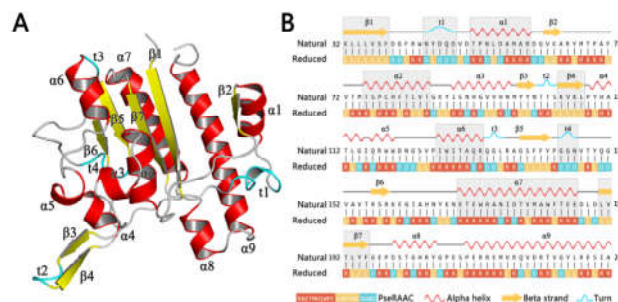


[设为首页](#) [English](#) [蒙文版](#)[全国文明校园风采展示网站](#) [图书馆](#) [校友会](#) [教育发展基金会](#) [内大概况](#)[组织机构](#)[院系设置](#)[教育教学](#)[科学研究](#)[招生与就业](#)[合作交流](#)[校园生活](#)

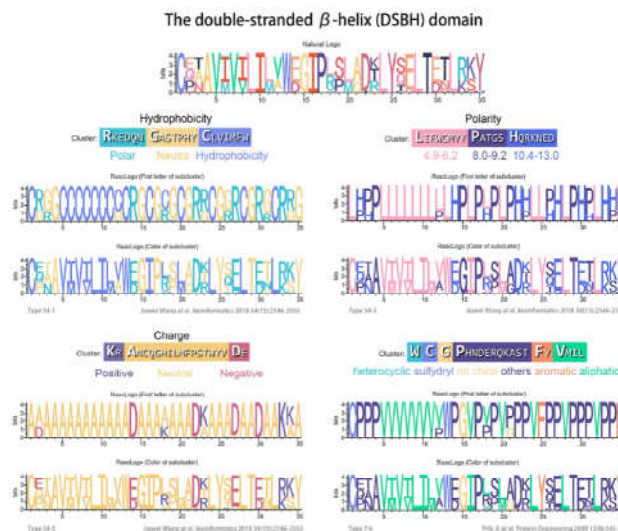
## 左永春教授课题组在蛋白质生物大数据分析平台研究中取得新进展

2020年6月10日,国际生物信息学权威期刊Briefings in Bioinformatics (影响因子IF: 9.101) 在线发表了我校生命科学学院、省部共建草原家畜生殖调控与繁育国家重点实验室左永春教授为通讯作者,博士生郑磊为第一作者的关于约化氨基酸序列logo的最新数据分析平台RaacLogo的科研成果,题目为“RaacLogo: a new sequence logo generator by using reduced amino acid clusters”。本研究得到国家自然科学基金(61561036, 61702290, 61861036)、内蒙古杰出青年培育基金(2017JQ04)和内蒙古青年科技英才(NJYT-18-B01)项目的资助支持。

蛋白质是生命活动主要的承担者,生命的进化最终都体现在蛋白质功能的多样化上。蛋白质是由20种氨基酸编码的,相比于ATGC的DNA遗传信息,氨基酸字母表显示出极大地复杂性和多样性。这就产生了一系列非常有趣的问题:生命为什么会选择20种氨基酸作为标准字母表?更少的氨基酸能否组成或满足一个蛋白质执行功能的基本组成单元?我们能否用更少的氨基酸创造出摆脱当前遗传法则,设计更为独特的生命?理论上,具有相似性质、原子排列或频率分布的天然氨基酸可以被合并,并用更小数量的约化字母表来表示编码的氨基酸。蛋白质氨基酸约化分析是对生命进化本质的探索,同时对于蛋白质合成、拓扑结构预测,以及药物设计等具有重要的意义。



