

Evolving new protein-protein interaction specificity through promiscuous intermediates

--发现:新的蛋白间特异性互作的产生需要通过复杂的突变中间物

翻译 : 徐垚, 郎昊, 梁鹏远
PPT : 徐垚

INTRODUCTION

同一细胞内的许多相互作用的蛋白质，特别是信号蛋白质，是通过复制和突变而扩展的大型旁系同源家族的成员。为了扩大数量，旁系相互作用蛋白质通常在复制后必须变得特异性，以避免不必要的串扰。

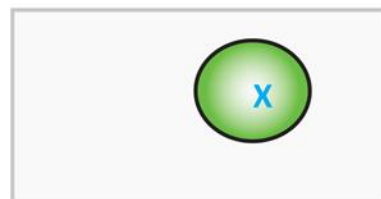
在大多数系统中蛋白质 - 蛋白质相互作用的特异性决定因素还没未可知。即使在已被鉴定的情况下，我们也不了解进化过程中出现新的，绝缘的蛋白质 - 蛋白质相互作用的关系，更普遍地说，是在蛋白质进化过程中遵循的突变途径我们还不知道。

A**model 1:
compensatory mutation**

↓ specificity-switching
mutation in protein A



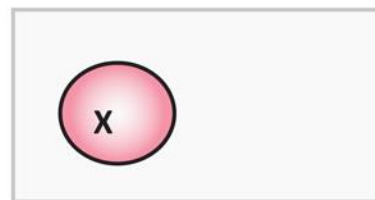
↓ compensatory
mutation in protein B



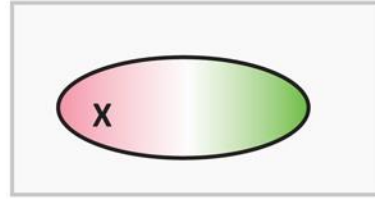
○ set of protein B variants
bound by protein A

X ancestral protein B

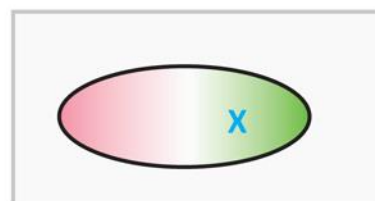
X derived protein B

B**model 2:
promiscuous intermediate**

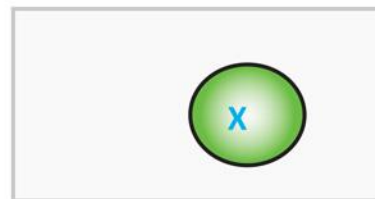
↓ promiscuity-inducing
mutation in protein A



↓ mutation in protein B



↓ specificity-narrowing
mutation in protein A



思考

两个模型中的哪一个适用于大多数相互作用蛋白质？

compensatory mutation
model

VS

promiscuous intermediate
model

细菌毒素 - 抗毒素 (TA) 系统

---解剖蛋白质 - 蛋白质相互作用的协同进化动力学的很好的模型系统

通常，毒素是稳定的球状蛋白，其可以抑制细胞生长或活力，除非直接结合和螯合毒素的同源抗毒素拮抗。

抗毒素降解速率或合成速率的变化可以引发毒素的释放。通常毒素与抗毒素由相同的操纵子编码，毒素 - 抗毒素旁系同源物通常通过操纵子的重复产生。

问题：

TA系统是以独一无二的方式相互作用的吗？

相互作用可能是特异性的 ([Fiebig et al., 2010](#))

毒素的生长抑制作用通常仅通过表达其共同作用的抗毒素来获得 ([Haliez et al., 2010](#); [Ramage et al., 2009](#))

不同操纵子中编码的毒素和抗毒素能够在体内和体外相互作用，可能形成大的杂交网络([Yang et al., 2010](#); [Zhu et al., 2010](#))

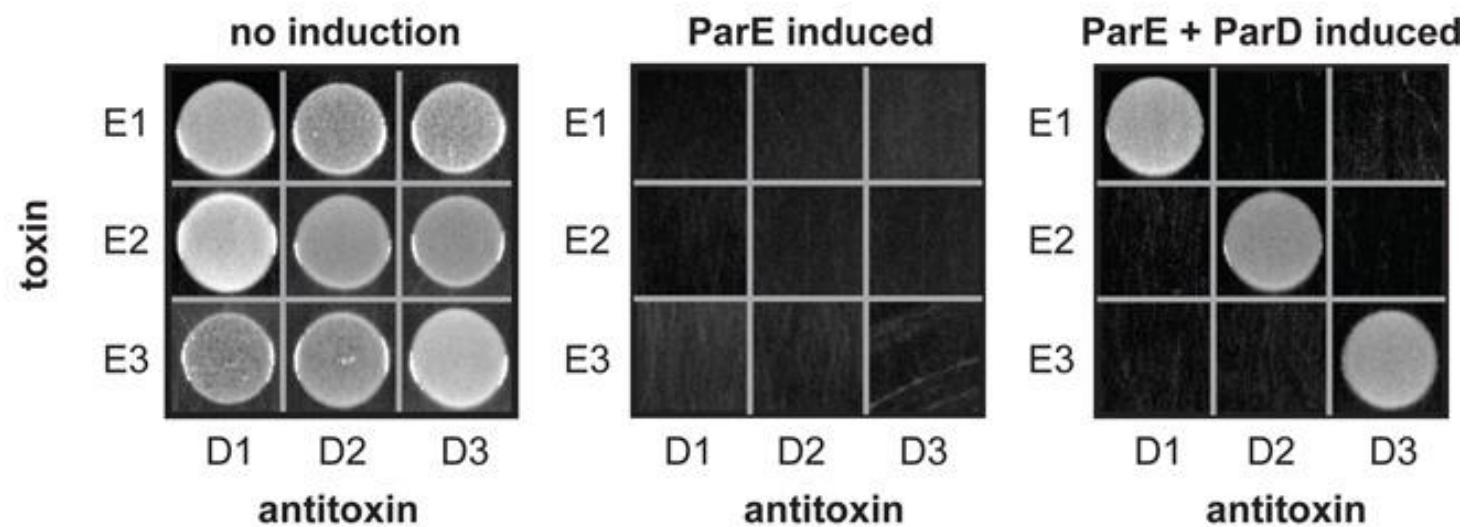
实验对象：

为了系统地测量TA系统的相互作用特异性，我们选择通常存在于细菌染色体上的多个拷贝的ParD-ParE家族

实验步骤：

我们最初将编码的ParD-ParE三个染色体的基因从 α -proteobacterium *Mesorhizobium opportunistum* (α -蛋白杆菌) 克隆到允许分离和诱导表达ParE毒素和ParD抗毒素的载体中

