数据仅对应于蛋白酶-抑制剂、抗体-抗原和 pMHC/TCR 相互作用。虽然这些类别中的许多相互作用共享共同的结合位点或同源结合位点（图 2），但这些组之间也存在联系，例如通过结合蛋白酶活性位点的抑制抗体，或由于免疫球蛋白超家族中蛋白质的共同结合区域，如抗体、TCR、MHC 和 β -2 微球蛋白。数据中还存在较小的共享和同源相互作用的聚类，例如 Ras-效应物聚类。这些关系在数据库中有所记录，并可能在开发模型时避免过拟合，或用于验证和估计泛化误差，如前所述 47。

条目在结构有序程度上也有所不同。虽然大多数条目对应于折叠域之间的相互作用，但数据库中也包含在结合时发生构象变化的条目，例如蛋白-肽相互作用，以及在极端情况下，ACTR/NCBD 相互作用，其中两个结合伙伴在结合时变得有序 32。实际上，数据集要求具有结构以便被纳入，*ipso facto* 偏向于这些数据，这意味着“模糊”复合物（在平衡状态下保持扩散结构）没有代表性。这种相互作用形成了可构象或可解析的晶体。

除 pMHC-TCR、抗体-抗原和蛋白酶-抑制剂类别外，大多数相互作用是内源性的。尽管如此，集合中也包括外源性相互作用，从同一物种中不同个体之间的蛋白质相互作用，即人类精子和卵子中的性融合蛋白 Juno 和 Izumo1 4，到宿主-病原体相互作用，例如腺病毒和冠状病毒在病毒进入过程中与人类受体的相互作用 65, 27，再到蛇毒神经毒素 fasciculin 对乙酰胆碱酯酶的抑制 2。对于 pMHC-TCR 相互作用，存在多种呈递抗原，包括外源性病毒肽和谷蛋白，以及内源性自身免疫和癌症肽。抗体相互作用包括病原体抗体，以及针对细胞外治疗靶点的抗体。蛋白酶相互作用大多是外源性的，源于它们固有的交叉反应性，这是由于其经典抑制环的趋同进化。

3.2 组成 SKEMPI 2.0 的显著研究

SKEMPI 数据来源的研究是多样的，涵盖了许多生物过程，共报告在 295 篇出版物中，包括系统扫描、丙氨酸和同源扫描、设计研究（包括计算设计和源自噬菌体展示的设计）、双突变循环研究、抗体工程、生物药物设计和病理突变的评估。最大贡献来自已故的 Michael Laskowski Jr. 研究组，他们系统地研究了火鸡卵白素第三域在选定位点的所有可能突变及其与四种蛋白酶的抑制相互作用 43，以及其他鸟类物种中同一领域的相互作用和超高亲和力广谱抑制剂的设计。来自 Gideon Schreiber（例如 62）和 Timothy Palzkill（例如 10）研究组对 A 类 β -内酰胺酶抑制相互作用的研究也提供了大量数据，此外还有细胞因子受体相互作用的研究，特别是来自 Schreiber 研究组（例如 63）和 K. Christopher Garcia 研究组（例如 ??）对 I 型干扰素的研究，以及来自 Michael W. Parker 研究组对 GM-CSF / GM α 相互作用的研究 9。其他显著的数据来源是对激素受体相互作用的研究，特别是来自 Jim Wells 研究组（例如 14）的人生长激素受体和来自 Michael E. Hodsdon 研究组（例如 35）的催乳激素受体的研究，以及对抗原识别的研究，包括来自 Herman van Vlijmen 研究组（例如 12）的综合计算和实验设计研究，以增强 AQC2 抗体对整合素 α -1 的亲和力，来自 Richard G. Fensler 研究组对广泛中和抗体靶向 HIV gp120 的相互作用的分析 11，以及来自 Roy A. Mariuz 研究组的各种研究。

