

Expore the interaction proteins of rice HDAC1 by bioinformatics

College: College of Life Sciences

Subject : Life sciences and Biotechnology

Name : Jieqing Chen

Directed by : Pro. Lijia Li

ABSTRACT

The continuous development of biotechnology and information technology are booming in 21st century, proteomics research has become an important direction to explain life phenomena. Gene expression, gene interaction, phenotypic traits, and trait inheritance are all based on PPI (protein-protein interaction). Since the introduction of bioinformatics, finding the potential combination between different proteins has been selected as an indispensable part of protein research. A fast, efficient and easy to operate analog matching system needs to be established. Appeared in the process of the development of computer technology and a group of virtual matching based on molecular docking software, in view of the different protein libraries all over the world, using genetic annealing, the SVM vector machine, and so on way of computing means has become more universal, which has been recognized by researchers all over the world. But this method has some limitations, simple preliminary screening cannot undertake further research and comparison, and because the protein database cannot cover all known protein, some structures cannot be determined. So, this application could make some mistakes, and needs to be optimized according to other research facts. Rice histone deacetylase I (OsHDAC1) is proven closely connected with the root growth, but the PPI in this regulating pathway is not yet clear. In this article, with the help of AUTODOCK and documents from Swiss and other protein structure database, we try to predict OsHDAC1 and BR signal factor OsBZR1 mode of interaction through the structures of several proteins that have been shown to interact with HDAC1, and use BiFC (bimolecular fluorescence complementation) to confirm the above conjecture.

Key words: OsHDAC1 ; OsBZR1 ; Root growth ; Bioinformatics ; Structure prediction

目 录

1 绪论	1
1.1 研究背景	1
1.1.1 人类基因组计划的发展	1
1.1.2 生物大分子的相互作用方式	1
1.1.3 虚拟筛选	2
1.1.4 生物信息学	2
1.1.5 研究现状	3
1.1.6 水稻 HDAC1 蛋白	4
1.2 课题内容和意义	4
2 分子对接	5
2.1 理论基础	5
2.1.1 分子对接的模型发展	5
2.1.2 Autodock 软件与遗传算法	6
2.1.3 自由能函数	6
2.1.4 数据库介绍	7
2.2 Autodock 的对接实验	9
2.2.1 实验基本流程	9
2.2.2 确定目标蛋白	10
2.2.3 受体与配体的优化	13
2.2.4 活性序列的 Blast 对比	16
2.3 分子对接总结	17
2.3.1 Redocking	17
2.3.2 最佳对接模式的确定	18
2.3.3 对 BZR1 蛋白活性位点的猜测	19
3 BZR1 与 HDAC1 的相互作用	20
3.1 BZR1 蛋白	20

3.2 双分子荧光互补实验	20
3.2.1 双分子荧光互补	20
3.2.2 实验流程	21
3.2.3 实验结果	26
结论	27
参考文献	28
致谢	31
附录	32

1 绪论

1.1 研究背景

1.1.1 人类基因组计划的发展

人类基因组计划时由美国政府在 1984 年提出的一项与阿波罗登月计划等齐名的项目，于 1998 年正式启动 **Error! Reference source not found.** 其最终目标是完成人类核基因组的全部测序工作，并从物理特性以及功能组成等多个方面鉴定分析并绘制所有的基因图谱。政府资助的测序大部分在美国，英国，日本，法国，德国，西班牙和中国的二十多所大学和研究中心进行[2]。

2006 年 6 月份，人类基因组计划已经完成了覆盖率达到 97%的“工作框架图”，各国科学界于 2003 年 4 月正式宣布人类基因组图谱已经绘成[3]，自此，全世界的科学家便可以用这一成果进行社会学，生物学，伦理学的研究，甚至可以建立起庞大的蛋白质，DNA，RNA 库[3]。据测定，人类基因组大约包括 3.5-4.5 万个基因，而由这些基因产生的可编码的蛋白质的数量也在 400 个到 3000 个之间。在生命活动的进行中，这些蛋白质承担着重要作用。而对这些蛋白质结构解析已经成了目前蛋白质组学，生物信息学以及结构生物学的重要研究方向。

1.1.2 生物大分子的相互作用方式

在复杂的生命体中，大分子的种类繁多，其相互作用方式也多种多样。蛋白质与核酸，酶分子和底物，受体与刺激物等等，在这些分子的相互作用中，离不开结构的相互配合以及作用力的相互吸引。而在这些相互作用的背后，是一个个生命具体活动，从消化，呼吸，吸收到分泌激素，重吸收尿液等等。而蛋白质之间的相互作用作为最普遍也最直观的相互作用也一直时研究者们研究的重点。

在人类基因组计划完成以后，蛋白组计划开始被广大科研工作者所熟知，大

数据的发展，使得不同蛋白质结构的共通以及快速的筛选成为可能。根据 Boguski[4]的建议，已经注释过的单个基因功能应该定量化，以适应生物大数据的发展。这种基于数据挖掘，机器学习以及神经网络算法的动态模式也渐渐成为主流。与此同时，这种理论给蛋白质相互作用的方式的准确预测提供了基本的理论依据。

蛋白质-蛋白质相互作用（PPIs）是两个或多个蛋白质分子之间建立的高度特异性的物理接触，这是由静电力（包括疏水作用）驱动的生物化学事件[5]。许多是与特定生物分子的特殊结构发生在细胞内的物理化学接触。蛋白质往往需要与其他分子进行相互作用，这些相互作用构成了生物体内的所谓互动组学。相对的，异常 PPIs 是多种疾病的基础，如克雅氏病，阿尔茨海默氏病，并可能导致癌症。

目前，不同的角度研究 PPI 的方式有：生物化学，量子化学，分子动力学，信号转导等。这些信息的测定使得创建与代谢或遗传以及表观遗传网络类似的大蛋白质相互作用网络成为可能[6]。

1.1.3 虚拟筛选

在药物研发的各项流程中，目前利用计算机辅助进行大规模筛选的主流方法由基于小分子特定结构和活性残基的药物设计（Ligand-based drug design, LBDD）和基于大分子的药物设计（Structured-based drug design, SBDD）组成，这类方法通过受体分子的三维结构建立起受体-配体的特定配对方式，通过计算结合过程中发生的配对相互作用来所判断结合难度。而在生物大数据的发展下，虚拟筛选也逐渐成为目前用于大规模药物筛选的主要手段之一，而虚拟筛选则是上述两种方法的进一步延伸，目前，虚拟筛选正日益发挥出巨大的潜力。

1.1.4 生物信息学

生物信息学的概念在 1987 年提出，是一门研究生物系统和生命过程中的数据信息流的整合学科，在生物统计，计算机科学以及定量生物学和计算生物学的众多分支中，生物信息学都起到了不可或缺的桥梁和整合作用。在计算机科学的

角度,生物信息学利用自动化的分析程序将生物数据分割成独立的有意义的片区,为基因功能和蛋白质功能的解释提供了有利依据,在药物研发的角度,生物信息学将不同种类的蛋白进行虚拟配对,自动分析,大大缩短了药物研发的时间,提高了靶标的准确性,在生物进化的角度,生物信息学对遗传系谱以及遗传树的分析,将本来复杂无法确定的演化历史直观地有效地反映在二维乃至三维的图纸上,系统地解释了生物界演化的分子历程以及宏观历程。与以前相比,生物信息学在大数据时代,也将继续向前发展,为生物学的研究继续开辟更高效更准确的崭新模式。

1.1.5 研究现状

最初概念的提出是在 2001 年,Bock 等人提出了基于氨基酸残基的电荷数量,残基末端的疏水性[7],以及蛋白质序列的特征而获得特征向量的方法,并且采用 SVM 向量机预测蛋白质之间可能的相互作用。

随着生物信息学的进一步向前不断发展,三联体蛋白联合编码越来越成为编码蛋白质序列的主要方式,三联体蛋白序列编码的主要操作逻辑为三步,首先确定氨基酸侧链的极性,根据偶极和体积将目标蛋白分成七组,然后取其中相邻的三个残基为一个测定子集,根据三联体在序列中的分布状况来表示蛋白质序列。此种方式的正确率相较于之前的方法有了较大的提高,可以基本达到 83%