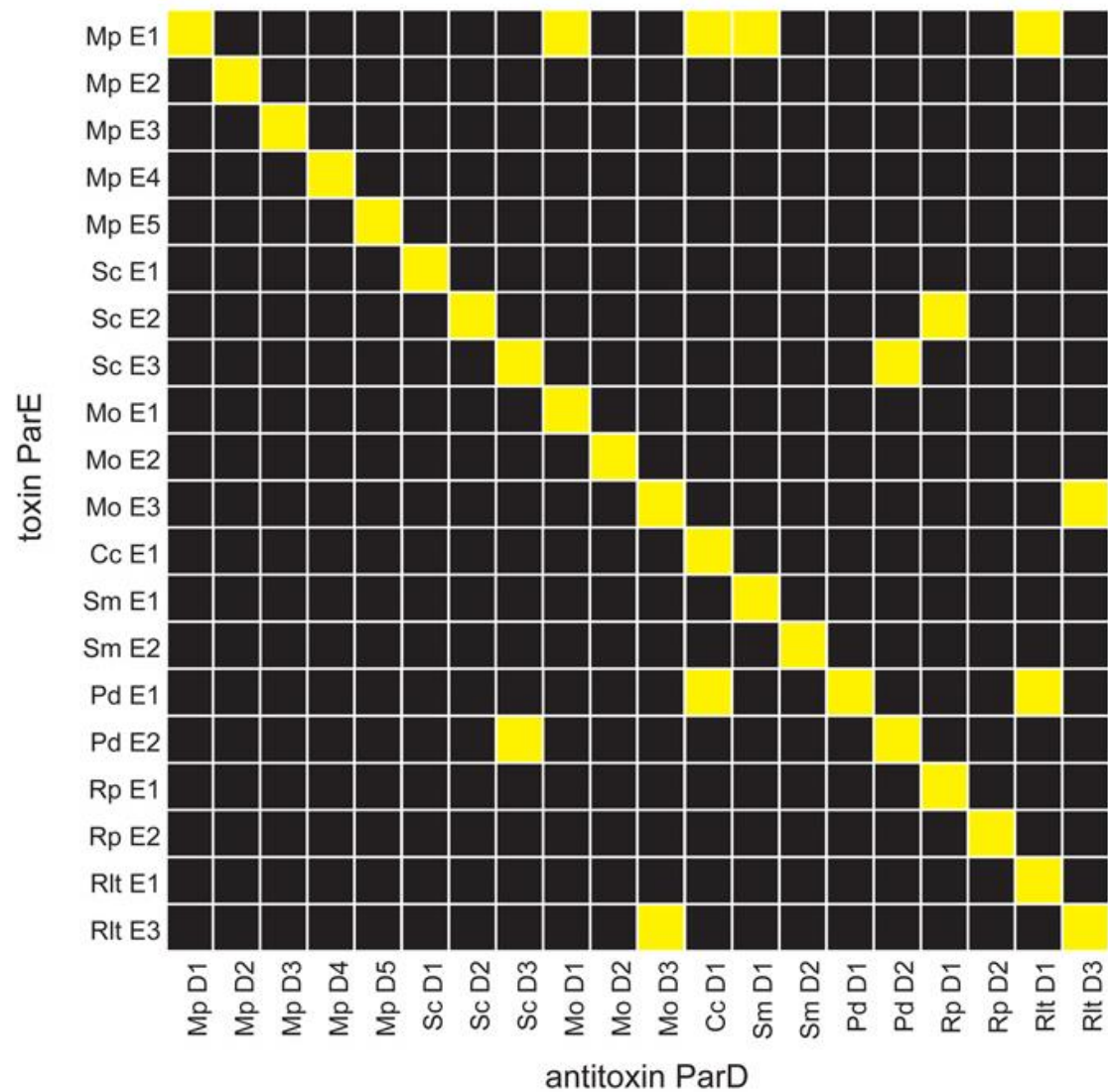
A



六个非同源对没有观察到生长，表明来自 *M. opportunistum* 的 ParD 抗毒素只能中和其同源的 ParE 毒素

PS:

ParD 抗毒素
ParE 毒素

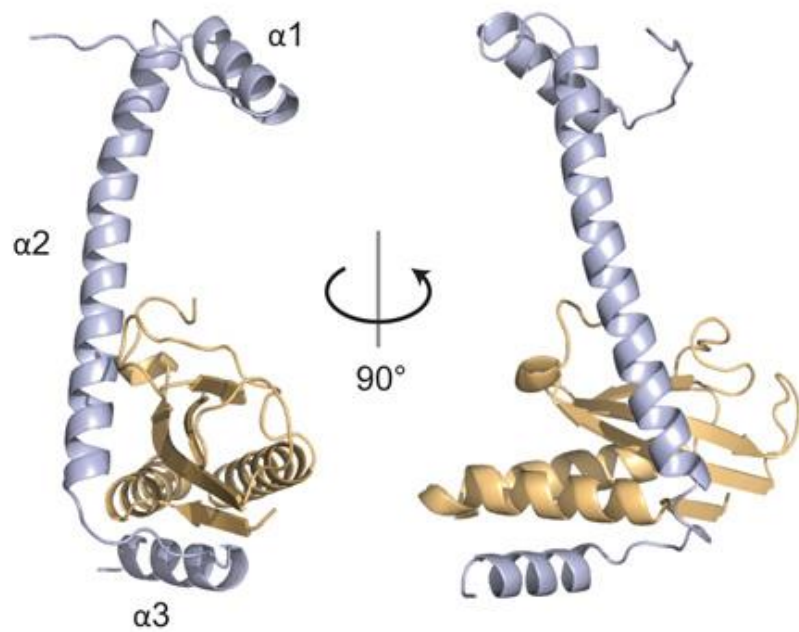
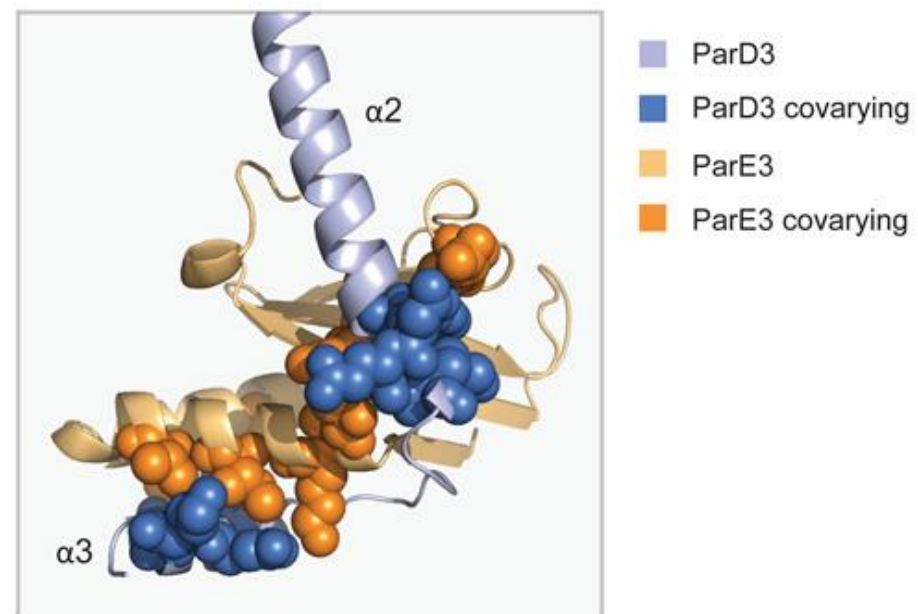
B

Y = visible colony growth B = no colony growth

结论：
ParD抗毒素
对于其同源的
ParE毒素具有高
度的特异性

first step

Identification of covarying residues in ParD and ParE

A**B**

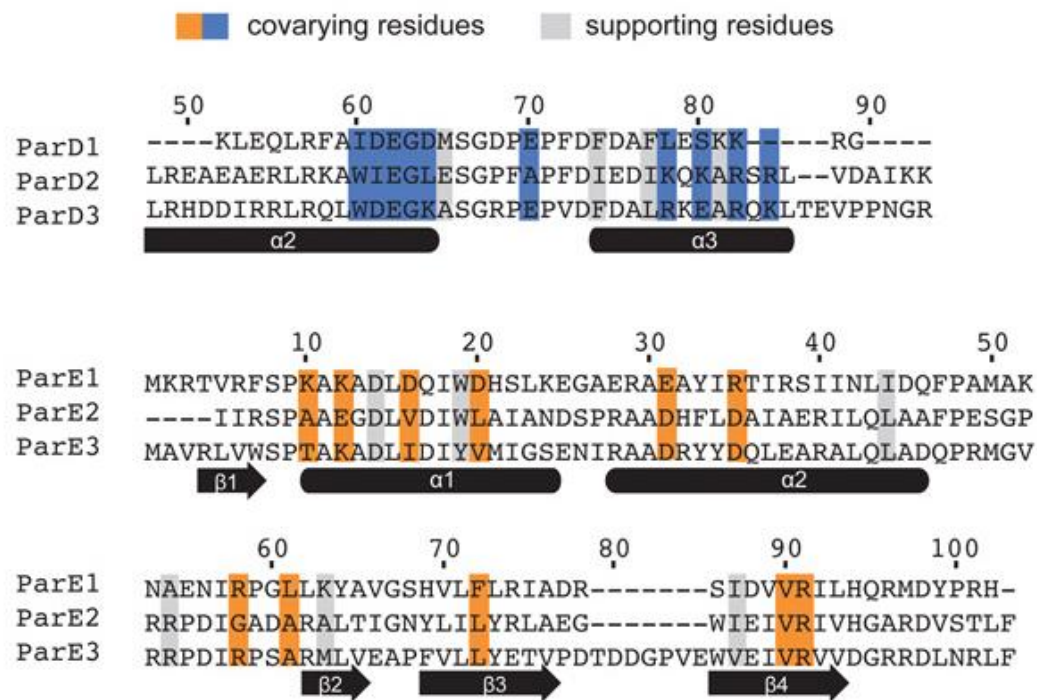
Identification of coevolving residues

使用**GREMLIN模型**鉴定ParDE家族中的共培养残基。

输入序列是来自 *M. opportunistum* 的ParD3和ParE3，我们将迭代次数设置为4，将E值截止值设置为1E-04。为了鉴定特异性残基，我们分离了比例大于1.25的比例分析的所有残基配对。为了识别支持残基，我们使用1.25的得分截距进行以下迭代过程：