理论上，具有相似性质、原子排列或频率分布的天然氨基酸可以被合并，并用更小数量的约化字母表来表示编码的氨基酸。最简单的约化方案是以极性或者亲疏水性作为标准，将20种氨基酸只约化为两个字母。当然，这种极端的方法产生的蛋白质，一定长度的序列所能承载的信息实在太少，几乎不可能完成复杂的生命活动。那如果是3个，5个，10个呢？另外，氨基酸字母表的大小直接决定的序列motif的复杂性和多样性。而氨基酸序列logo作为蛋白质序列保守度的可视化图形，对研究同源蛋白质的系统发育和功能差异，以及蛋白质定向分子设计具有非常重要的意义。然而传统的logo图形包括了20种氨基酸字母，显示出极大地复杂性和多样性。由于在logo生成过程中缺少对氨基酸性质等生物学因素的考虑，往往会导致对蛋白质功能进化保守性的评估出现偏差。左永春教授课题组基于氨基酸约化的方法，建立了能够简化氨基酸序列复杂性的RaacLogo，使基于信息学的logo图形具有了生物学意义，更好的反映了蛋白序列的功能保守性。RaacLogo作为一个新的简化氨基酸字母的logo生成器，包含了40多种聚类算法和74种约化后的氨基酸字母，这些字母被提取生成673个约化的氨基酸簇(RAACs)，用于简化复杂的logo字母。当20个氨基酸按照一定的相似性聚类到更小的字母表中，使蛋白质的复杂性将大大降低一些功能保守区域将被更清晰地显示出来，从而减少了过拟合的机会，降低了计算障碍，减少了信息冗余。



近年来,左永春教授围绕“基于氨基酸约化的蛋白质结构和功能解析”展开了一系列研究工作和大数据平台开发,2017年,课题组构建了国际上首个氨基酸约化蛋白质序列生物大数据在线分析平台,论文发表在生物信息学核心期刊Bioinformatics选为当年首期目录出版论文(Bioinformatics. 33(1):122-124),该平台引起国际同行们的广泛关注和使用,被国际生物信息权威软件库OMICTOOLS收录(收录编号:OMICS\_08205),当年给予5星推荐指数,连续两年被ESI 1%高被引数据库收录。在此基础上,2019年,课题组构建完成了包含更全氨基酸约化字母表涉及700余种约化类型的数据库RAACBook,并开发基于人工智能机器学习的蛋白质特征提取和功能分类预测平台,成果发表在国际生物数据库权威期刊Database杂志上(2019:baz131,1-12),相关软件获批计算机软件著作权(2019SR0467812)。另外,课题组还解析了DNA去甲基化关键蛋白TET家族功能发挥的序列基础(Briefings in Bioinformatics. 2019, 20 (5), 1826-1835)。

通讯作者简介:左永春,教授,博士生导师,2011年毕业于内蒙古大学,获生物物理学理学博士学位,国际Frontier期刊出版社的Topics Guest Editor, Current Gene Therapy期刊编委,国家自然科学基金评审专家,教育部学位论文通讯评审专家等。以内蒙古大学为第一(通讯)作者单位在专业领域内Briefings in bioinformatics、Bioinformatics、GigaScience等国际期刊上发表学术论文60多篇,成果受到美国艺术与科学院院士、台湾中央研究院院士等同行专家的积极评价,被Nature和PNAS等累计引用1000多次(Google, H=18),多篇成果入选ESI 1%高被引论文,建立了草原家畜生殖调控与繁育生物大数据分析平台。内蒙古自治区优秀博士学位论文和杰出青年培育基金获得者;内蒙古自治区高等学校青年科技英才,内蒙古自治区草原英才,以及新世纪321人才工程等。

论文链接: <https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbaa096/5855392>

上一条: 校党委理论学习中心组召开专题研讨会

下一条: 我校国际学生热议习近平主席给巴基斯坦留学生的重要回信



版权所有 内蒙古大学地址: 内蒙古呼和浩特市大学西路235号|邮编: 010021|邮箱: webmaster@imu.edu.cn |  蒙ICP16002391号-1