3.3 范围和误差

范围：突变后结合自由能的变化范围从 -12.4 到 $12.4 \text{ kcal.mol}^{-1}$ ，与 SKEMPI 1.1 中一样，其中 $\Delta \log_{10} k_{\text{on}}$ 范围从 -3.6 到 2.4 ， $\Delta \log_{10} k_{\text{off}}$ 范围从 -6.0 到 6.8 ， $\Delta \Delta H$ 范围从 -18.3 到 $26.5 \text{ kcal.mol}^{-1}$ ， $\Delta \Delta S$ 范围从 -61 到 80 cal.mol^{-1} 。约 60 个突变体非常不稳定，使结合能降低 8 kcal.mol^{-1} 或更多。这些都存在于酶/抑制剂复合物中，例如蛇毒束蛋白对乙酰胆碱酯酶的抑制，或对酶的抑制，如果它们解离并在错误的位置变得活跃，则会产生有害影响，例如核酸酶（barnase/barstar、大肠杆菌素 E9 DNase/Im9、RNase A/血管生成素）和蛋白酶（如胰蛋白酶/BPTI）。这些相互作用往往在皮摩尔亲和力左右，并且由于测量时间限制，处于可检测到的上限。

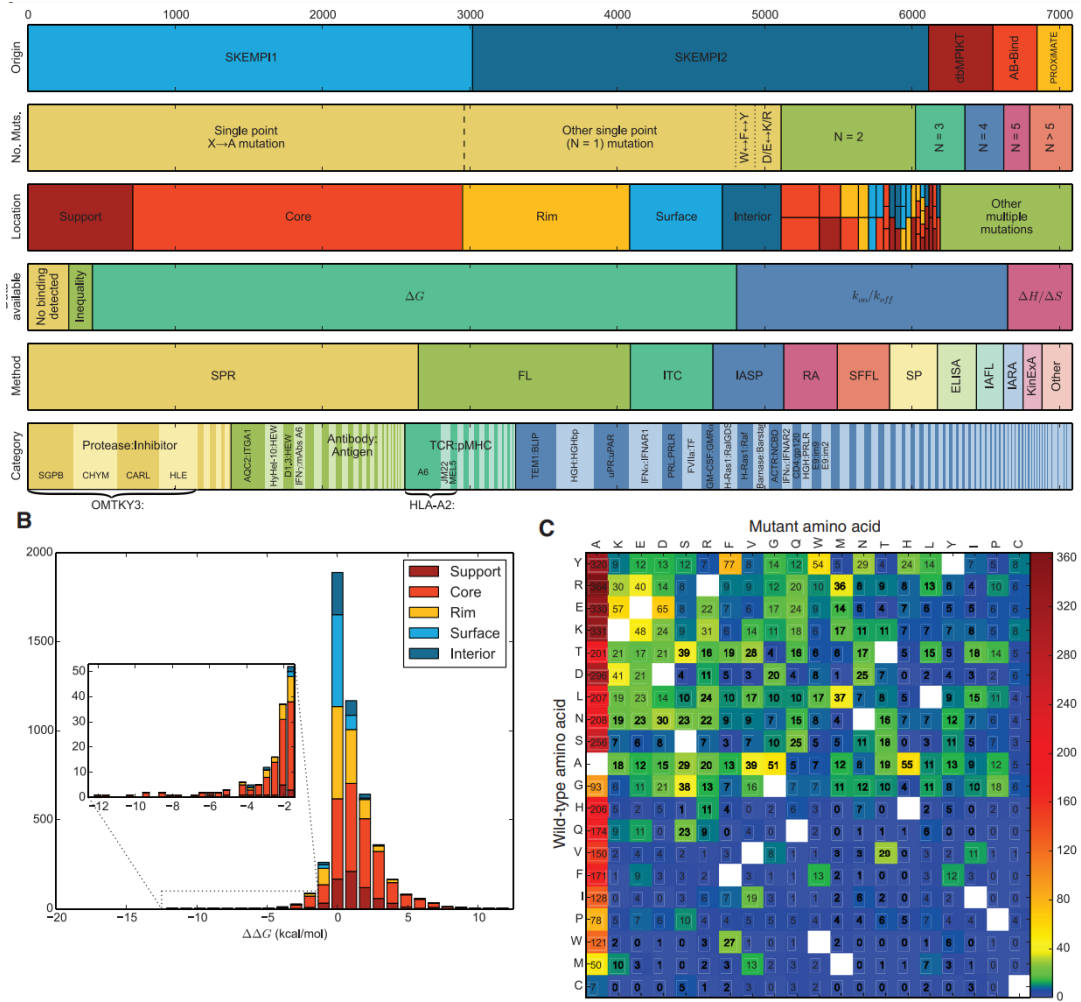


图 1: SKEMPI 2.0 概述。(A) 根据突变的来源、改变的残基数量、复合物中的位置、额外的动力学和热力学数据的可用性、所用的实验方法和类别对突变进行划分。(B) $\Delta \Delta G$ 的分布。(C) 单点突变的源氨基酸和目标氨基酸。