在国外,Agata 等人[7]不依赖于蛋白质序列,改进了基于配体的筛选方法,并且利用改进后的方法对一系列的蛋白质进行了多次大规模的重新筛选,最终得到了多个独特结构,与此同时,在他们的研究中,其实验中发现了可以抑制艾滋病病毒在靶细胞中复制的新化学型。而 Daniel 等人[8]利用已知的配体进行药效团搜索,根据结构相似度的高低进行评价打分。目前,准确率已经基本可以达到 30%的虚拟筛选,相较于传统的高能高通量测序和分子生物学实验,具有更高效更准确的特点。在药物研究中,虚拟筛选越来越重要。

在国内,我国的药物筛选基本与国际接轨,发展较快。全国各地目前由将近 100 多家高校和研究机构建立了自己的虚拟筛选平台,针对市面上的近万中化合物都进行了相应的研究。Wang 等人[9]利用 3D-QSAR 手段对多个 ZINC 数据库

中的乙酰胆碱抑制酶进行三维结构的建模，并得到了近 70 个亲和力高的双位点抑制分子。同样的，Li 等人[10]利用机器学习和支持向量机的方法对 ZINC 数据库的多种蛋白子库进行虚拟筛选，也获得了一系列潜在的药物。

计算方法和虚拟筛选在蛋白质相互作用以及药物筛选的应用已经非常广泛而且有效，不仅如此，在可靠性方面，由文本挖掘，高通量计算为主要的的数据收集方式也保证了数据来源的可靠性和唯一性，相对于传统的化学生物学方法，这样的数据收集更加有效和便捷。在评价方面，由于实际应用时，单单只依靠计算机辅助筛选所得到的结果往往需要实验的进一步验证，但是虚拟筛选的方式为后续实验的设计提供了明确的方向，相对于以前的盲目筛选和全基因组筛选，在进行后续实验前，先进行大规模的虚拟筛选将会是以后研究蛋白质相互作用的主要趋势。

1.1.6 水稻 HDAC1 蛋白

水稻 Os06g0583400 在 2002 年被研究人员正式确定为 OsHDAC1。组蛋白去乙酰化酶主要调节染色质的结构以及基因的转录。OsHDAC1 基因编码约 57kDa 的蛋白质，分别与 OsHDAC2，OsHDAC3，玉米 RPD3 和人 HDAC1 蛋白质分别具有 73.5%，72.7%，79.9%和 57.1%的氨基酸序列同一性[11]。OsHDAC1 在叶，根和愈伤组织细胞中以相似的水平表达，而 OsHDAC2 和 OsHDAC3 在根和愈伤组织细胞中表达，但在叶中不表达，表现出不同的组织特异性。基因组结构和 Southern blot 显示，OsHDAC1，OsHDAC2，OsHDAC3 分别含有 7 个，6 个和 7 个外显子，并且构成了水稻基因组中的 I 型家族。

1.2 课题内容和意义

目前蛋白质虚拟配对的主要研究方向是药物的大规模筛选，这与其多线程多模式处理有关，而在植物学的研究中，因为植物栽培和控制的困难，植物相关蛋白质小分子产品的发展要略逊一筹，植物蛋白的相互作用与动物一样，大规模而且主动虚拟筛选也是研究植物蛋白质相互作用的重要方式，在生物信息学的帮助

下,通过比对已知相互作用蛋白的配对机制来预测未知蛋白的结构正是其中的一个方向。本文将 BZR1、TPR3 两种蛋白与 HDAC1 之间相互作用作为例子,尝试将这一方法进行验证,以及评价这种方法的可能性,为之后的相关实验打好基础。主要包括以下内容:

1. **分子对接** 本文介绍几种流行的分子对接软件及其相关理论,论述目前分子对接的流程以及主要面临的挑战和出现问题。
2. **打分函数和自动配对分析** 在目前已有的分子对接软件中选取成功率较高,易于操作的软件进行分子对接的尝试,通过判断所假定的两种蛋白相互作用的可能性(即亲和程度)的高低,初步尝试利用改进的 SVM 向量机来构建大规模筛选的学习集。
3. **同类型蛋白的结构预测** 在分析了几种已知的互作蛋白后,根据这些蛋白的三维结构对目前仍没有预测结构的蛋白质结构进行初步预测。
4. **实验辅助验证蛋白相互作用** 在基本确定目标蛋白的活性区域以及相互作用可能性高低之后,用相应的分子实验进一步确定实验猜想,并最终以此评价本方法的可靠性。

2 分子对接

2.1 理论基础

2.1.1 分子对接的模型发展

锁钥模型是 1890 年所提出的一种解释酶与底物相互作用的理论。其基本思想是:蛋白质与其特异性的配体之间的结合是刚性的。I.D.Kuntz 在此基础上进行了分子对接的定义[12]。随着计算机的逐步发展,将实验数据的结果转化成可操作的,规律的数据集成了目前分子对接的重点。因此,在锁钥模型的基础上,分子对接的概念变成了先假定再验证的亲和力计算并打分的方法。

随着分子生物学的不断往前发展,一种更加实用也更加普及的理论被提出来

了,“诱导-契合”模型与锁钥模型的不同在于,这种学说否认了在最初状态下酶便与底物以互补的形式存在,在两者结合的过程中,酶蛋白会发生构象变化从而适配不同的底物或者以构象的变化来获得更加稳定的结合。

因此,基于上述的两种结合方式,分子对接的方式一般大抵可分为半柔性对接,刚性对接,和柔性对接。考虑到本文利用的蛋白的一般作用方式,而在本文中,主要关注的对象均为刚性对接和半柔性对接。

2.1.2 Autodock 软件与遗传算法

AutoDOCK 是基于 C 语言的半柔性对接软件,主要采用遗传退火算法来进行分子结合力的优化以及寻找合适的最低能态构象[13]。

在计算机科学和运筹学研究中,遗传算法是一种随机的搜寻算法,是模仿达尔文生物进化中的自然选择和遗传学机理寻觅最优解的算法。遗传算法将所要解决的问题的答案视为某种优势种的染色体,在一系列假设的结果中,利用生物进化的原则选出最优解。主要由这几个步骤组成:编码,群体设定,适应度函数的确定,遗传算子的设定。与其他方法相比,这种方法所针对的对象是编码后的参数,适用于大规模复杂性较高的计算[14]。

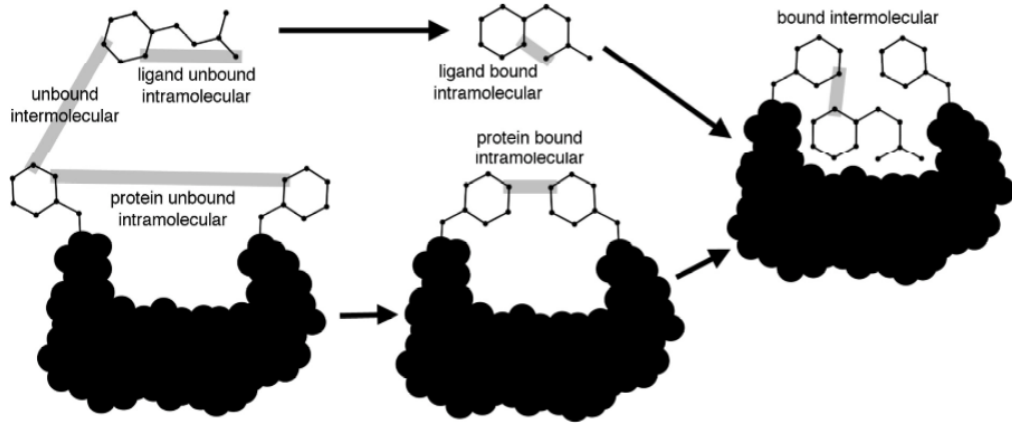
每个候选的遗传算法的解决方案的标准表示形式是一个位数组。其他类型和结构的数组可以以基本相同的方式使用。遗传算法便利表现的主要共性是它们大都固定大小,易于对齐,利于简易的穿插操作。

遗传退火算法是一种广泛应用于自然科学,经济学,物理工程学等一系列学科的算法,该算法基于概率学。在退火过程中,物体的温度上升时,内部分子的运动加速,内能升高,之后进入降温途径,粒子速度降低并最终趋于平稳,能量也降到最低值[15]。

2.1.3 自由能函数

本文采用的 Autodock 软件使用的是一种基于经验的估算模拟。这个估算模拟的自由能是经过大量的蛋白质重配对和重拟合而得到的,而且这些重新拟合的

抑制常数已知。



这样的自由能结合立场用于评价亲和能力的高低通常分为两步：在第一步中，从未结合状态过渡到结合状态时估算配体和受体中分子内能的变化，第二步则是利用处于结合模式下的两个蛋白之间的彼此作用能来初步估算分子内能的值 [16]。