- (1) 鉴定ParD或ParE中与特异性残基相关的残基;
- (2) 识别ParD或ParE中与步骤 (1) 中鉴定的特异性残基或支持性残基共存的残基;
- (3) 重复步骤 (2) ， 直至没有鉴定出新的支持残基。

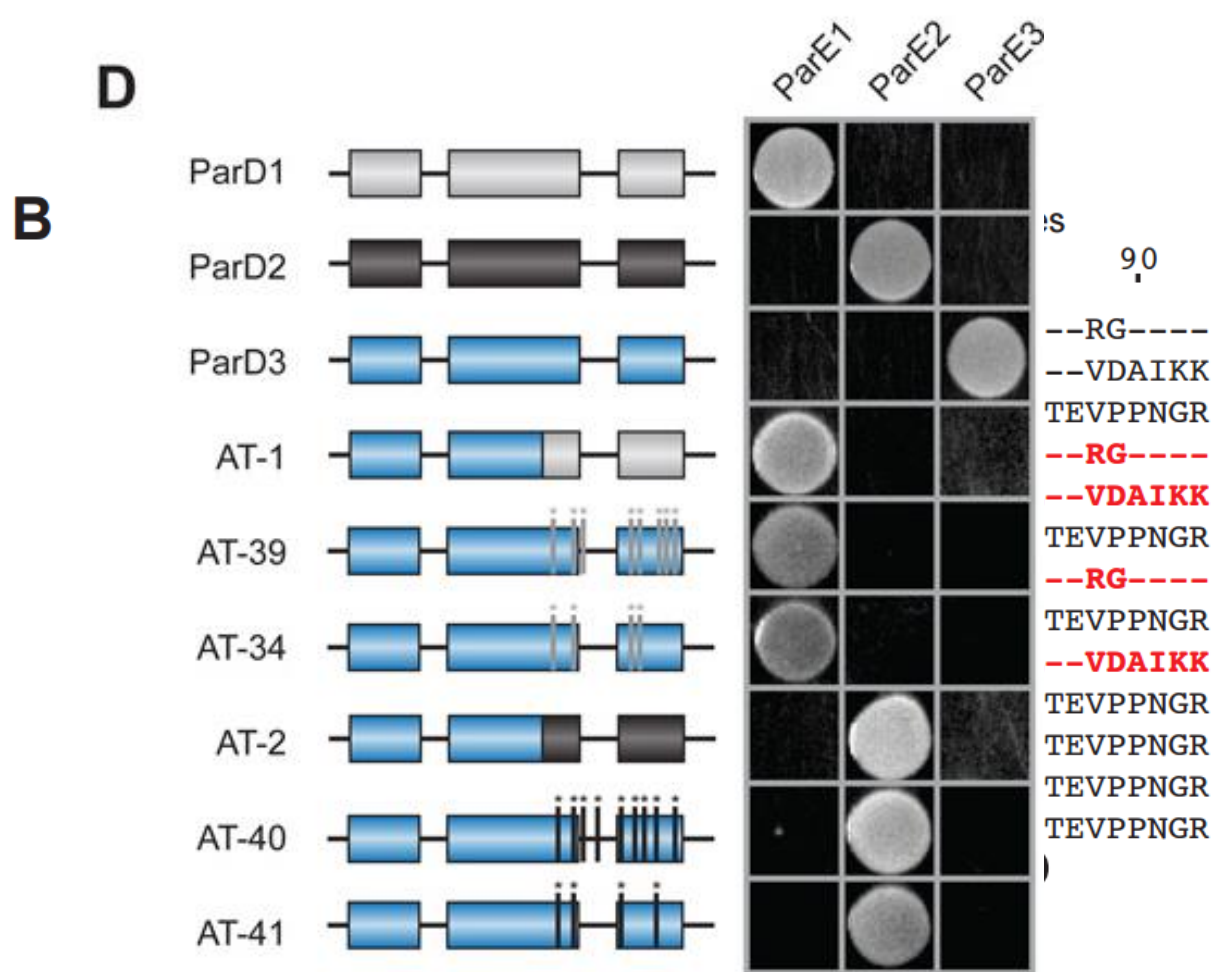
C



该分析确定ParD中的10个残基和ParE中最强烈共有的11个残基之后，我们称这21个氨基酸为“特异性”残基

我们还使用GREMLIN来识别与特异性残基共存残基（ParD中的四个，ParE中的六个）

这些“支持”残基可能通过影响界面特异性残基的取向或包装间接地促成ParD-ParE相互作用特异性



例如，我们发现了四个替换的集合，足以重新编程ParD3与ParE1或ParE2进行交互。

综合起来，我们的研究表明，替换抗毒素中突变率高的特异性残基足以重新编程其相互作用特异性，并且在某些情况下，仅突变这些残基的一个也可以完全切换伴侣的特异性。

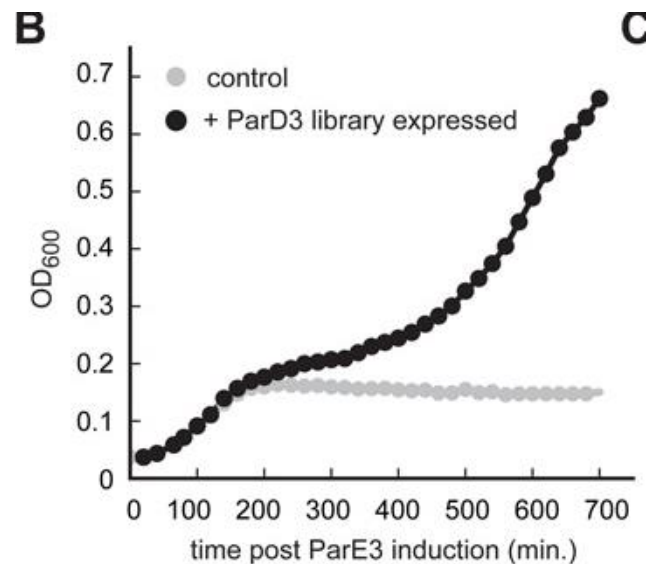
上述结果表明，通过仅改变四个残基可以重编抗毒素之间相互作用的特异性。但是，如果只替代四个残基中的一个，特异性如何变化，替代顺序是否带来影响？从ParE毒素对另一种毒素的拮抗特异性的转变是突发的还是经历了突变中间体？