所需的平衡状态和质量作用定律所需的低浓度，以探测结合曲线的信息区域。这些非常去稳定的突变将亲和力降低到微摩尔范围，接近使用标准方法可量化的下限。由于突变和野生型亲和力都接近检测阈值，这些条目的误差通常较大。此外，尽管某些突变可能导致亲和力变化超过七个数量级，但在基准测试中缺少此类突变的亲和力可以用以下事实解释：这些突变会涉及超出上限或下限的亲和力。实际上，存在一些新条目，其中单一或双重替换将结合力从数十皮摩尔降低到没有可检测的结合。对于许多高度去稳定的突变，已经解决了突变的晶体结构，而在数据库中 30

个最稳定的突变 ($\Delta\Delta G < -3 \text{ kcal.mol}^{-1}$) 大多数是应用于这些结构的反向突变。这些大多数是单突变或双突变, 但包括对非同源 Colicin E2/Im9 复合物的多达 27 个残基的突变, 使其在序列空间中向同源的 E9/Im9 移动 40。

误差: K_D 的标准误差通常报告为 50% 的量级, 约为 $0.25 \text{ kcal.mol}^{-1}$ 。这些估计是通过在相同设备、环境和协议下进行重复测量得出的, 因此不包括源于系统偏差的误差。然而, SKEMPI 已通过不同组或使用不同技术评估相同突变的条目对进行估算。在 1741 对这样的条目中, 有 84% 的条目给出的 $\Delta\Delta G$ 值相差不超过 1 kcal.mol^{-1} 。对于 704 对同时可用 k_{on} 的条目, 80% 的 $\Delta\log_{10}(k_{\text{on}})$ 相差不超过 0.5。对于 702 对可用 k_{off} 的条目, 83% 的 $\Delta\log_{10}(k_{\text{off}})$ 相差不超过 0.5。在 62 对同时具有 ΔH 和 ΔS 值的情况下, 61% 的 ΔH 相差不超过 $3.0 \text{ kcal.mol}^{-1}$, 58% 的 ΔS 相差不超过 $10 \text{ cal.mol}^{-1} \text{ K}^{-1}$ 。

3.4 突变循环

在 SKEMPI 中, 一些条目可以结合起来构建突变循环, 以量化残基之间的相互作用、这些相互作用对其他残基的依赖性以及其他高阶效应。最常见的情况是双突变循环, 其中野生型、A、B 和 AB 突变的亲和力可用, 存在 610 个示例。其中 53 个涉及至少一个未观察到结合的突变, 或者仅提供不等式, 而 235 个涉及在同一参考文献中报告的突变, 因此这些亲和力很可能是使用相同技术和条件测量的。还可以在第三突变的背景下构建 218 个双突变循环 (即可用 C、AC、BC 和 ABC 突变), 其中 209 个不由不结合突变或带有不等式的突变组成, 131 个涉及来自同一参考文献的亲和力。在 766 个不包含 $\Delta\Delta G_{A \rightarrow AB}$ 或 $\Delta\Delta G_{B \rightarrow AB}$ 的双突变循环中, 因一个突变而导致的结合自由能变化以相反的方式报告, 以及 $\Delta\Delta G_{A \rightarrow B \rightarrow AB}$ 和 $\Delta\Delta G_{B \rightarrow AB}$, 即在另一突变的背景下单一突变的能量, 以及 $\Delta\Delta G_{\text{int}} = \Delta\Delta G_{A \rightarrow B \rightarrow AB} - \Delta\Delta G_{B \rightarrow AB} = \Delta\Delta G_{A \rightarrow B}$, 两个突变的相互作用能量 26。由此可推断, 345 个是加性的 ($\Delta\Delta G_{\text{int}} < 0.5 \text{ kcal.mol}^{-1}$)。在非加性循环中, 293 个在双突变中的结合力比单突变的总和更紧密 (正表观基因互作), 其中 6 个导致的结合力甚至比加强相互作用的两个单突变的个体效应更紧密 (协同正效应), 而 273 个对应于双突变, 其降低结合力的程度小于两个降低结合力的单突变的总和 (对抗性正效应)。同样, 128 个循环的双突变表现出比两个单突变的总和更弱的结合力 (负表观基因互作), 其中 58 个包含两个去稳定的单突变 (协同负效应), 26 个包含两个稳定的单突变 (对抗性负效应)。 $\Delta\Delta G_{\text{int}}$ 值的范围很少超过 -5 到 3 kcal.mol^{-1} 的范围。在 421 个非加性循环中, 151 个显示出明显的符号表观基因互作, 即 A 或 B 突变的效应符号会根据背景突变的存在或缺失而翻转 (即, 对于 A 突变, $\Delta\Delta G_{A \rightarrow B} > 0.2 \text{ kcal.mol}^{-1}$ 和 $\Delta\Delta G_{A \rightarrow B \rightarrow AB} < 0.2 \text{ kcal.mol}^{-1}$, 或反之)。