其熵变算法以及结合能算法如下：

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf})$$

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{e(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)}$$

2.1.4 数据库介绍

目前的蛋白质数据库主要集中于蛋白质内部的氨基酸序列，以及对应的基因序列。部分数据库有完整的已经被证实和解析过的蛋白质三位结构，另外的一些数据库则是提供根据蛋白质功能和氨基酸序列给出推测的三维结构，这种结构虽然不是完美符合实际情况，但是在大规模的数据处理下，部分高相似度的分子模型是可以被采用的。以下是本文所用到的几种蛋白来源的数据库介绍。

1. **PDBbind 数据库** 是一个完整的利用亲和力数据测量蛋白质之间相互作用的数据库。收录了大多数蛋白质的空间结构，DNA，RNA 以及配体之间的相互作用数据，首次公开于 2004 年 5 月，当前版本为 2014 年版。

2. **String 数据库** 数据库的容量要低于 PDBbind, 但是该库的一大特点是提供了网状图和树形图辅助研究人员进行多种蛋白质大分子以及功能性多肽的相互作用分析, 并且提供了多种处理端口, 以及预测的相互作用数据。以颜色的区别进行相互作用可能性高低的得分。是本论文实验数据的主要来源之一。
3. **Swiss-prot 蛋白结构库** 是一个带有注释的蛋白质序列数据库, Swissport 与其他蛋白数据库的不同之处在于其采用的三个不同标准: 注释, 最小冗余, 与其他实验室具有对应关系。在本文的结果处理以及后期的对比序列中有所应用。
4. **Uniprot 数据库** Uniprot 是一个可自由访问蛋白质序列、功能、相互作用、基因序列等一系列信息的数据库, 主要数据来源于各国学者的研究文献。是连接蛋白质与基因, 查询原始文献的中转数据库。Uniprot 主要提供四个核心数据库: UniprotKB (包括了 swiss-prot 和 TrEMBL)、Uniparc 和 UniRef。
5. **SC-PDB 数据库** SC-PDB 是在蛋白质数据库 (PDB) 中发现的 6415 个结合位点的三维结构的集合。该数据库可从所有高分辨率晶体结构中提取结合位点用于鉴定和评价蛋白质腔和小分子量配体之间的复合物的稳定性。重要的是, 在此数据库中, 配体是从药理学角度考虑而不是从结构角度考虑。因此, 溶剂, 洗涤剂 and 大多数金属离子往往不会被储存在 SC-PDB 中。在 SC-PDB 中, 能成为配体应具备以下条件:
 - a) 小质量分子或者核苷酸 (<4-mer), 肽 (<9-mer), 辅因子和有机化合物。
 - b) 相应的结合位点由距离配体原子的 6.5 埃以内的一个或多个原子的所有蛋白质残基 (包括氨基酸, 辅因子和重要金属离子) 形成。
 - c) 通过非共价的方式结合。

SC-PDB 可用于多种目的: (i) 筛选用于预测任何配体的最可能靶标的结合位点的集合, (ii) 分析不同空腔之间的分子相似性, 和 (iii) 导出规则描述配体药效学点和活性中心性质之间的关系。

2.2 Autodock 的对接实验

2.2.1 实验基本流程

本文从筛选已知的相关蛋白质的三维构型出发，通过匹配目前主流数据库中的相关蛋白质，在软件辅助下进行一系列的批量筛选。基于配体-受体相互作用理论，采取以下的实验方式确定目标蛋白的亲和力大小[17]：

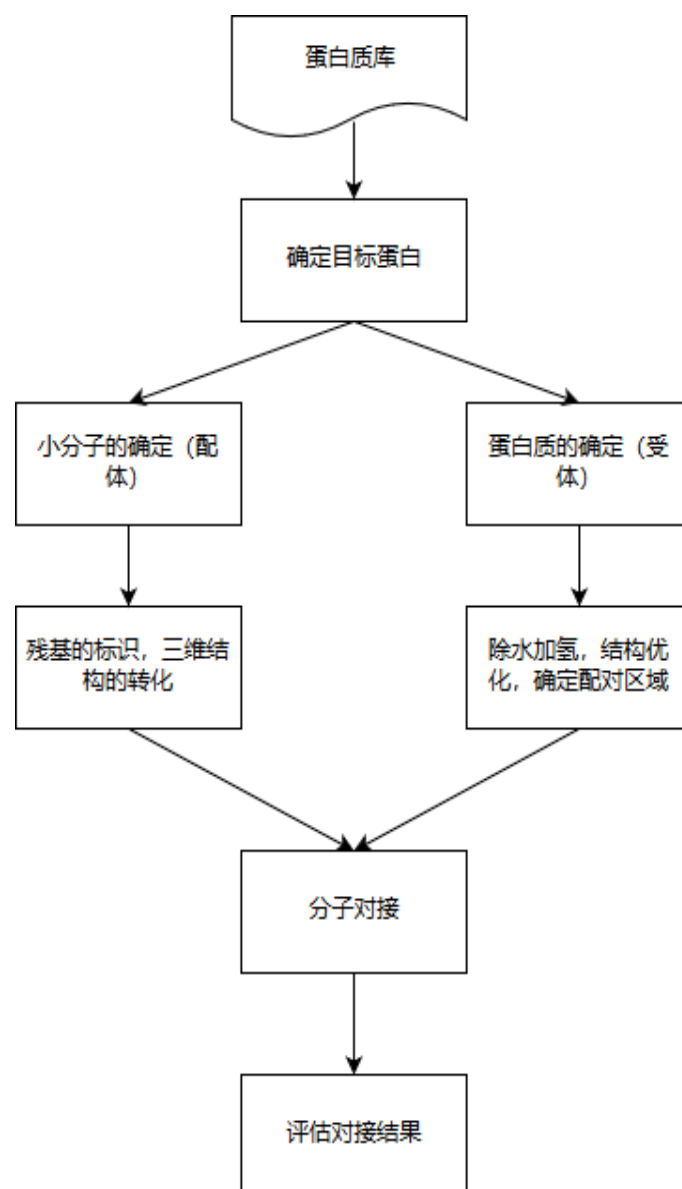


图 2.1 实验基本流程

2.2.2 确定目标蛋白

1. String 筛选

利用 Huynen 等人的方法[18], 本文选取了 string 数据库中与 OsHDAC1 紧密相连的 7 个蛋白及分子进行对接实验。各自的亲和作用相对高低见表 2.1, 构型及基本功能特点见表 2.2, 结构统计见表 2.3。

表 2.1 与 OsHDAC1 相互作用蛋白的亲和力相对值表

node	相对亲和力
4330479	0.97
4324980	0.956
4326133	0.953
4339773	0.953
4337520	0.953
4341387	0.953
4333540	0.945
4341387	0.945
4327826	0.945

表 2.2 各个蛋白质的构型及基本功能

#node	构型/基本功能
4326133	转录抑制因子, 表达蛋白
4327826	WD 结构域, 包含 G-beta 重复序列
4324980	ZOS1-20 - C2H2 锌指蛋白
4330479	SET 结构域蛋白
4333540	包含蛋白质的 WD 结构域, G-β 重复结构域
4337520	转录抑制因子
4339773	表达蛋白
4341387	组蛋白脱乙酰化酶
LOC_Os08g25200.1	两亲性螺旋重复蛋白
4347727	包含蛋白质的 WD 结构域, G-β 重复结构域
4351981	ZOS12-04 - C2H2 锌指蛋白

表 2.3 String 提供的亲和蛋白结构特征统计

#pathway ID	结构描述	蛋白统计	阴性概率
IPR003822	两亲性螺旋重复结构	3	5.36E-07
IPR001680	WD40 结构域	3	0.00731

IPR017986	WD40 重复结构域	3	0.00731
IPR015943	WD40/YVTN 重复/类似结构域	3	0.00898

2. WD40 结构域

WD40 重复序列（通常又称之为 WD 基元或者 β -转导蛋白重复序列）是一个约有 45 个氨基酸的较短序列，通常以色氨酸（Trp）-天冬氨酸（ASP）（W-D）组成的二肽为末端[19]。具有这样重复序列的分子通常会折叠在一起形成一种被研究者称为 WD40 结构域的环状螺旋蛋白结构域。

