**工作学习内容：**

工作：进行队伍管理、人员安排，参与讨论HP相关事宜，进行项目构思。

学习：学习了Microbiology overview，Virology 5, 6, 7章（主要第七章 流感病毒），浏览了部分免疫学相关文献，后转变方向，研究2017年iGEM Heldelberd项目。

**学习内容分享：**

海德堡大学17年项目做的是加速噬菌体进化以做到——“使噬菌体 体内定向的蛋白质进化更快，更容易，更强大，并扩大其应用范围”的目的，此项目基于PACE(噬菌体持续辅助进化)，然而PACE具有很多缺陷：

一次偶然的机会，我们有幸会见了PACE的共同发明人Kevin Esvelt，他于2017年6月1日即在我们开始湿实验室工作之前在海德堡进行了演讲。凯文（Kevin）很友好地参观了我们的团队，并详细讨论了他的PACE方法。除了具有定向蛋白质进化和概念美的功能外，我们还确定了当前PACE设置的两个主要限制：

1. 它需要一个复杂的，定制的生物反应器（图1），其中包括复杂的流量控制装置，由于存在多个故障点（潜在的噬菌体污染；如果流速太高，则有可能被噬菌体冲刷的风险），组装起来极具挑战性且难以稳健运行；宿主细胞生物膜形成（适当的噬菌体进化）
2. PACE几乎完全限于改善已经存在的蛋白质功能。为了发展真正新颖的功能，需要进化的垫脚石，到目前为止，只有通过缓慢地将选择压力适应待发展的活动才能创造出这种垫脚石。[1]。实际上，这意味着不断发展的新功能需要极其复杂的实验设置，其中包括多个按天顺序顺序应用的微调合成选择电路。在大多数情况下，从头开始发展一种新颖的功能仍然是几乎不可能在PACE中实现的任务。

他们的整体项目是基于弥补PACE系统的缺陷所建立的，此项目生物实验和硬件部分改善了实验条件和环境，证明了他们所设计的项目的确行之有效并且可以投入使用。

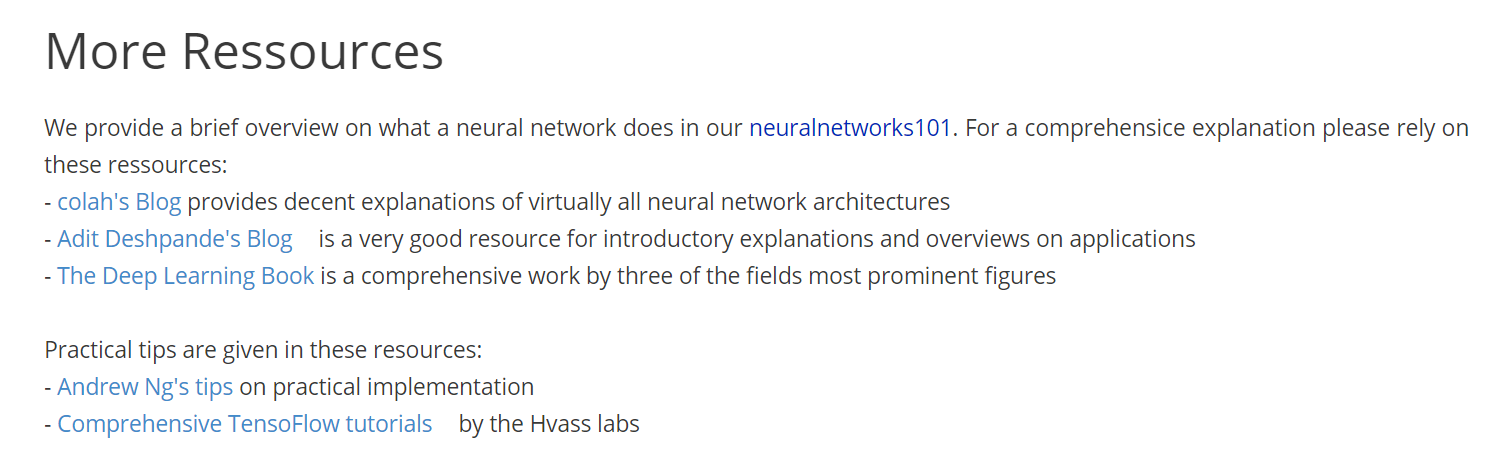
我关注的项目重点在于**计算机软件部分**，他们的设计并非想要进化噬菌体，而是利用PACE系统(噬菌体的高突变性)，实现噬菌体内某几种蛋白质的进化。为了在现实中直接进化出想要的蛋白质，省去大量随机突变的时间，得到所谓进化的垫脚石，他们在计算机里设计了一套庞大的基于神经网络和深度学习的系统，用于模拟并筛选出理想的蛋白质(主要是酶活性)的进化方向。

他们首先基于深度学习、卷积神经网络设计了一套训练方式（DeeProtein 系统），用数百万个蛋白质的序列及其功能信息对算法进行训练，以基于ResNet架构将蛋白质序列分类为功能，进行蛋白质表征学习，目的是为了得到一套算法：我们输入一段蛋白质序列，计算机可以对此序列评分，计算出该蛋白质的功能性究竟如何。

此外他们又设计了一套GAIA系统，来进行将已有的蛋白质进行随机突变（其实是有目的性的随机突变），与上述提到的DeeProtein系统对接，为DeeProtein提供序列材料（进行分子进化时使用，并非训练）。

GAIA是负责序列突变和选择的遗传算法，它与DeeProtein（提供功能蛋白序列分类的预训练深层神经网络）接口。GAIA 通过迭代随机氨基酸取代，然后通过最大化DeeProtein分类得分来选择所需目标功能，实现了计算机指导的蛋白质进化。通过对DeeProtein进行微调以适应特定的进化任务，可以对GAIA进行优化，以实际上促进给定输入蛋白质序列上的任何功能转移。因此，GAIA通过纯计算的方式为成功进行定向进化实验提供了必要的进化垫脚石。

它们提供了一些指导项目的深度学习、神经网络、序列评分分析的算法教程：



**想法：**

海德堡大学项目提供的思路：

1. 我们所设计出来的项目(软件)，一定不能是为了设计而设计的，它需要有自己独特的功能，它的存在可以帮助科学界(或IGEM界)获得提升，因此项目设计需要满足两点其一：

如果说我们设计的是一套全新的系统，前人没有做过类似的事情，我们需要保证这套系统在限定的大环境下发挥起到它应有的作用。

如果说我们的项目是做前人已经进行过的研究，我们需要基于前人没有做到的部分，去进行补充，此过程中必须体现出我们项目设计的优越性、不可替代性（比如在项目设计中，咱们设计一款软件，我们几乎需要检视现有所有类似的算法、软件，加以评估，得出咱们软件的优越性）。

1. 我们可以借鉴海德堡大学的软件这块儿的思路、甚至是算法，他们非常完美、清楚地（比任何一个软件队项目都要完美）介绍了软件及算法的设计思路及每一个组成部分。一来可以成为我们后来项目工作内容、wiki编写的模板，二来也许我们真的可以实打实的用到他们的一些算法、思想，比方说设计一套类似的基于神经网络评分系统，针对病毒某些关键酶的功能或者是病毒表面蛋白的结合效率进行评分，从而设计一套智能评测系统，分析某类病毒的毒性、传染性。