可以通过更高阶循环获得更高阶相互作用项, 例如三重突变立方体, 这些立方体由野生型、三个单突变、三个相应的双突变和三重突变的能量构成 26。在 SKEMPI 中, 可以制作 45 个三重突变立方体, 其中 10 个来自同一参考文献。对于这些, 三阶相互作用能量的范围在 -1 到 1 kcal.mol^{-1} 之间。对于四阶相互作用, 由以下能量构成: 野生型、四个单突变、六个双突变、四个三重突变和一个四重突变, 在 SKEMPI 中存在 14 个示例。然而, 在解释四阶残基耦合能量时应谨慎, 因为误差的累积在这些情况下会加剧, 尤其是由于亲和力在不同的出版物中报告。没有五阶或更高阶的相互作用。

3.5 SKEMPI 网站

该数据库可在线访问, 网址为 <https://life.bsc.es/pid/skempi2/>, 可以下载包含所有数据的原始 CSV (逗号分隔值) 文件。数据也可以在线浏览、排序和搜索, 用户可以根据任何字段 (如实验方法或突变位置) 进行搜索, 或通过名称或 PDB 代码搜索特定蛋白质, 并可视化结构。网站上的其他页面提供数据摘要、常见问题解答和帮助页面, 以及一个用户贡献页面, 这些贡献将被评估, 以便在未来版本中出现。在开始格式化论文之前, 首先将内容写入并保存为单独的文本文件。请将文本和图形文件分开, 直到文本格式化和样式化之后。不要使用硬制表符, 并将硬回车的使用限制为段落末尾的一个回车。不要在论文的任何地方添加分页。不需要给文本标题编号——模板会为您处理这个。

Reference

1. R. Agius, M. Torchala, I. H. Moal, J. Fernandez-Recio, and P. A. Bates. Characterizing changes in the rate of protein-protein dissociation upon interface mutation using hotspot energy and organization. *PLoS Comput Biol*, 9(10):e1003354, 2013.
2. Y. Azia, O. Levy, J. Shamir, G. R. Dadwar, M. Risman, O. Avraham, and J. Shifman. Mapping of the binding landscape for a picomolar protein-protein complex through combinatorial sequence scanning. *Structure*, 22(4):561–571, Apr 2014.
3. V. Arakcheeva, G. Yosef, J. Shifman, T. Cohen, Y. Horesh, M. Grossman, I. Sagi, J. Wachtel, and A. M. Frenkel. Structural and Material Mechanics. *J Biol Chem*, 292(8):3481–3495, Feb 2017.
4. H. Aydin, A. Sultana, S. L. A. Thavalingam, and J. E. Lee. Molecular determinants of human paramyxovirus entry and fusion receptor complex. *Nature*, 534(7608):562–565, 2016.
5. K. A. Barrow, S. C. Tokuhisa, S. Thompson, S. F. Juras, E. J. Lea, B. H. Moont, and C. Kremer. Binding Free Energy Estimation of Antibody Mutants Using Relative Free Energy Perturbation Calculations. *J Mol Biol*, 398(1):81–101, 2018.
6. M. Barlow and P. Amigues. Crystallization and Structural Analysis of Acta Structural Proteins. *Proc Natl Acad Sci USA*, 118(41):10411–10419, 2014.
7. N. Berliner, T. Jogler, S. Garc, and P. Menet. Mapping of Biophysical Interactions at the Protein Surface. *J Mol Biol*, 104(1):119–132, 2019.
8. H. M. Berger, N. Shevtsov, A. Rego, and C. Guillet. Comparative Study of Binding Free Energies of the Protein Data Bank. *Nat Struct Biol*, 23:195–202, 2005.
9. S. E. Daughdrill, T. R. Reust, T. L. M. Moore, S. P. Leach, M. J. McClure, M. D. Roughton, K. A. Kremer, and P. Amigues. A Comprehensive Study on the Mechanisms of Protein-Protein Interactions. *Biophys J*, 25:117–125, 2012.
10. J. K. Tho, M. E. Bronk, C. J. Rippner, and T. O. Cunnell. Structural Determinants of the Viral-Host Binding Mechanism. *J Virol*, 33:249–257, 2011.
11. A. A. Jones, G. R. Johnson, A. R. Upton, R. F. Q. Upton, R. H. Harrison, M. L. Scott, M. B. Tarver, and M. Smith. Comparative Study of Biological Data Repositories. *Annu Rev Biophys*, 47:103–120, 2018.
12. J. B. Urena, G. Q. Cabral, M. I. Sanchez, and M. L. Dodger. Machine Learning Applications in Drug Design. *Annu Rev Biophys*, 47:97–110, 2018.
13. P. H. Hanken, D. H. Allison, J. T. Forbes, and R. A. Brown. New Developments in Drug Interaction Studies. *Nat Rev Chem*, 8:99–106, 2015.
14. R. J. Christensen, D. W. Goldstein, T. M. Cummins, J. W. Hamilton, and C. W. Karan. Mechanisms of Cellular Binding Interactions. *Curr Opin Struct Biol*, 23:77–92, 2019.
15. K. R. Smith, D. A. Samuelson, M. K. Sanders, and E. A. Lewis. Proteomics Analysis of Cellular Binding Sites. *Cell Microbiol*, 15:1121–1135, 2020.