含有 WD40 结构域的蛋白质通常具备 4 到 16 个重复的结构单元，这些结构单元可以形成环化的 β 螺旋结构。WD40 结构域一般由开始约 20 个残基的可变区和接下来的一系列重复残基簇构成。这些重复的结构通常会形成四股反向平行的 β 折叠片。这些 β 折叠片的组合一起构成螺旋结构，其中，7 个 β 折叠片互相锚定，使得一个重复区域的最后一个 β 折叠片与下一个重复区域的前三个 β 折叠片形成 3D 叶片结构[20]。

WD40 重复结构域是在所有真核生物中均有发现的大家族，其功能从信号转导和转录调节到细胞周期控制，自噬和细胞凋亡等都有涉及。WD40 结构域具有重复保守的特点，可以作为蛋白质与蛋白质相互作用的刚性支撑，并且还可以进一步协调蛋白质与蛋白质之间的相互关系[20]。

从表 4 中的数据可以初步看出，WD40 结构域的存在对于 HDAC1 的蛋白相互作用起到一定的作用，文献报道[11]，HDAC1 蛋白质因为其具有分子基团大，电荷量多的特点，在与其他蛋白结合时，通常需要保持其刚性结构的稳定。由此可初步推断，WD40 结构域在 HDAC1 蛋白与其他蛋白发挥相互作用时发挥了一定作用，之后的实验也以此为基础，着重分析具备 WD40 或者类似结构域蛋白与 HDAC1 的相互作用。

3. TPR3 蛋白

Tang 等人关于 OsZIP46 与 MODD 的相互作用的研究表明，MODD（OsZIP46 失活和降解的介体）与拟南芥 ABSCIIS-ACID-INSENSITIVE5 结合蛋白 AFP 同源。而且可以通过与 OsTPR3-HDA702 共抑制因子复合物的相互作用和 OsZIP46

靶基因的组蛋白乙酰化水平的下调来抑制 OsZIP46 的活性[21]。在该项研究中，双荧光素酶报告系统显示 HDAC 抑制剂 TSA 与 MODD 相接触后，后者的转录抑制效应降低至 30%，即 MODD 的转录抑制效应很大程度上取决于 HDAC 的活性。而且该研究使用 Cop-IP 实验证实了 OsTPR3 和 HDA702 之间的相互作用，表明 OsTPR3 将 MODD 与 HDA702 连接起来，在整个反应中起到桥梁作用，且主要活性部位位于在 N 末端。Cop-IP 结果如下图。

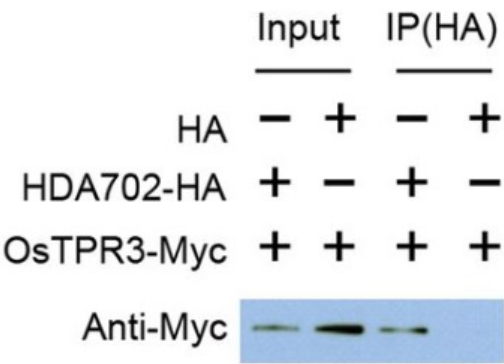


图 2.2 HDAC1 与 TPR3 的免疫共沉淀实验[21]

本文主要讨论 HDAC1 与相关蛋白的相互作用，目前已经解析的 OsTPR3 的相似蛋白及其相似度见图 2.3

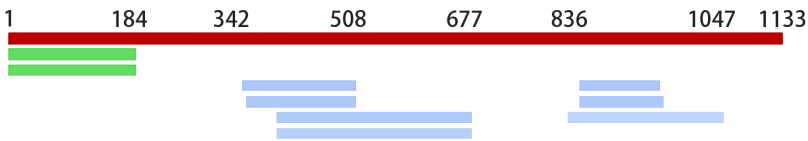


图 2.3 PSI 中 OsTPR3 全序列的结构模型匹配

在 PSI 中，OsTPR3 总共由 1133 个氨基酸构成，其中部分序列的三维结构与已知蛋白的结构类似。其中 N 末端的结构与 TOPLESS 蛋白的相似度在 93%以上 [22]。其三维结构如图 2.4：

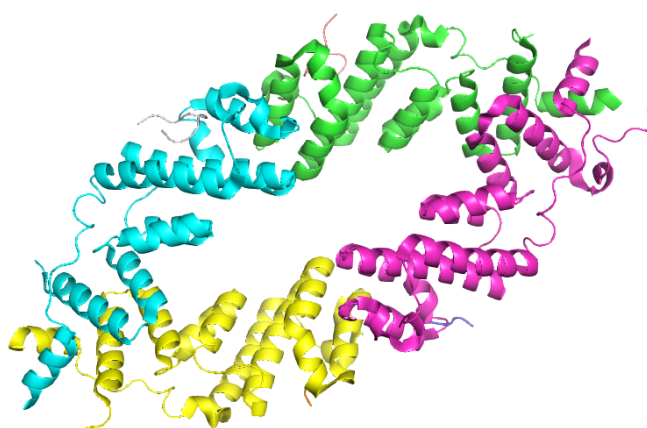


图 2.4 OsTPR3 的 N 末端的模拟结构，相似一致性 93.9%

2.2.3 受体与配体的优化

对于两种靶蛋白，本文基于 SC-PDB 的分类标准，在 HDAC1 中截取活性残基作为小分子配体，以其他蛋白作为配体进行后续的优化操作。其中，HDAC1 的结构以及活性残基如下（图 2.5、图 2.6、图 2.7）：

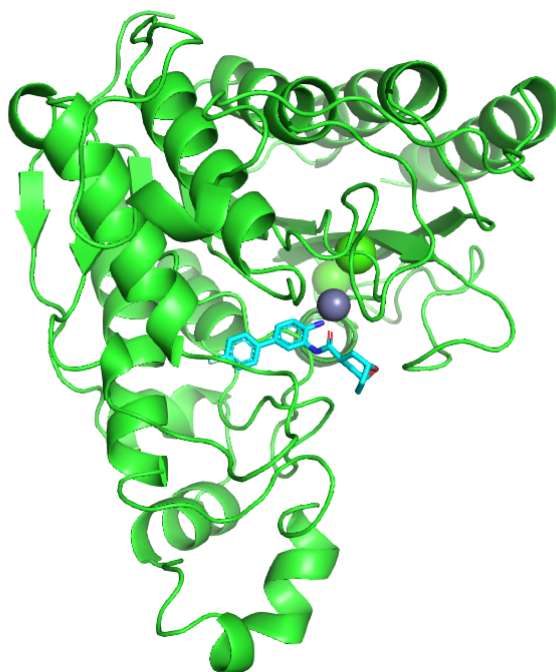


图 2.5 OsHDAC1 的蛋白质结构

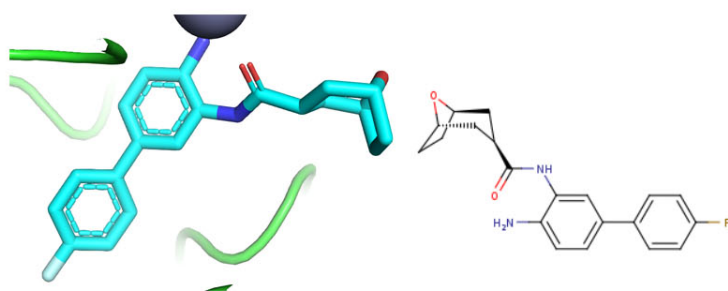


图 2.6 OsHDAC 的活性残基部分（6EZ，在人 HDAC2 里也有存在），具有较强的疏水性

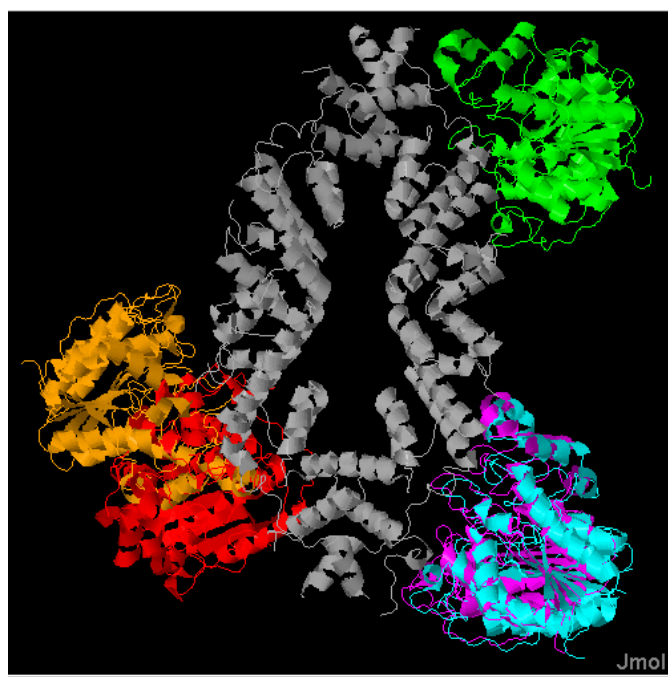


图 2.7 利用 ZDOCK 模拟的高亲和力的对接构象（不同颜色代表不同亲和度的结合位点）

上述 OsHDAC1 蛋白构型及其活性残基的来源为 PDB（Protein data bank），分辨率为 2.53Å，PDB 数据库之前已经介绍过，作为一个大分子的晶体结构数据库，其主要数据的来源是各国研究者通过核磁共振，X 射线晶体光谱，低温电镜