为了回答这些问题，我们试图生成一个大型的ParD3变体文库，其中包括显示为特异性拮抗ParE3或ParE2的残基的组合，以及分离这些特定状态的突变中间体。

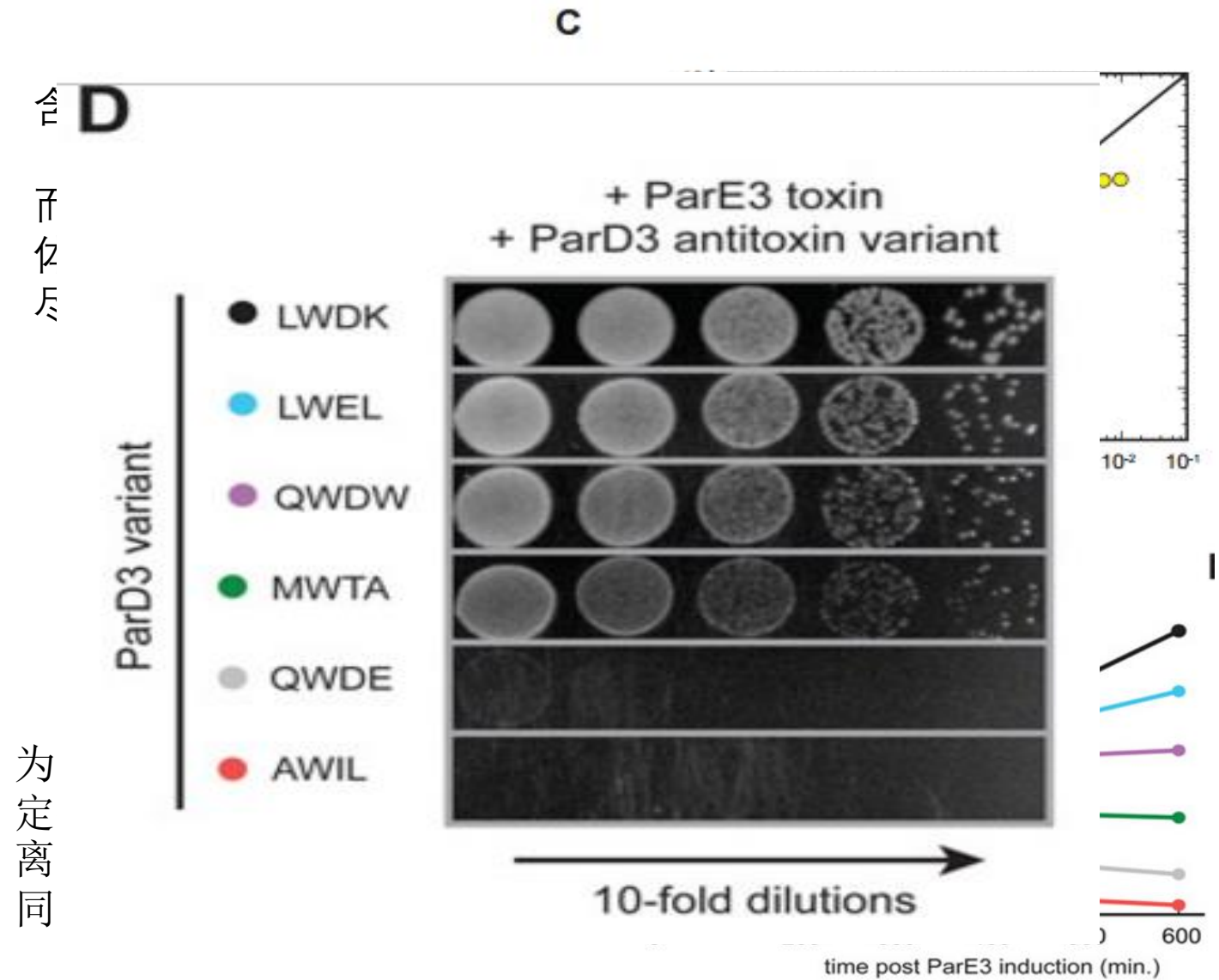
为此，我们在ParD3抗毒素的四个关键特异性氨基酸残基Leu59, Trp60, Asp61和Lys64 (LWDK) 位置产生了一个突变体库。为了减少我们的图书馆的复杂性，我们每个文库的位置只允许在天然存在的ParD同源物中通常存在的残基

所得文库具有9,360个变体的理论多样性，其中在文库的四个相应位置编码的12,6,13和10个可能残基

为了评估每个ParD3变体结合和拮抗ParE3的能力，我们用ParD3文库和可诱导的ParE3载体共转化大肠杆菌。当在不诱导ParD3的条件下培养时，细胞生长在诱导ParE3毒素后200分钟内停止。相比之下，当ParD3文库被表达时，诱导毒素之后的生长减缓，但最终恢复，表明一些部分群体可以中和ParE3毒性。



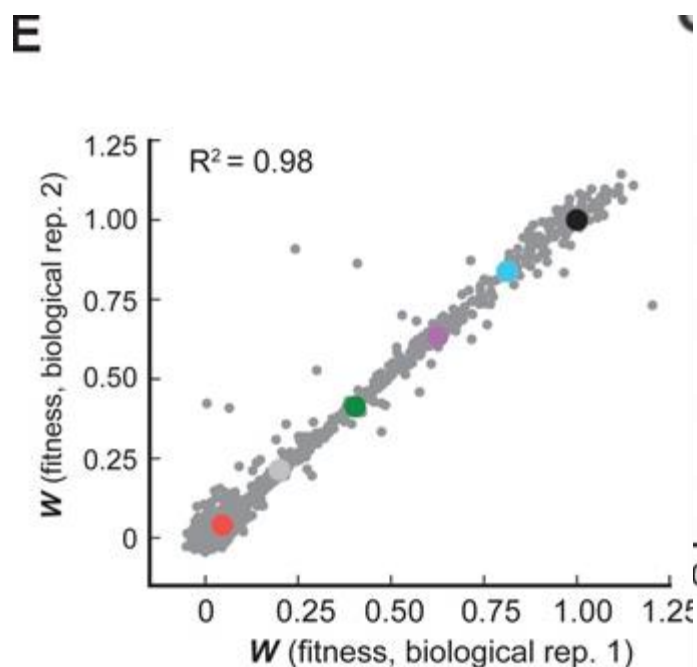
为了确定哪个突变体中和了ParE3，因此在本实验过程中被富集，我们每100分钟收获一次样品，并对parD3的相关区进行深度测序。



含子体片
为定离同

为了量化竞争性生长期间变异行为的差异，我们产生了与每个突变体的频率与时间的线性拟合函数，然后取其与其余群体的对数，产生原始健康值 (W_{raw})。然后我们转化了这些原始适应度值，使得移码变体的 W 值为 0，野生型 (LWDK) 序列的 W 值为 1，所得到的 W 值分布范围为 -0.04 至 1.13，并且在生物重复之间具有高度的可重复性。

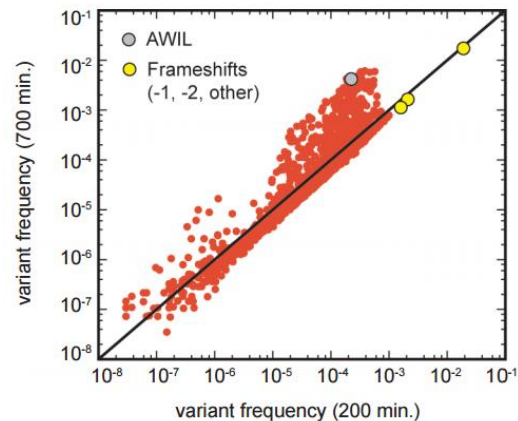
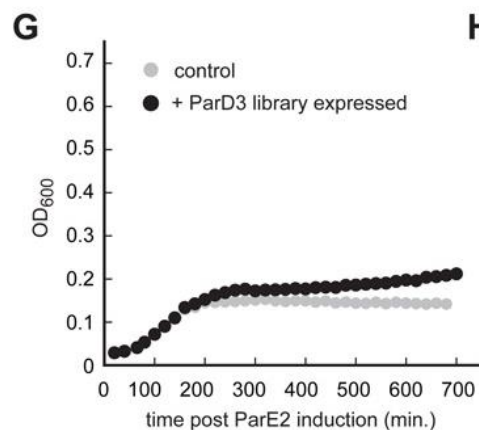
我们发现总共 252 种变体， W 值 > 0.5 ，占总数的 2.7% (图 F)。该组包括相对于野生型序列的残基 (LWDK) 和 31 个单个，189 个双重和 31 个三重突变体的野生型组合 (图 D)。没有四倍体突变体，因为位置 60 是不变的色氨酸。该组中最常见的残基是野生型的残基。然而，鉴定可以有效拮抗 ParE3 的 252 种变体表明 ParD 特异性面残基中的相当程度的功能退化。