16. W. Li, L. W. Ma, A. Qasim, S. Anderson, J. Apostol, A. Walder, S. M. Bigler, Y. W. Chiang, J. Cook, M. N. James, J. I. Kato, C. Kelly, S. Shust, K. Tomiyama, K. Inoue, R. Keegan, S. M. Aswad, S. J. Park, S. Qasim, M. Qasem, M. Tahsin, D. W. Nawe, E. R. Tiyati, N. Askenazi, W. M. Wescott, M. Witsur, R. Wym, R. Wayne, and D. A. Hu. Predictive Modeling of Binding Interactions in Protein Complexes. *Proc Natl Acad Sci USA*, 104:8109–1415, 2018.
17. S. Lambert, S. A. M. Du, and K. J. Davies. The Evolution of Cellular Signaling Mechanisms. *Nat Rev Mol Cell Biol*, 21(3):210–218, 2017.
18. E. J. Lawrence, R. C. Devereux, and M. L. Rogers. Structural Dynamics of Protein Interactions. *J Mol Biol*, 429:1469–1476, 2016.
19. D. M. Glotzer, L. J. Larson, and P. S. Hagman. Machine Learning Applications for Drug Design. *Nat Biotechnol*, 37(7):781–792, 2019.
20. D. M. Perez, G. M. Rodrigues, M. C. Sanders, J. W. Ford, T. O. Folly, and H. D. Witte. Proteomic Analysis of Cellular Binding Sites. *Nat Struct Mol Biol*, 18:213–220, 2018.
21. B. T. Perez, D. H. Murphy, G. K. Hall, J. E. Cooper, K. J. Freeman, T. N. Jones, and P. A. Smith. Target Identification and Validation in Drug Discovery. *J Med Chem*, 62(15):6947–6956, 2019.
22. C. Zorowka, G. Vanegas, and A. M. J. J. Borwin. Exploring the interplay of reactivity and binding affinity in protein complexes. *Process Eng Des*, 298(2):299–329, Aug 2016.
23. L. Gossage, D. E. Pires, A. Olivera-Nappa, J. Asenjo, M. Bycroft, T. L. Blundell, and T. Eisen. An integrated computational approach can classify VHL missense mutations according to risk of clear cell renal carcinoma. *Hum Mol Genet*, 23(22):5976–5988, Nov 2014.
24. G. Guerois, J. E. Nielsen, and B. Kuhnham. Engineering a protein-protein interface using a computationally designed library. *Proc Natl Acad Sci USA*, 101(24):7969–19301, Nov 2010.
25. W. Guo, J. A. Wessels, and R. H. Hot. Host species and dose-specific anti-molecule inhibitors for protein-protein interactions. *Bioorg. Med. Chem. Lett.*, 21(15):2462–2554, Jun 2014.
26. D. Hammond and P. Baumann. Structure and dynamics in the overcoating of intraviral-advection interactions in peptides and proteins. *J Mol Biol*, 14(3):613–61, Apr 2009.
27. T. Huynh, M. V. Georoy, V. Graziano, J. M. Flanagan, and P. Freund. Structural basis for variations in adenovirus affinity for the cellular co-receptors and adenovirus receptor. *J Biol Chem*, 278(26):26028–26151, Jul 2003.
28. J. H. Lei, J. N. Chen, and X. Zhang. Conservation of hot regions in protein-protein interaction in evolution. *Methods*, 110:73–80, Nov 2016.
29. X. Huang and D. L. Brutlag. Dynamics of multiple parameter sets in sequence alignment. *Nucleic Acids Res.*, 35(21):676–686, 2007.
30. G. Kuhn, P. Verzyl, and J. Wang. Binding interface recognition by scanning protein-protein docking results. *Proteins*, 82(15):57–66, Jan 2014.