等手段得来。此结构为直接实验结果，而非 TPR3 活性部位的模拟结构，故此结构可以认为是十分精确的，通过 PyMol 软件的图形界面，可以很直观地确定有效的活性部位，在不考虑其他形式的相互作用的前提下，该活性部位唯一的且亲和力最强。

靶蛋白选用 TPR3 的 N 端结构，该结构来源于 PSI (Protein Model Portal)，N 端前 184 个氨基酸已经囊括了在文献中报道的 TPR3 活性部位，选用的相似蛋白为 PDB 中编码为 5nqv 的蛋白质，属于拟南芥的 TOPLESS 蛋白的 N 末端，其分辨率为 1.95Å。

配体于受体的自动对接使用 AutoDock 软件的 autogrid 和 autodock 模式，其中蛋白质通过 PyMol 去除水分子，并且优化氢原子的位置。在 Autodock 中，计算出配体的电荷数用于后续的退火算法以及序列分析，确定对接的范围为受体蛋白的活性残基部位，根据结果来看，总共有 50 个可能的位点，这里选取结合能 <-3.0 的前 19 个位点作为目标位点。其相应的结合能如下：

表 2.4 TPR3 与 HDAC1 的位置结合能分析 (binding energy<-3.0)

位点标记	binding energy	refRMS
1	-3.95	34.16
2	-3.84	47.95
3	-3.77	30.3
4	-3.66	5.4
5	-3.61	20.06
6	-3.57	7.57
7	-3.49	20.89
8	-3.42	40.13
9	-3.38	43.14
10	-3.33	9.43
11	-3.32	32.95
12	-3.25	21.93
13	-3.2	25.46
14	-3.17	6.36
15	-3.16	2.58
16	-3.14	92.41
17	-3.12	28.17
18	-3.09	23.49
19	-3.08	49.9

从表中的数据来看，大部分的残基配对亲合能都在-3 以下，重配对 Redocking 的结果显示，前后两次对接的差值为 1.4，低于 2 这个经验准确 cut off 值，故以上结果可以认为代表了与 HDAC1 相互作用的蛋白的基本亲和情况，在 Autodock 的使用手册中，对于亲和能的大小没有详细的规定，据前人的研究表明，Autodock 在第四个版本更新中放弃了之前广泛使用的模拟遗传退火算法，故本次实验所得到的结果在新一版本的 Autodock 中可能有一定的出入。。

2.2.4 活性序列的 Blast 对比

1. Blast

NCBI 中的序列局部比对搜索工具（BLAST）是生物信息学分析中的一个有力工具，用于找到序列之间局部相似性的区域[23]。这个程序会自动利用已知的蛋白质与蛋白质之间的相互作用数据计算目标蛋白序列的统计显著性。BLAST 还可以用来推测序列与序列之间的功能和进化关系，进一步来说，通过对进化关系的分析和研究，利用 Blast 便可以鉴定不同基因家族的具体组成成员。目前的 BLAST 工具主要有 4 种，分别为 tBlastn，Blastx（主要用于蛋白和核酸的之间的序列比对），Blastp（蛋白和蛋白之间氨基酸序列比对），Blastn（用于核酸序列于核酸序列之间的序列对比）。

2. TPR3 的 Blast 结果

利用 NCBI 提供的 Blastp 程序[24]，本文对 TPR3 全蛋白序列进行了序列匹配分析，其结果如图 2.8。

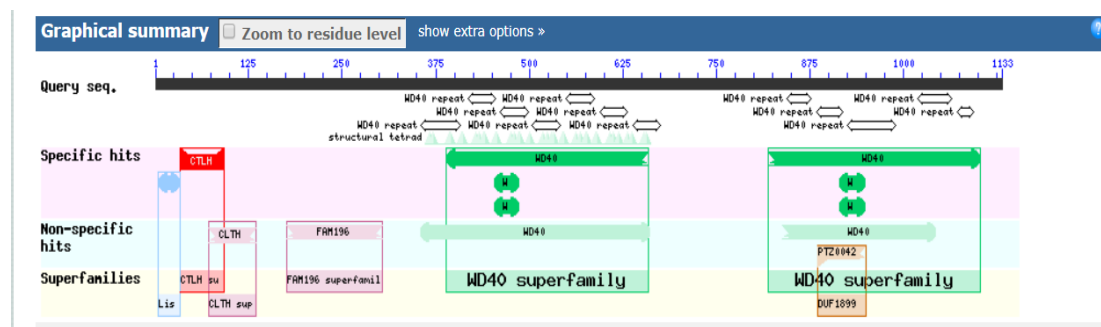


图 2.8 TPR3 的 Blast 结果，其中 300-1100 的氨基酸片段含有两端高度保守的 WD 重复序列

与此同时，在 TPR3 的 N 末端的 184 个氨基酸残基中，发现了与其类似序列的相关蛋白 BZR1。在目前的实验中，对三级结构的相似性是目前验证结论的困难所在。由于目前的主流数据库中，对于水稻 BZR1 蛋白的解析程度不够，推定的三级结构模型的序列一致性不到 20%，所以仅仅凭借目前的分子对接软件无法准确判断两者的相互作用。本文尝试从一级结构侧面推测其与 HDAC1 的相互作用。

BZR1 与 TPR3 的活性位点的序列匹配如图 2.9。

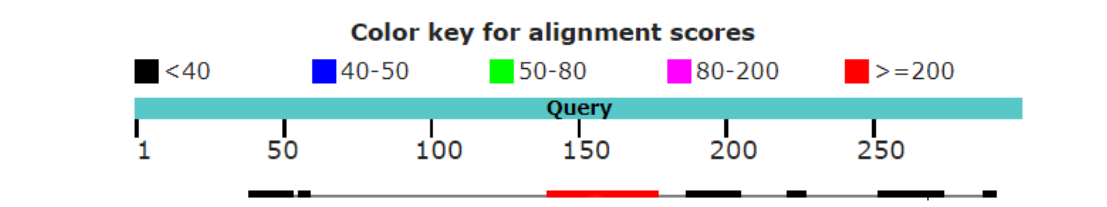


图 2.9 TPR3 与 BZR1 的 Blast 分析，其中第 150 个氨基酸左右的残基具有较高的相似度。(阈值设定为 50，E value 值为 10^{-6})

根据以上实验结果，可以推测 BZR1 与 HDAC1 发生相互作用的可能性较大，而 BZR1 本身不含有保守的 WD40 氨基酸序列，故可能存在其他含有类似 WD 结构域的辅助蛋白帮助 HDAC1 与其结合。

2.3 分子对接总结

2.3.1 Redocking

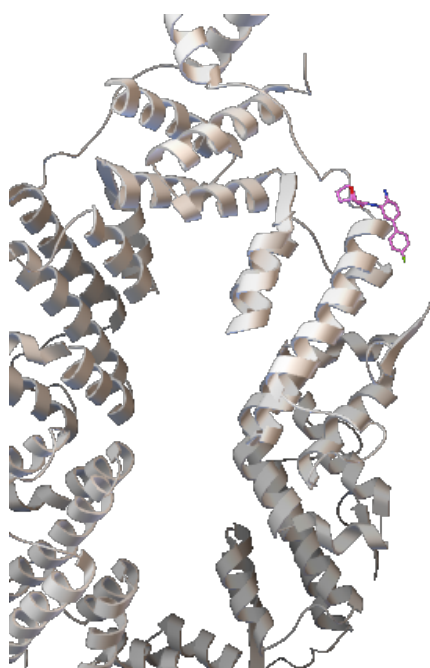
目前的分子对接技术的难点在于各种分子晶体的结合模式多种多样，而初步确立对接可靠性的方法是 redocking[25]，即将成功对接的复合物的配体拆出，先生成三维结构构象，再使用对接技术将其重新对接入配体，根据对接结果的结合模式推断之前对接的可靠性，本文所使用的 HDAC1 活性部位与 TPR3 重新对接后的 RMSD 差值最高为 1.4，一般来说，差值小于 2.0 被认为时可靠的。故本文提出的活性残基与 TPR3 的对接结果是可靠的，即表示所选取的活性残基有很大