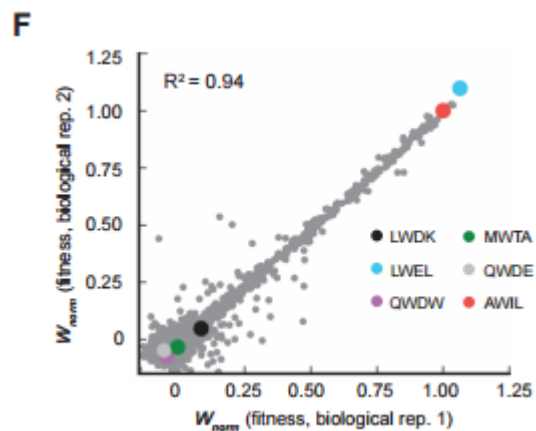
D

mutations	scored	functional ($W > 0.5$)
0	1	1 (100%)
1	37	31 (84%)
2	499	189 (38%)
3	2798	31 (1%)
4	5859	0 (0%)

含有在天然ParD2抗毒素（AWIL）中发现的特异性残基的变体在ParE3文库中被富集，ParC3文库针对ParE2进行筛选，但是当对ParE3进行筛选时被耗尽



如前所述，我们量化变体适应度，发现总共151种变异体（占总体1.6%， W 值 > 0.5 ）能够拮抗ParE2。最常见的残基是Ala59，Trp60(不变量),Leu61和Leu64



fit vs. ParE3

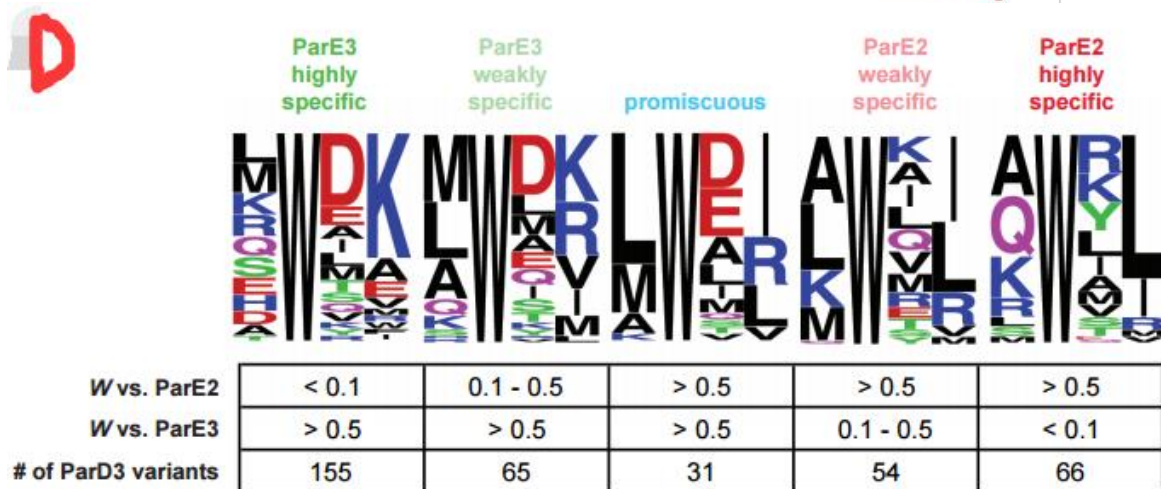
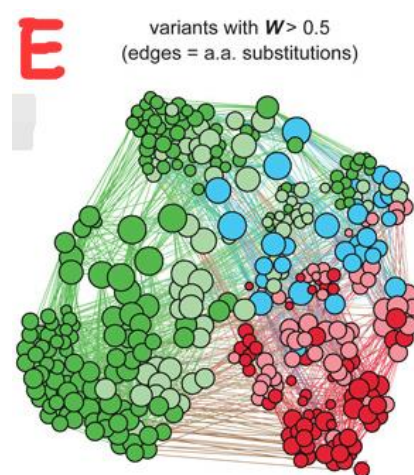
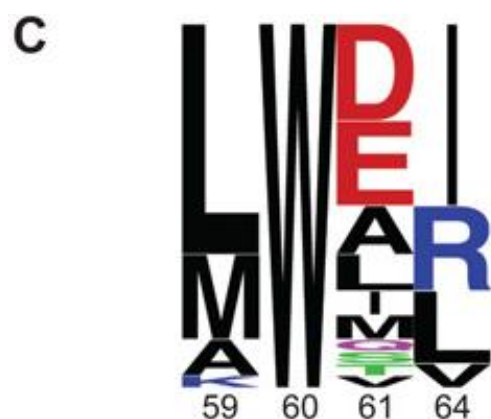
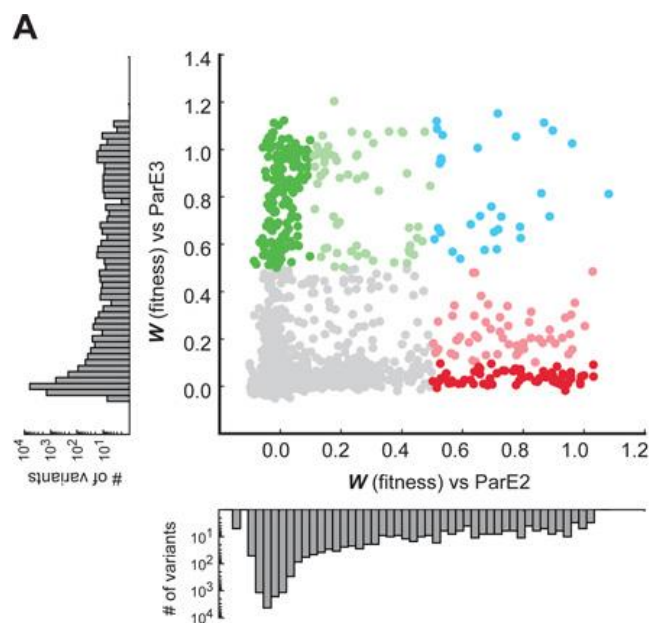