31. S. Jamshidi, K. Youngdahl, and M. Michael Gromnin. PROXIMATE matrices for flexible docking algorithms. *Biochemistry*, 33(17):278–278, Sep 2017.
32. P. Jemtehu, M. A. T. Jensen, and B. Ambo. A frustrated binding surface across diverse inherently disordered proteins. *J. Biol. Chem.*, 289(9):5523–5538, 2014.
33. L. Kasstic, I. H. Moal, H. Hwang, Z. Weng, P. A. Bates, A. M. Brown, and J. Iann. A structure-based benchmark for protein-protein docking affinity prediction. *J Mol Biol*, 25(30):482–491, Mar 2011.
34. C. Keidel, L. Serrano, and C. Hennerman. A detailed thermodynamic analysis of reflector complex interfaces. *J Mol Biol*, 34(05):1039–1055, Jul 2004.
35. M. Kulkarni, M. C. Tettemanti, J. W. Murphy, C. Keeler, D. G. Myszk, N. E. Chayen, E. J. Lois, and M. E. Hodkson. Two independent roles of the human monoclonal antibody receptor in eaearcell self-dependent receptor recognition and activation. *J Biol Chem*, 25(49):594–3624, 5833, Dec 2000.
36. M. D. Thunn and M. Denbach. Nuclear Proteoin-Interaction Databases. *Nucleic Acids Res.*, 34:Database issue:D195-D198, Jan 2006.
37. K. Ji, J. A. Doh, Z. Wang, and P. Bardh. Enhancing Structure Prediction for a Design of Soluble and Membrane Proteins with Explicit Solvent-Protein Interactions. *Structure*, 25(11):1578–1770, Nov 2017.
38. E. D. Levy. A simple definition of structural divergence and its relevance to protein binding interface evolution. *J Mol Biol*, 40(35):660–674, Nov 2019.
39. M. L. L. Simonetti, A. Gommcan, and A. R. Fanchedna. Mutabiling entries and interpreters for multiple effects of sequence variations on protein-protein interactions. *Nucleic Acids Res*, 44(W1):494–501, Jul 2016.
40. M. Cota, A. Dennis, G. R. Moore, R. James, and C. Kleanthous. Protein-protein interactions specificity of Im9 for the endonuclease protein colicin E9 determined by molecular modeling. *J Biol Chem*, 285(19):422–423, Jul 1997.
41. Q. Lin, Z. Chen, B. Wang, and J. Liu. dbMPIKT: A web resource for the interaction hot spots, dynamics of mutation protein interactions. *Nucleic Acids Res.*, 35:D148–D153, Jan 2011.
42. J. L. Miethke, M. White, D. N. Chandon, and P. Khan. A new global drug database for studies in oncology. *Nucleic Acids Res.*, 44:D220–D230, Feb 2015.
43. G. Chang, W. Cheng, C. T. Lewis, and M. L. Rogers. Analyzing molecular evolution of proteins. *BMC Bioinformatics*, 21:49–76, 2016.
44. C. J. Robinson, P. Ehmer, P. Robinson, G. M. Verbi, and M. Z. Upton. Analysis of Interface-Restructure Changes. *Nat Struct Biol*, 45:333–357, 2017.
45. H. I. Moal, B. Barradas-Bautista, B. Jimenez-Garcia, M. Torchala, M. van der Velde, T. Vreven, Z. Weng, P. A. Bates, and J. Fernandez-Recio. IRaPPA: information retrieval based threading for the prediction of biophysical models for protein assembly selection. *Bioinformatics*, 33(12):1806–1813, Jun 2017.
46. B. Jimenez-Garcia, M. Moal, and J. Fernandez-Recio. pyDockSAXS: protein-protein complex structure by SAXS prediction. *Bioinformatics*, 26(17):1956–1957, Jul 2014.