概率为目标蛋白的结合残基。

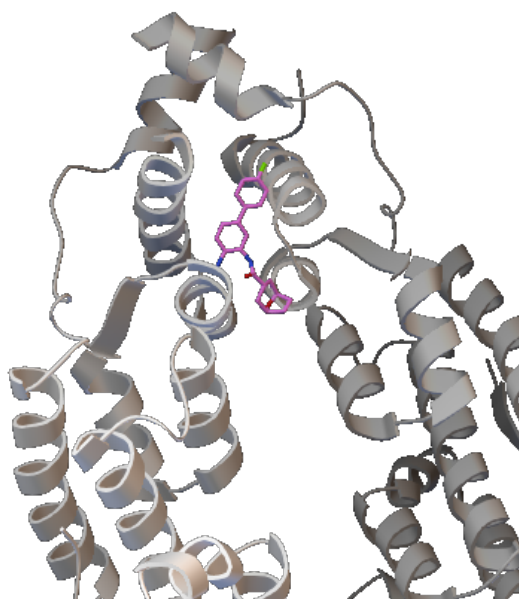
2.3.2 最佳对接模式的确定

基于之前的配对分析,在 ZDOCK 对 TPR3 和 HDAC1 全蛋白的对接预测中,高亲和力的结合部位位于四周,这与 Autodock 的结构类似,在排除其他干扰位点后,最终确定的对接位点有如下两个:

1. 位点标记 4



2. 位点标记 6



以上两个位点综合根据结合能的大小和分子空间位置给出[26], Autodock 的结果可靠性依赖于经验判断和分子实验的辅助,故本文得出的两种构型还需进一步的实验验证。蛋白质的解析依赖于生物化学和结构生物学的发展,蛋白质之间的相互作用可以为分子结构的解析提供侧面旁证[25]。本文将 HDAC1 和 TPR3 蛋白的结合为例,提出了一个从活性残基入手,并最终推断结合位点的方法。在以后的进一步研究中,尝试用打分函数和 SVM 向量机进行大范围大规模的筛选,以本次的结果作为学习集的一部分可以大大加快虚拟筛选的准确性和速度。

2.3.3 对 BZR1 蛋白活性位点的猜测

根据两次 Blast 的结果, BZR1 的蛋白质序列与 TPR3 的前 200 个氨基酸有着很大程度的相似性,可以认为 BZR1 的活性位点也具有与 TPR3 蛋白相似的特征。而且 WD40 结构域的存在也似乎暗示着其他蛋白对 HDAC1 的辅助作用,在水稻组蛋白去乙酰化的调节通路中,具有 WD40 重复结构域的蛋白质的数量不少,作为组蛋白去乙酰化途径中的重要的支持作用蛋白,这样类似与 WD40 结构的蛋白质可以为以后与 HDAC 系列蛋白质相互作用提供依据。而 BZR1 本身所不具备的 WD40 结构域的特性也从侧面证明,在蛋白质的相互作用中,两种蛋白结合形成的共聚物也可以在某些特殊分子的帮助下发挥重要的生物学作用。这也为之

后进一步准确确定 BZR1 蛋白的功能特点提供了实验依据。

3 BZR1 与 HDAC1 的相互作用

3.1 BZR1 蛋白

OsBZR1 是油菜素内酯信号传导的下游信号分子[27],其可以在植物生长和发育的过程中调节多种激素的合成从而进一步调控植物的生理活动。其结合细胞表面的受体激酶 BRI1 以激活调控核基因表达和植物生长的信号转导途径。有文献报道[28], 通过 RNAi 抑制 OsBZR1 表达可以导致转基因水稻植株的矮化, 使其对 BR 信号的敏感性降低并且改变了 BR 响应基因的表达。前人的种种研究证明了 OsBZR1 在水稻 BR 反应中起到十分重要的作用。

Bai 等人的研究[27]利用酵母双杂交的方法, 筛选确定了 14-3-3 蛋白质为 OsBZR1 相互作用蛋白。且 14-3-3 蛋白可通过减少其核定位直接抑制 OsBZR1 功能。同样的, Qiao 等人在 2017 年的而研究表明 RLA1 / SMOS1 可以与 OsBZR1 相互作用以增强其转录活性。

3.2 双分子荧光互补实验

3.2.1 双分子荧光互补

双分子荧光互补 (也称为 BiFC) 是一种普遍用于验证两种蛋白质能否发生相互作用的技术。被拆分的荧光蛋白片段与大分子复合物组分相连, 推测相互作用的蛋白质可以与荧光蛋白的未折叠互补片段融合, 并在活细胞中表达。这些蛋白质之间的相互作用使不同的荧光片段接近, 使得荧光报道蛋白重新形成其天然的三维结构, 并发射出荧光信号。利用可以反映出细胞中的荧光的倒置荧光显

显微镜可以定位并检测该细胞内的荧光信号强弱与位置。此外，一般认为测定的荧光强度与相互作用的亲和程度成正比。较强的荧光表达水平即表明相互作用较为密切甚至可以直接进行相互作用，较低的荧光水平一般表明复合物内可能存在相互作用。因此，双分子荧光互补实验可以直接通过观察和分析细胞中荧光的强度和分布来鉴定两种或多种目标蛋白质的作用位置和相互作用情况。

3.2.2 实验流程

3.2.2.1. 目标 RNA 的提取

1. 取 0.1 相应材料，置于已经预冷并高温灭菌的研钵中，液氮研磨 1 分钟。
2. 将样品的粉末转入 1.5 ml EP 管中，并加入 1 ml Trizol 溶液，充分混匀，于室温放置 10 分钟。
3. 于 4°C，12000 rpm 离心 10 分钟，并将上清转入新的 EP 管中
4. 向 EP 管中加入 200 μ l 氯仿溶液，充分将其混匀，并于室温下放置约 2 分钟左右。
5. 于 4°C，12000 rpm 离心 10 分钟，将上清转入新的 EP 管中。
6. 向 EP 管中加入 500 μ l 异丙醇，充分混匀，于室温放置 10 分钟。
7. 于 4°C，12000 rpm 离心 10 分钟，弃上清，并向沉淀加入 1 ml 70%乙醇，吹洗。
8. 于 4°C，7500 rpm 离心 5 分钟，弃上清，于室温放置 10 分钟。
9. 待样品完全干燥后，加入 30 μ l ddH₂O，充分溶解后，于-70°C 保存。

3.2.2.2. 逆转录 PCR

1. 反转录 PCR 运用的为 Fermentas 反转录 PCR 试剂盒，冰上配置以下反应体系：

表 3.1 反转录反应体系

名称	用量
水稻 RNA	2 μ g
oligo dT	1 μ l
ddH ₂ O	8 μ l

2. 混合均匀, 65° C 温育 5 min, 使 mRNA 二级结构展开, 完成后立即将 PCR 管放在冰上, 防止二级结构恢复。

3. 冰上依次加入以下试剂:

表 3.2 反应体系

名称	用量
5×逆转录酶反应 buffer	4 μ l
dNTP (10 mM)	1 μ l
RNase H 抑制剂	1 μ l
MLV 逆转录酶	1 μ l

4. 充分混匀后, 于 37° C 进行 1 小时延伸反应, 加入 1 μ l 0.5 M EDTA, 置于 70° C , 10 分钟终止反应, cDNA 产物置于-70° C 保存。

3. OsHDAC1 和 OsBZR1 CDS 片段扩增

使用引物如下:

OsHDAC1 F: GTACCGAGCTCgga^{tcc}ATGGACGCCTCCGCCGGA;

OsHDAC1R: GGCCGTTACTAGTgga^{tcc}TGTCTTCTGGTGCACCGAT;