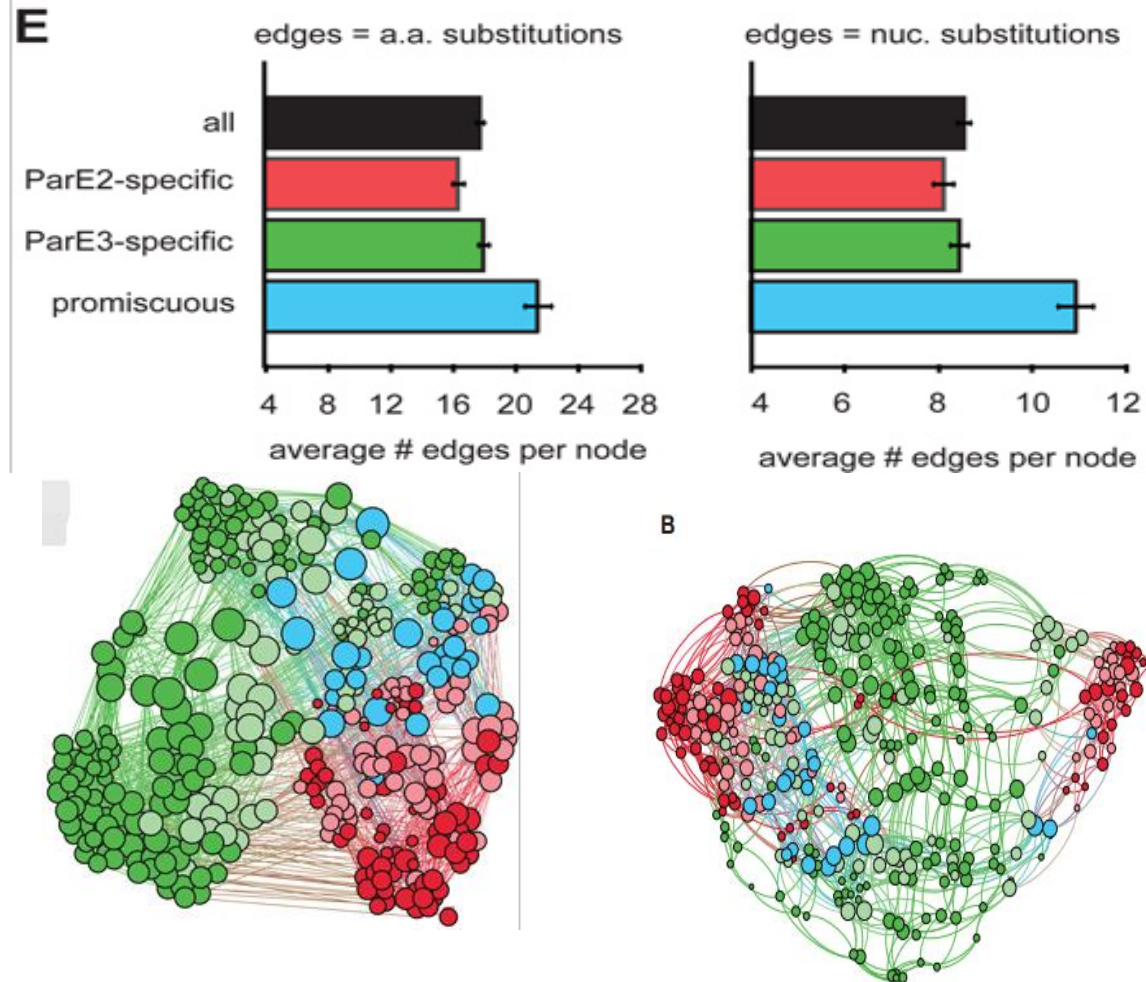
fit vs. ParE2



通过对比，我们注意到与ParE2和ParE3反应的变体之间的重要差异，特别是在我们文库的最后两个可变位置。ParE2特异性变体在61位倾向于具有小的疏水性或带正电荷的残基，而ParE3特异性变体在该位置有利于负电荷残基。此外，ParE2特异性变体在64位更可能含有小疏水残基，而ParE3特异性变体往往具有带正电荷的残基

为了更系统地探索控制ParD3特异性的序列空间，当对ParE2或ParE3毒素进行筛选时，我们产生ParD3变体适应度的散点图（图A）。该分析显示跨越了所有适合范围的突变体，包括能够同时拮抗ParE2，ParE3或两种毒素的变体。我们确定了共有31个混合变体（两种毒素都为 $W > 0.5$ ），其代表252个ParE3反应性和151个ParE2反应性变体的一个子集（图B）。然后，我们然后通过特异性类别分组变体（图D），发现混合变体，如LWEL，倾向于携带来自ParD3和ParD2的序列元件，通常具有负电荷残基(位置61)和脂肪族残基(位置64)（图C）。



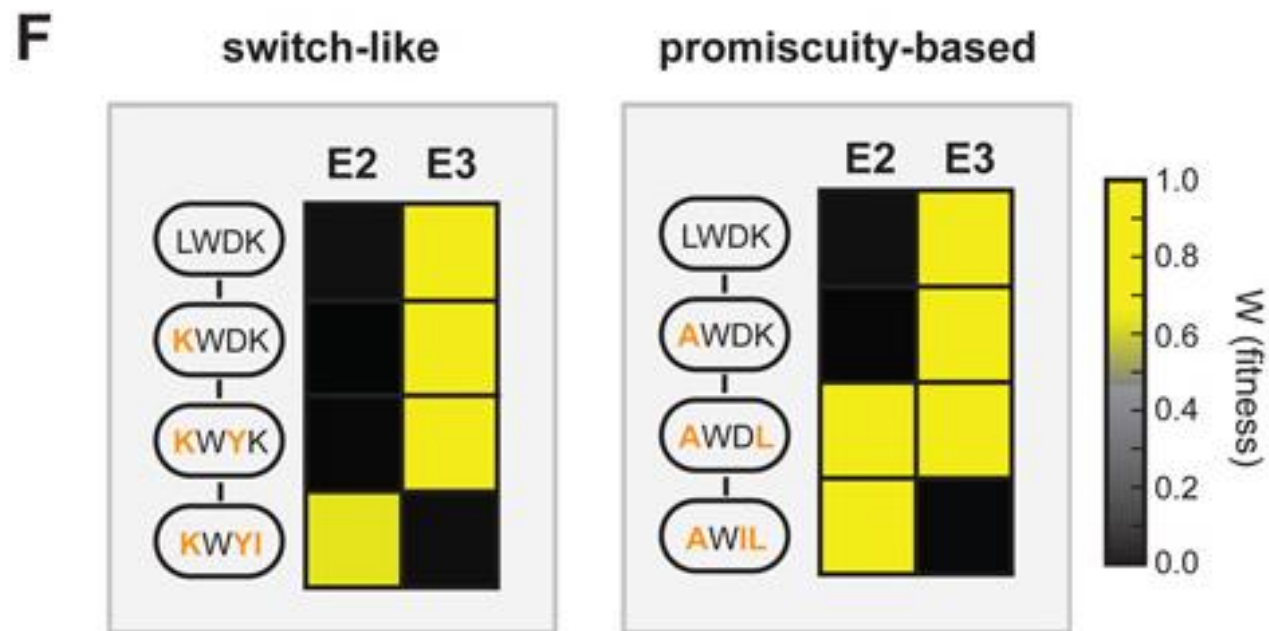


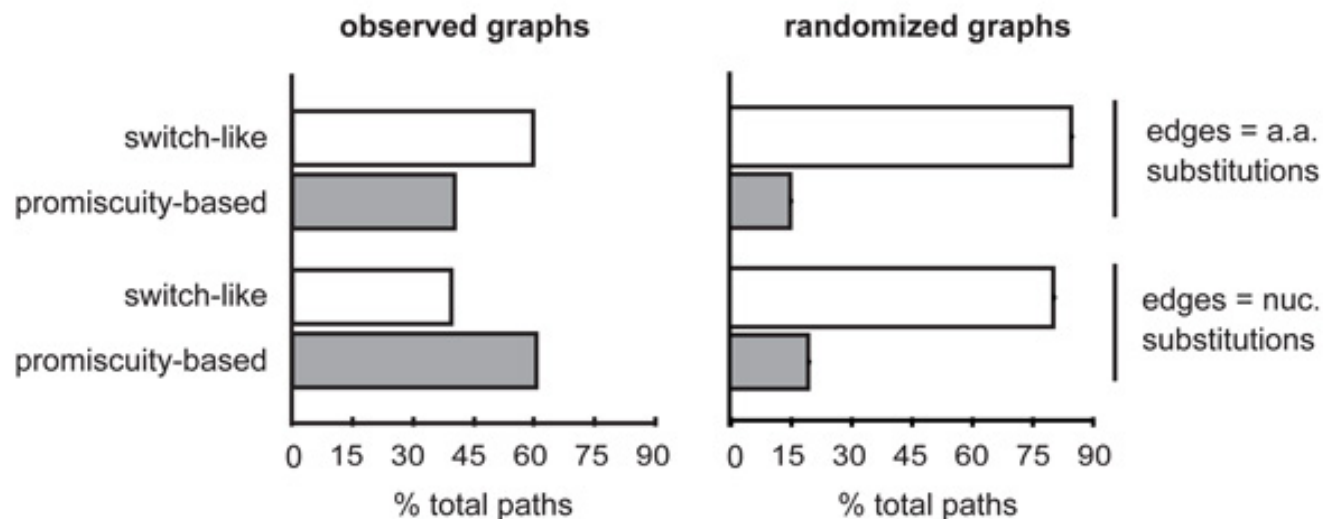
每个节点的平均边缘数或度数为17.8，范围为7到31。

然而，我们注意到混杂变体的平均边缘数比针对ParE2或ParE3特异的变体高出23%

混杂变体平均比其ParE2-或ParE3特异性对应物更多地连接到其他节点31%

定义了两种类型的特异性重编程路径。请注意，对于以下分析，我们排除ParD3无法与ParE3和ParE2进行交互的路径(另请参阅讨论)。**第一类路径是“开关式”**，仅涉及到ParE2或ParE3特定的中间体，而**第二类路径是“混合式”**，并且通过至少一个可以抑制ParE2和ParE3的中间体（图F）。