47. H. I. Moal, B. Jimenez-Garcia, B. Barradas-Bautista, B. Jimenez-Garcia, M. Torchala, and J. Fernandez-Recio. Protein-protein docking with pyDock using shape complementarity. *Bioinformatics*, 29(8):1081–1083, May 2013.
48. B. Jimenez-Garcia, M. Moal, B. Jimenez-Garcia, and J. Fernandez-Recio. CCharPPI: Characterization of Protein-Protein Interactions Using Docking Algorithms. *Nucleic Acids Res.*, 42(Web Server issue):W229–W233, Jul 2014.
49. D. A. Smith, L. Verbeeck, J. A. Wells, C. Karason, and W. R. Dirkx. A new paradigm for drug repositioning: structure-based optimization. *Drug Discov. Today*, 21(1):25–33, 2016.
50. K. A. Barth, A. S. Engelman, R. L. Lockman, M. G. Dingwall, and G. Scarcella. Binding specificity and stability of therapeutic antibodies against SARS-CoV-2. *Proc Natl Acad Sci USA*, 112(20):3319–3324, Apr 2017.
51. M. K. Foster, B. Varghese, J. L. Leung, J. Weiller, R. P. Stack, and G. Adams. Functional characterization of a candidate therapeutic molecule with broad antiviral properties. *J Virol*, 79(17):10349–10352, Jul 2016.
52. G. L. Roberts, R. Gray, K. J. Harrow, M. Whitby, and P. Jones. Analysis of gene expression patterns in liver carcinoma. *J Exp Med*, 13(3):493–502, Jun 2017.
53. M. D. Andre, B. Biederman, R. D. Lavendoski, and M. P. Wagener. Construction of high-affinity inhibitors for bacterial and viral pathogen targets. *Science*, 336(6):453–456, Apr 2014.
54. R. T. Lorentz, M. Pegram, M. He, and M. Heinemann. Characterization of the receptor-ligand binding properties in hepatocellular carcinoma. *Biochem J*, 4(1):110–114, May 2018.
55. S. M. Bernofsky, R. Ke, and T. Do. Evaluation of the impact of viral dynamics on the immune response. *Nature*, 485(12):506–510, Jan 2012.
56. B. L. Paterson, E. Weinberg, and M. E. Mann. Mechanistic insights into interactions with interleukin 6 and its receptor. *Proc Natl Acad Sci USA*, 120(15):1055–1061, Mar 2014.
57. J. W. Lee, B. J. Castañeda, and M. A. Aleem. Targeting viral entry mechanisms with monoclonal antibodies. *Cell Host Microbe*, 18(7):325–328, Feb 2019.
58. F. M. Ferrer, E. T. Harris, and J. L. Tams. Structural modeling of antibody interactions with viral proteins. *Trends Immunol.*, 34(2):185–190, Apr 2020.
59. M. P. Stone, C. L. Cheng, and A. M. Berman. Comparative analysis of viral infection pathways. *Antiviral Res*, 88(10):561–564, Mar 2016.
60. R. T. Greenwald, G. P. Lewis, and D. A. McCarthy. Interactions between the immune system and the viral envelope proteins. *Virology*, 275(3):120–124, Dec 2017.
61. K. A. Goodspeed, A. H. Lee, and T. L. Patterson. Molecular dynamics simulation for evaluating receptor binding. *J Comput Chem*, 30(4):678–685, Oct 2018.
62. P. L. Matthews, J. E. White, and S. C. Waldman. Development of computational methods for high-throughput screening. *Chem Biol Drug Des.*, 57(1):88–95, Jan 2016.

63. J. W. Kim, R. G. Richards, J. F. Hugenschmidt, and T. L. Ashmore. Comparative evaluation of the efficacy of engineered monoclonal antibodies. *Nat Biotechnol.*, 27(8):801–809, Sep 2013.
64. D. J. Patel, S. P. Ford, and M. R. Gagnon. An investigation of antibody interactions with glycoproteins. *J Med Chem*, 59(11):5151–5161, Jun 2018.
65. K. A. Hirose, R. S. Hirota, M. G. Nishiyama, and H. L. James. Discovery of protein inhibitors against bacterial pathogens. *Bioorg Med Chem Lett*, 16(4):237–242, Jul 2017.
66. M. P. Becker, J. S. Aaron, and J. D. Franklin. Design of dual targeting molecules for enhancing drug delivery. *Mol Pharm*, 11(2):372–377, Feb 2019.