OsHDAC1 F: GTACCGAGCTCgga^{tcc}ATGACGTCCGGGGCGGGCG;

OsHDAC1 R: GGCCGTTACTAGTgga^{tcc}TTTCGCGCCGACGCC;

所有引物均添加了 BamH I 限制酶酶切位点 GGATCC,上游为克隆载体 pEASY-T1 的同源序列,浓度为 10 μ M。按照如下反应体系进行 50 μ l 体系 PCR 反应:

表 3.3 反应体系

名称	用量
dNTP (10 mM)	1 μ l
cDNA	2 μ l

Primer F	2.5 μ l
Primer R	2.5 μ l
Primer R	1 μ l
5X HF buffer	10 μ l
ddH ₂ O	31 μ l

PCR 反应程序:

表 3.4 PCR 反应体系

温度	用量
95℃	2 min
95℃	10 s
63℃	10 s (35 cycles)
72℃	30 s
72℃	5 min
12℃	10 min

3.2.2.3 凝胶回收 PCR 目的产物

使用 AxyGen 凝胶回收试剂盒进行 PCR 产物的回收实验, 测定回收好的目的片段浓度, -20℃ 保存备用。

3.2.2.4 目的片段和载体的连接

连接反应的方法与操作依据 Vazyme 同源重组试剂盒(C112)进行:

1. 根据载体和插入片段浓度计算所需 DNA 用量;
2. 于冰上配制以下反应体系:

表 3.5 反应体系

组分	重组反应	阴性对照 1	阴性对照 2
线性化载体	X μ l	X μ l	0 μ l
插入片段	Y μ l	0 μ l	Y μ l
5x CE II buffer	4 μ l	0 μ l	0 μ l
Exnase II	2 μ l	0 μ l	0 μ l
ddH ₂ O	to 20 μ l	to 20 μ l	to 20 μ l

X/Y 根据载体使用量情况 0.03 pmol, 插入片段 0.06 pmol 计算使用量;

3. 轻轻吹打混匀, 短暂离心;

4. 37℃温育 0.5 h，完成后立即置于冰上备用

3.2.2.5 连接产物转化大肠杆菌感受态及阳性克隆鉴定

1. 将大肠杆菌感受态 DH5 α 置于冰上融化；
2. 加入 3 μl 连接产物，轻轻混匀，冰上静置 30min；
3. 热激：将以上 EP 管放入 42℃水浴中，90 秒后立即取出；
4. 迅速加入 1 ml LB 培养基，放入 37℃摇床上复苏 1-2 小时，转速 180rpm；
5. 在超净台上使用灭菌的枪头吸取 200 μl 的菌液，均匀涂布在 LB 平板上 (Kan+)，倒置培养 20 h，37℃。
6. 利用金牌 PCR mix（擎科）配置 50 μl 菌落 PCR 反应体系：

表 3.6 PCR 反应体系

名称	用量
Primer F	1 μl
Primer R	1 μl
PCR mix	48 μl

7. 用无菌 10 μl 枪头挑取菌落，在新的 LB 平板 (Kan+) 上接种后，枪头没入 PCR 反应体系，轻轻吹打混匀。
8. 进行菌落 PCR 反应，新接种的平板继续培养。

表 3.7 反应体系

温度	时间
95℃	2 min
95℃	20 s
63℃	20 s (35 cycles)
72℃	1 min
72℃	5 min
12℃	10 min

9. 电泳检测阳性菌落，培养后送测序。

3.2.2.6 35S-pSPYNE-OsHDAC1 和 35S-pSPYCE-OsBZR1 表达载体构建和受体转化

1. 目的基因和 YFP 互补载体的酶切：将上一步连接在 T1 载体上的目的片段和 YFP 互补载体分别用 BamH I 内切酶酶切；
2. 电泳回收目的片段和线性化载体；
3. 连接，T4 连接酶 16℃连接 12 h；

3.2.2.6 质粒转化 GV3101 土壤农杆菌

1. 取出 GV3101 农杆菌，用手捂至融化，加入约 2 μ g 的 DNA 连接产物，轻轻摇晃并混合均匀，在冰上放置 30 min；
2. 液氮低温处理 90s，迅速转入 37℃水浴锅中，1 min 后取出；
3. 加入 1 ml LB 培养基，放在摇床上培养 4h，200rpm，25℃；
4. 将上述菌液涂布在含 Kan 和 rif 抗性的 LB 固体培养基上，28℃倒置培养 48 h 左右；
5. 菌落 PCR 鉴定阳性克隆，挑取阳性菌落接种至 rif+/kan+ LB 液体培养基中。培养至 OD600 \approx 0.5。

3.2.2.7 农杆菌 GV3101 转化烟草叶片

将经过转化的农杆菌注射到烟草叶片中，具体操作如下：

1. 将农杆菌菌液常温 4000rpm 离心 15min，收集农杆菌，重悬至 OD600 \approx 1；
2. 用针头将烟草叶片扎一个小洞；
3. 使用去掉针头的 1ml 注射器，吸取菌液对准小洞，同时手指在叶片另一侧堵住小孔，注射菌液至叶片上下表皮之间；
4. 将烟草植株暗室培养一天左右，转入光照下培养两天；
5. 撕下烟草叶片表皮，在激光共聚焦显微镜下观察荧光。

3.2.3 实验结果

最后置于荧光显微镜下观察，其结果如下：

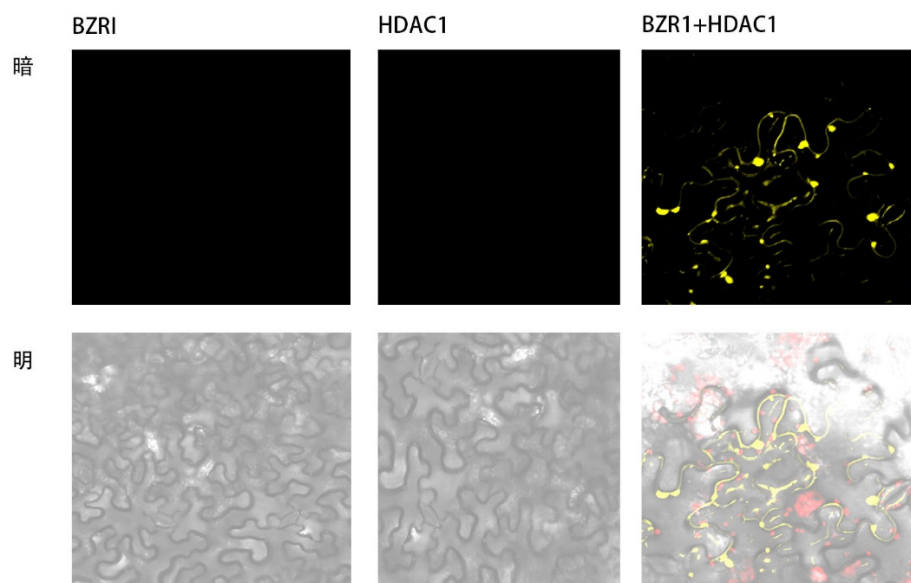


图 3.1 荧光互补实验结果

该图左侧为单蛋白的空白对照，最右为导入了荧光蛋白之后的结果，其中，在中心以及右上都出现了荧光反应，说明 BZR1 与 HDAC1 发生了相互作用，而且荧光信号较强。

结论

本文通过对 HDAC1 蛋白质的蛋白质序列分析，三维构型分析，在不同的生物信息学工具的帮助下，有针对性地筛选出了若干个存在相互作用可能性的蛋白，并且基于序列特点，找出了一个成功发生相互作用的蛋白质 BZR1，最后辅以分子实验验证，其结果可靠程度较高，这样的方法同样可以作为后续打分函数以及分类器的训练集使用[30]。

其中，在 TPR3 蛋白质与 HDAC1 的相互作用以及对 HDAC1 的 String 分析的过程中，发现与 HDAC1 相互作用的蛋白往往都具有 WD40 或类似的重复结构域，这是前人的实验所没有发现的。WD40 结构域的存在，从侧面反映除了 HDAC1 蛋白本身所具有的分子生物学特性，可以用作后续更加深入研究其相互作用方式以及其在代谢通路中的具体角色的实验依据之一。