G

为了确定改变ParD3的相互作用特异性的途径的方式，我们确定了野生型ParD3变体(LWDK)到ParE2高度特异性的66种变体中的每一种的所有最短突变路径($WE_2 > 0.5$, $WE_3 < 0.1$)。对于该分析，每个突变步骤涉及单个氨基酸取代。我们共发现了370条最短路径，其中40%涉及到混杂中间物。当仅考虑涉及单核苷酸取代的途径时，通过混杂中间体的路径百分比增加到61%。

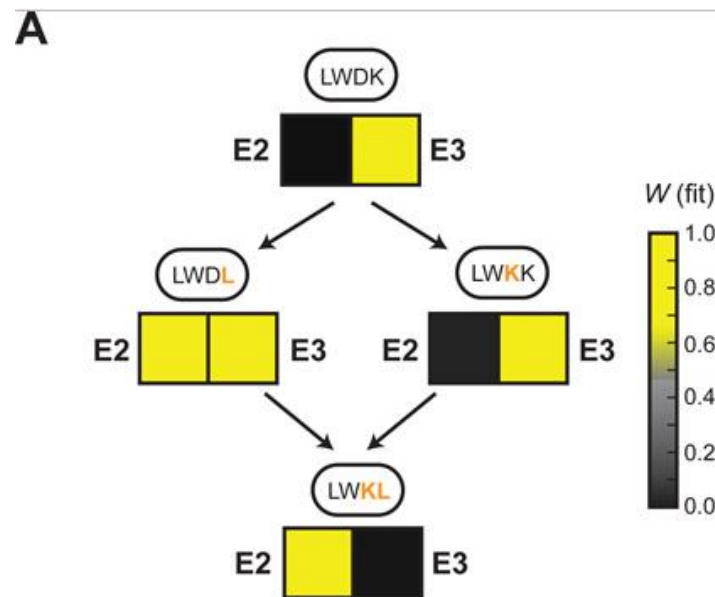
为了确定涉及混杂变体的路径数是否大于偶然预期的路径数，我们生成了[随机洗牌的图形](#)，并再次计算了ParD3(LWDK)到ParE2的每一类路径的百分比。具体的变体。对于具有随机边缘的这些图，涉及混杂中间体的路径的百分比对于氨基酸相邻图降低至15%，对于核苷酸相邻图则降低20%(图G)。

因此，**在观察图中富集基于混合模型的路径是显著的**。总而言之，我们的研究结果证明了功能变体在控制ParD-ParE相互作用特异性的序列空间中的密集连通性，并且揭示了特异性 - 重编程路径对于那些涉及混杂变体的那些来说是高度丰富的，这可能有助于ParD-ParE系统的发展与新的特异性。

epistasis

突变成不同特性的ParD3突变体的路径表明ParD3中的第三和第四库位置，即残基61和64，对ParD-ParE系统的绝缘有重要作用。例如，ParD3，**LWDK**中的野生型残基组合使其特异性结合ParE3，而双突变体**LWKL**对ParE2是特异性的。然而，引人注目的是，连接**LWDK**和**LWKL**的两个可能的路径是不同的类。单个ParD3取代（**LWDL**中的K64L）导致与ParE2和ParE3的混杂结合，而在该背景中的第二个取代（**LWKL**中的D61K）导致对ParE2的特异性。

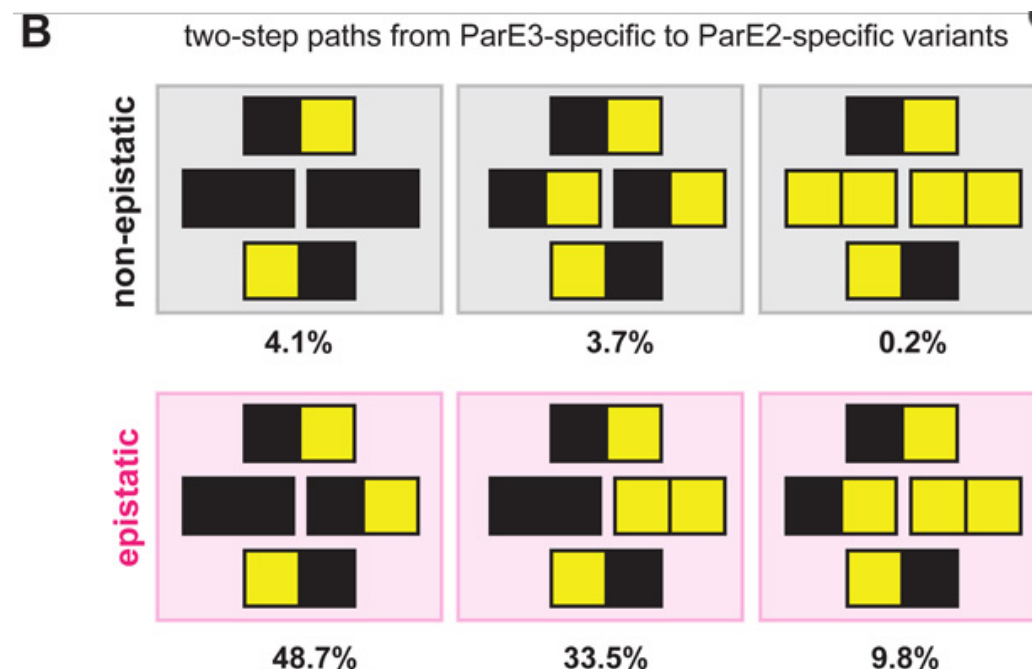
相比之下，以相反的顺序合并这些取代，D61K和K64L导致了特异性的开关样变化，其中初始D61K取代保留了对ParE3的特异性，但随后使随后的K64L取代产生ParE2特异性抗毒素。这些结果强调了少量突变如何完全重新编程蛋白质-蛋白质相互作用特异性，并证明突变的顺序可以强烈地影响到新特异性状态的途径是否涉及混杂中间体或快速转换。



特异性的变化在很大程度上取决于取代的顺序代表一种上位形式。为了更广泛地量化ParD3界面残基的这种上位性，我们首先定义了涉及两个氨基酸取代的六种特异性重编程途径。

六种路径类型中的三种是上位性的，这意味着替代顺序影响从ParE3到ParE2特异性的变化。我们量化了每个病例的路径类型，其中两个取代将ParD3 ($WE3 > 0.5$, $WE2 < 0.5$) 特异性重编程ParD3特异于ParE2 ($WE3 < 0.5$, $WE2 > 0.5$)，并发现总共2,653例，其中92%为上位性。

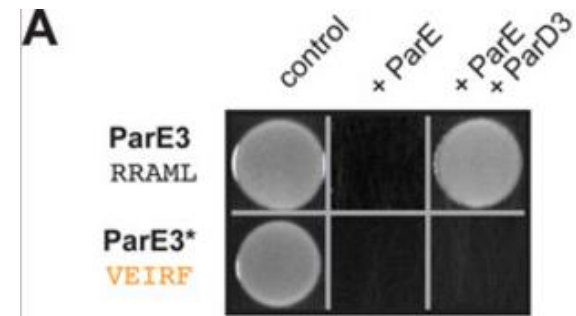
总之，我们的结果突出了上位性对ParD功能的普遍影响。这种上位形式可能会显著影响新的ParD-ParE系统的演进。混杂的中间体使蛋白质-蛋白质相互作用特异性发生变化，而不会通过非功能状态，其中释放的毒素会抑制生长和增殖(图1)。因此，这里记载的上位状态可能会将进化过程中的突变轨迹从根本上限制为涉及突变中间体的突变轨迹。