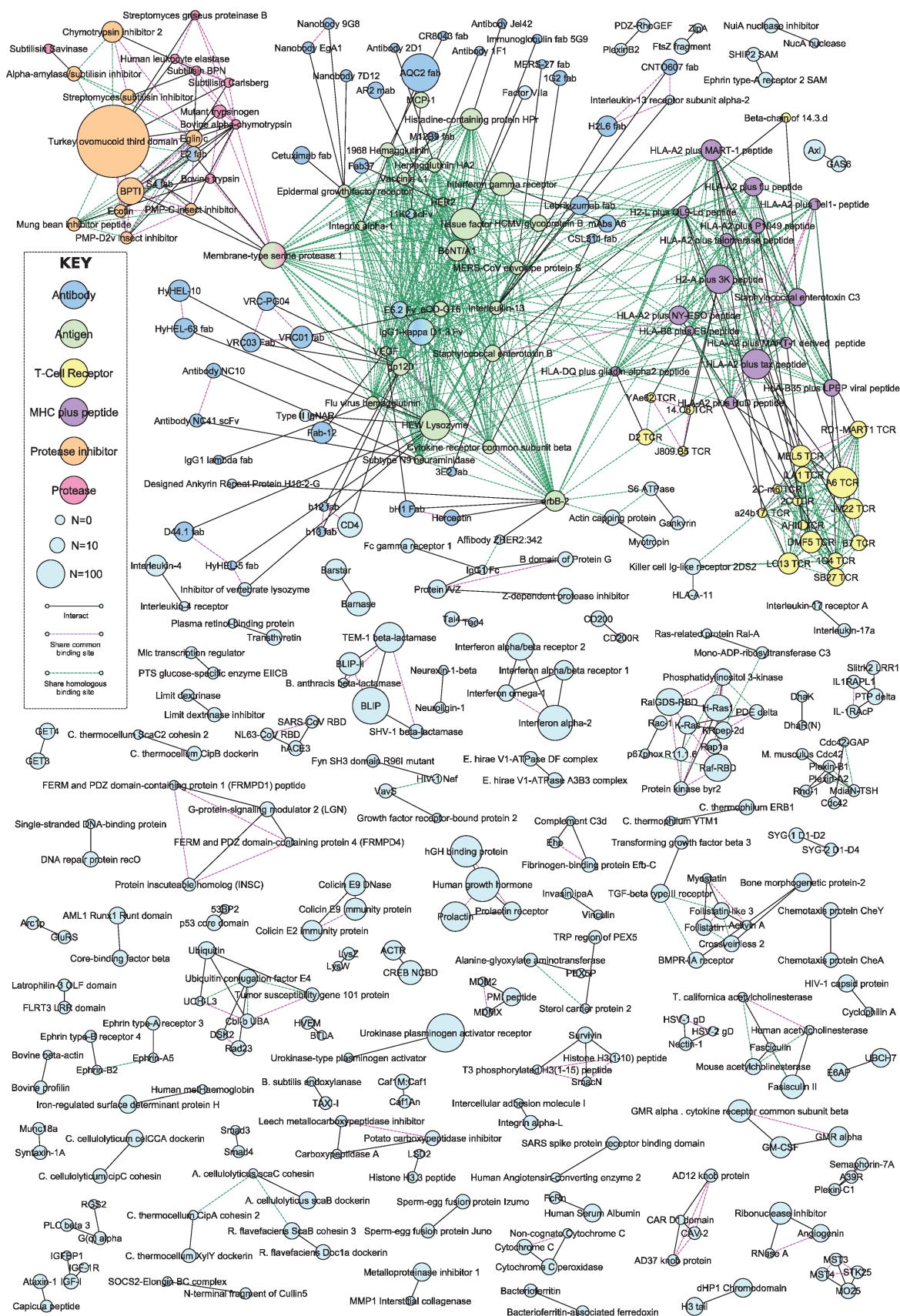


图 2: SKEMPI 中的相互作用概述。节点表示蛋白质, 按该蛋白质的突变数量缩放并根据类别着色。边缘显示直接相互作用, 以及共享共同或同源结合位点的蛋白质之间的关系