在分子对接的实验中，仅凭打分函数的分析往往不能说明准确的相互作用位点和方式，但是在 BLAST 实验和基于活性残基的 Autodock 实验的分析之后，从侧面可以看出，三维结构的精确性对相互作用对接的实验结果的影响很大，而且，在后续寻找类似结构的蛋白（如 BZR1）的过程中，仅凭三维结构的自动对接并不能很好地说明相互作用的成功率。故多方位且全面地针对目标蛋白的各个特性（如序列一致性，特殊结构域，相互作用稳定性等）是研究蛋白质之间的相互作用的关键所在[29]。

本文试图通过以 HDAC1 蛋白的相互作用为例，寻找一种批量寻找合适的靶蛋白的方法，鉴于目前已经有好的分类器存在，在药物筛选，抑制剂筛选的过程中，利用分类器是一个十分有效而且快速的方式[30]，除了单个蛋白的对接筛选之外，大规模地分子筛选也将成为以后确定未知蛋白的三维结构，确定小分子靶标的主流。

当然本文也遇到过各种各样的问题，本文所用到的 Autodock 软件目前基于模拟遗传算法，在其早期的版本中有用到传统遗传算法，粒子群算法等等[31]。不同的算法最后所得到的亲和程度（在本文中即结合能的大小）都会不同。所以在条件允许的情况下，应使用多种不同的打分函数和打分方式对结果进行重新测算以求最准确的对接方式。第二则是在进行 Blast 的过程中，BZR1 和 TPR3 的 N

端重叠程度不是很高，而且由于片段较小所以采用了较大的阈值，这样的结果存在一定风险，而目前由于各种目标蛋白之结构的不确定，在最终进行对接时也得到了很多有缺陷的数据。在三维结构不明确的情况下，应当尽量与相关联的同系蛋白或者同组蛋白进行 redocking，以追求最稳定的数据结果。

参考文献

- [1] Lone Dog L. "Whose genes are they? The Human Genome Diversity Project". J Health Soc Policy. (1999) 10 (4): 51–66.
- [2] Snyder M, Du J; Gerstein M (2012). "Personal genome sequencing: current approaches and challenges". Genes Dev. 24 (5): 423–431
- [3] "Economic Impact of the Human Genome Project – Battelle". Retrieved 1 August 2013. DeLisi, Charles (October 2008). "Conferences That Changed the World". Nature. 455 (7215): 876–877.
- [4] Benson D, Boguski M, Lipman D J, et al. The National Center for Biotechnology Information[J]. Genomics, 1990, 6(2):389-391.
- [5] Ardejani, Maziar S.; Powers, Evan T.; Kelly, Jeffery W. "Using Cooperatively Folded Peptides To Measure Interaction Energies and Conformational Propensities". Accounts of Chemical Research. (2017)50 (8): 1875–1882.
- [6] De Las Rivas J, Fontanillo C. "Protein-protein interactions essentials: key concepts to building and analyzing interactome networks". PLoS Computational Biology. (June 2010)6 (6): e1000807.
- [7] Smogorzewska, Agata, Desetty, et al. A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair[J]. Molecular Cell, 2010, 39(1):36.
- [8] Alley S C, Ishmael F T, Jones A D, et al. Mapping Protein–Protein Interactions in the Bacteriophage T4 DNA Polymerase Holoenzyme Using a Novel Trifunctional Photo-cross-linking and Affinity Reagent[J]. American Journal of Hypertension, 2000, 122(25):291-293.

- [9] Tang X, He L, Wang T, et al. Toward the Identification of Novel Carbonic Anhydrase XIV Inhibitors using 3D-QSAR Pharmacophore Model, Virtual Screening and Molecular Docking Study[J]. Letters in Drug Design & Discovery, 2014, 11(4):-.
- [10] Zheng G, Qian Z, Yang Q, et al. The combination approach of SVM and ECOC for powerful identification and classification of transcription factor[J]. Bmc Bioinformatics, 2008, 9(1):1-8.
- [11] Jang I C, Pahk Y M, Song S I, et al. Structure and expression of the rice class-I type histone deacetylase genes OsHDAC1-3: OsHDAC1 overexpression in transgenic plants leads to increased growth rate and altered architecture.[J]. Plant Journal, 2003, 33(3):531-541.
- [12] 王梦禹. 基于机器学习的药物蛋白虚拟筛选方法研究[D]. 哈尔滨理工大学, 2016.
- [13] Morris G M, Huey R, Olson A J. Using AutoDock for ligand-receptor docking.[J]. Curr Protoc Bioinformatics, 2008, Chapter 8(14):Unit 8.14.
- [14] Eiben, A. E. et al. "Genetic algorithms with multi-parent recombination". PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: (1994)78–87.
- [15] 吉根林. 遗传算法研究综述[J]. 计算机应用与软件, 2004, 21(2):69-73.
- [16] Morris G M, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.[J]. Journal of Computational Chemistry, 2009, 30(16):2785-91.
- [17] NG: a database of predicted functional associations between proteins[J]. Nucleic Acids Research, 2003, 31(1):258.
- [18] Mering C V, Huynen M, Jaeggi D, et al. STRI
- [19] Smith TF, Gaitatzes C, Saxena K, Neer EJ. "The WD40 repeat: a common architecture for diverse functions". Trends Biochem. Sci. (May 1999) 24 (5): 181–5.

- [20]Li D, Roberts R (December). "WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases". *Cell. Mol. Life Sci.* 2001.58 (14): 2085–97.
- [21]MODD Mediates Deactivation and Degradation of OsbZIP46 to Negatively Regulate ABA Signaling and Drought Resistance in RiceNing Tang, Siqi Ma, Wei Zong, Ning Yang, Yan Lv, Chun Yan, Zilong Guo, Jie Li, Xu Li, Yong Xiang, Huazhi Song, Jinghua Xiao, Xianghua Li, Lizhong Xiong *The Plant Cell* Sep 2016, 28 (9) 2161-2177; DOI: 10.1105/tpc.16.00171
- [22]Szemenyei, H., Hannon, M. & Long, J. A. TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. *Science* (2008) 319, 1384–1386
- [23]Lobo. Basic Local Alignment Search Tool (BLAST)[J]. *Journal of Molecular Biology*, 2012, 215(3):403-410.
- [24]Lavigne R, Seto D, Mahadevan P, et al. Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools.[J]. *Research in Microbiology*, 2008, 159(5):406-414.
- [25]Uchikoga N, Matsuzaki Y, Ohue M, et al. Re-docking scheme for generating near-native protein complexes by assembling residue interaction fingerprints.[J]. *Plos One*, 2013, 8(7):e69365.
- [26]Seeliger D, Groot B L D. Ligand docking and binding site with PyMOL and Autodock/Vina[J]. *Journal of Computer-Aided Molecular Design*, 2010, 24(5):417-422.
- [27]Bai M Y, Zhang L Y, Gampala S S, et al. Functions of OsBZR1 and 14-3-3 proteins in brassinosteroid signaling in rice.[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(34):13839-13844.
- [28]郝亚琦. BZR1 靶基因的筛选及功能研究[D]. 西北农林科技大学, 2012.
- [29]李洪林, 沈建华, 罗小民,等. 虚拟筛选与新药发现[J]. *生命科学*, 2005, 17(2):125-131.

[30]张健沛, 徐华. 支持向量机(SVM)主动学习方法研究与应用[J]. 计算机应用, 2004, 24(1):1-3.

[31]甄文祥, 王文田. 遗传算法及其应用[J]. 云南民族大学学报:自然科学版, 1997, 11(4):9-10.

致谢

感谢实验室的张昊师兄为本论文提供的建议, 感谢李立家老师对课题的指导以及吴英亮老师对我提出的问题的耐心解答。感谢在 Autodock 论坛上帮助本人实现分子优化以及大规模自动对接的代码修改的人们, 没有你们的无私帮助, 本论文的核心数据便无法顺利得到。

也感谢武汉大学生科院提供的优秀实验环境和网络端口, 为虚拟筛选提供技术支持的计算机学院。

附录

1. Autodock 软件（Windows 版本 x64）源地址:

<http://autodock.scripps.edu/downloads/autodock-registration/tars/dist426/autodocksuite-4.2.6.i86Windows.exe>

使用版本为 Autodock4.2。

2. Autodock 使用参考论坛:

<http://mgl.scripps.edu/forum/viewforum.php?f=12&sid=68a416544d6f4553e246b44c7b2137ae>

<http://mgl.scripps.edu/forum/viewforum.php?f=9&sid=68a416544d6f4553e246b44c7b2137ae>

3. Zdock 配对结果:

<http://zdock.umassmed.edu/results/aa3064f159/>