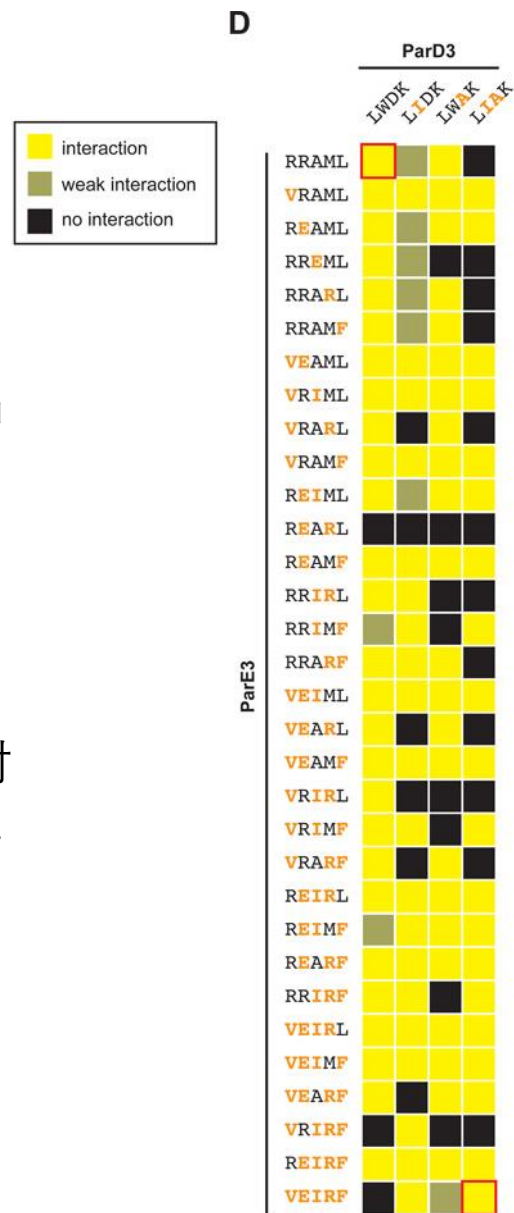
Mutational trajectories to an orthogonal ParD3-ParE3 pair

我们的研究结果表明导致ParD特异性变化的突变途径往往会通过混杂的中间体。因此，我们想确定野生型ParD3-ParE3和正交ParD3* -ParE3*系统之间的突变途径是否也通过混杂的中间体从而改变两种蛋白质的特异性而不破坏它们的相互作用。

因此，我们生成了ParE3变体ParE3*，其保留毒性，但不能与ParD3抗毒素结合。特别地，我们突变了与ParD3中特异性强烈残基的五个ParE3位点（Arg54, Arg58, Ala61, Met63和Leu72或RRAML）。我们突变了RRAML→VEIRF，因为在ParE3同源物中经常观察到这个变体残基其不同于相应的野生型残基。如预期，我们发现ParE3*保留毒性，但不再被ParD3中和（图A）。

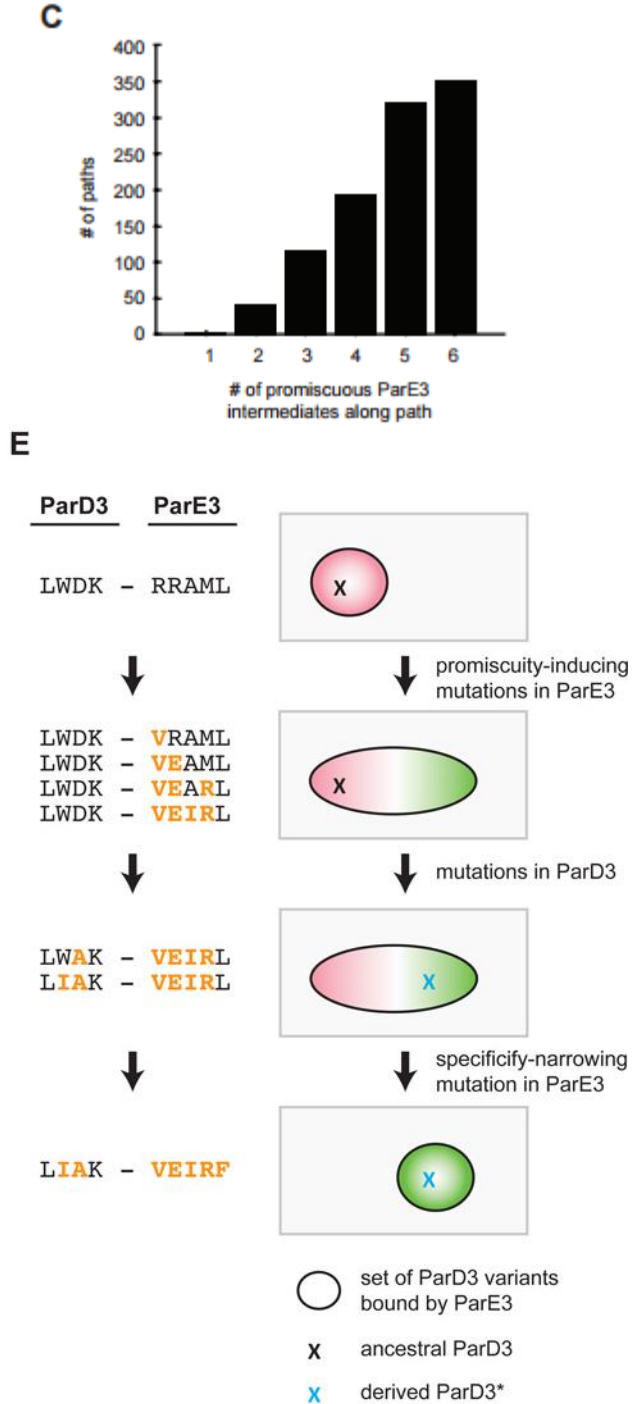


我们生成了包含ParE3 * (32个突变体) 中所有可能取代子集的ParE3变体, ParD3变体包含ParD3 * (4个突变体) 中所有可能的取代子集。然后, 我们将ParD3和ParE3变体(总共128对)的每个可能配对共同转化到大肠杆菌中, 并使用毒性拯救试验(图D)评估相互作用。有趣的是, 128对ParD3和ParE3变体中有90个能够进行相互作用, 可能是因为ParE3变体中的大多数(17/32)变体是混杂的, 我们定义为与ParD3和ParD3 * 两者强烈相互作用(图D)。



为了确定野生型和绝缘ParD-ParE对之间的路径是否倾向于通过混杂的中间体，我们首先列举了这些系统之间的轨迹总数。假设每个步骤有一个残差被改变，并且没有考虑到反转，从ParD3-ParE3到正交ParD3* -ParE3*对有5040条路径;的这些路径，1,030在每个中间步骤保留功能。令人惊奇的是，我们发现，所有这些1,030个功能路径都通过了ParE3的至少一个混杂中间件，每个路径平均有五个混杂的ParE3中间体（图C）。

这些混杂状态的流行可能使得ParD-ParE系统能够轻松地发展新的相互作用特异性。ParE3特异性的初步扩展使得ParD3在序列空间中的移动，随后在最后一步中ParE3特异性的变窄（图E）。相比之下，根据定义，ParD或ParE中的替代产生“切换”特征变化的突变路径将被破坏，直到第二个替换恢复相互作用。因此，我们的结果支持ParD-ParE家族的共同演化和扩张通过混杂的中间体发生的概念。



Discussion

我们通过测试所有可能的ParD3和ParE3突变中间体之间的128对互动来评估野生型ParD3-ParE3和正交ParD3 * -ParE3 *之间的完整的突变轨迹。令人惊异的是，32个ParE3中间体变体中的17个是混杂的，或者能够与ParD3和ParD3 *变体相互作用。因此，ParD3-ParE3和ParD3 * -ParE3 *之间的所有功能路径都涉及至少一个混杂中间体，其中大多数涉及五个以上。因此，我们的结果表明，ParD和ParE的混杂变体在序列空间上是丰富的，并且使用混杂中间体使能突变可以促进新的相互作用特征的演变，同时仍然使用相同的特异性残基。

类似的原理可能适用于整个生物学中的其他蛋白质 - 蛋白质相互作用，甚至不涉及有毒蛋白质的蛋白质。给定的蛋白质 - 蛋白质相互作用的破坏可以阻止执行必需的细胞功能或导致与另一种蛋白质的不必要的，有害的相互作用，从而有利于在每个步骤保持功能的共演化轨迹。

谢谢聆听