

# Progress on Methods of Subcellular Localization of Chloroplast Protein in Plant Cells

Yan Wang, Jie Na\*

School of Life Sciences, Liaoning Normal University, Dalian Liaoning  
Email: [jjienalnnu@126.com](mailto:jjienalnnu@126.com)

Received: Apr. 20<sup>th</sup>, 2020; accepted: May 19<sup>th</sup>, 2020; published: May 26<sup>th</sup>, 2020

---

## Abstract

Chloroplast plays an important role in photosynthesis. Chloroplast proteins distribute in different structures in chloroplast according to their various functions. Proteins could be effective and take part in cell activities only after they are transported to special loci. Therefore, it is necessary to know the subcellular localization of chloroplast protein to understand the relationship between chloroplast structure and function. Recently, in the research of plant protein localization, some experimental methods assisted by bioinformatic prediction were used to increase the accuracy of protein localization analysis. Basic theory and main steps of these two methods were summarized in this paper in order to provide an effective method for chloroplast protein subcellular localization research.

## Keywords

Chloroplast Protein, Subcellular Localization Methods

---

# 植物细胞中叶绿体蛋白质的亚细胞定位方法

王 颜, 那 杰\*

辽宁师范大学生命科学学院, 辽宁 大连  
Email: [jjienalnnu@126.com](mailto:jjienalnnu@126.com)

收稿日期: 2020年4月20日; 录用日期: 2020年5月19日; 发布日期: 2020年5月26日

---

## 摘 要

叶绿体是植物光合作用的重要场所。根据不同的功能, 叶绿体蛋白分布在叶绿体的不同区域。只有当蛋

\*通讯作者。

白质运输到特定的位置时, 才能发挥其功能并参与细胞的生命活动。因此, 确定蛋白质在细胞中的位置对于理解蛋白质的功能和结构是重要的。近年来, 在植物蛋白质定位研究中, 常用生物信息学预测法辅助实验方法, 增加蛋白质定位的研究的准确性。本文对近几年来常用的叶绿体蛋白质定位研究的预测方法以及实验方法的主要进展进行综述, 为高效进行叶绿体蛋白亚细胞定位方法提供参考。

## 关键词

叶绿体蛋白质, 亚细胞定位方法

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

叶绿体是植物细胞特有的细胞器, 是进行光合作用的重要场所。叶绿体是半自主的细胞器, 叶绿体蛋白质约占高等植物蛋白质的 10%~25%, 仅有一小部分叶绿体蛋白由叶绿体基因编码, 大部分是核基因编码[1] [2] [3]。叶绿体蛋白主要分为叶绿体膜蛋白、叶绿体基质蛋白、类囊体膜蛋白和类囊体腔蛋白[4]。蛋白质的功能与其自身的结构和亚细胞定位密切相关。只有当蛋白质运输到特定的亚细胞位置时, 才能发挥其功能并参与细胞生命活动[5] [6] [7]。因此, 了解蛋白质的亚细胞定位有助于了解蛋白质的生物学功能, 解释蛋白质与其他生物分子之间的相互关系[8] [9]。叶绿体蛋白定位方法常见的有实验定位和生物信息学预测两种方法[10] [11]。本文介绍了近几年来常用的叶绿体蛋白质定位研究的预测方法以及实验方法, 为高效进行叶绿体蛋白亚细胞定位方法提供参考。

## 2. 生物信息学预测方法

随着后基因组时代的到来, 生物信息学分析技术的不断发展, 利用计算机系统理论预测蛋白质亚细胞定位的手段也越来越成熟, 涌现大量的预测方法和在线预测网站。计算机预测方法低成本、高效率, 为生物学工作者的实验研究提供了一种高效的辅助手段, 主要涉及以下五个步骤。

### 2.1. 叶绿体蛋白质预测相关数据集的建立

UniProt 数据库集合了 SWISS-Prot、PIR-PSD 和 TrEMBL 三大数据库的序列信息, 是序列信息资源最全面、最丰富的数据库[12]。除了这些综合性的数据库以外, 目前还出现了一些针对亚细胞定位的数据库, 例如, 由 Du 和 Li 针对于亚叶绿体蛋白定位所建立的 SubChlo 原始数据库和针对亚线粒体蛋白定位的 SUBMITO 原始数据库[13] [14]。

蛋白质数据集的建立方法有① 使用已有的数据集; ② 在已有的数据集的基础上增加或减少数据; ③ 创建新的数据集。而目前常用的预测亚叶绿体蛋白定位的几种理论数据集是由 Du 和 Li 创建的 SubChlo 原始数据库, 再经由 CD-HiT 程序处理后, 去除相似性大于 80%、60%、40%的同源性序列, 所得到的 S80、S60、S40 的数据集[15]。

### 2.2. 特征参数的提取

蛋白质特征参数的提取是蛋白质亚细胞定位预测的基础, 根据蛋白质的序列特征能够决定蛋白质在合成过程中分选到特定的位置并发挥正确的生物学功能, 参数提取方法有四:

一是根据氨基酸组分进行分析, 氨基酸是组成蛋白质的基本单位, 多种氨基酸的排列组合构成了多种多样的蛋白质序列, 根据蛋白质氨基酸的组成信息可以有效地预测蛋白质的亚细胞定位, 但这种方法忽略了蛋白质氨基酸的顺序和空间结构等信息;

二是根据氨基酸序列 N 端的信号肽进行分析, 蛋白质可根据蛋白质序列上的特殊的信号序列分选到特定的位置, 并被细胞器上的特异性分选受体识别。1999 年 Nakai 和 Kanehisa [16]首次提出依据氨基酸序列 N 端的分选信号预测亚细胞定位并结合机器学习的方法设计出蛋白质亚细胞定位的预测软件 PSORT。之后 Emanuelsson 等人[17]根据氨基酸序列 N 端的信号肽并结合神经网络的方法, 开发出 TargetP 定位预测软件。蛋白质序列 N 端的信号肽可以提高预测蛋白的亚细胞定位的效率, 但并非所有蛋白质都含有分选信号, 存在一定的局限性;

三是根据氨基酸残基性质, 位于不同的亚细胞区域的蛋白质, 其氨基酸序列含有不同的理化性质。2000 年 Chou 等人[18]首次提出将氨基酸组分信息与物理化学属性信息相结合, 将氨基酸分成亲疏水性、极性、酸碱性和 R 基等化学结构, 以此来降低蛋白质组分的 20 维序列信息, 增加精确度;

四是根据序列同源相似性、蛋白质功能结构域。利用 BLAST 同源相似性搜索工具进行序列对比, 可得到保守序列的同源信息。但该法需要大量蛋白质功能结构域数据的支持, 目前尚无较为完整数据库, 使用受限。

### 2.3. 预测算法设计

分类预测算法能够正确、高效地预测不同蛋白质的亚细胞位置的, 为确定蛋白质亚细胞定位提供了重要信息。目前常用的预测算法主要都是基于机器学习分类法, 其中包括人工神经网络、支持向量机(SVM)、马尔可夫模型、贝叶斯网络组合耦合等。而在机器学习分类法中, 支持向量机(SVM)算法是成功率最高、操作运行最为简便的算法之一[19]。该算法采用了大间隔的线性分离, 计算结果错误率极低, 能够处理多种的机器学习问题上, 因此被广泛地应用于各种生物信息学研究。

### 2.4. 预测结果评估

得到的预测结果的可信度需要用科学的方法进行检测和评估, 常用的方法包括自洽检验(Self-consistency)、交叉检验(Cross validation)和留一交叉检验(Jackknife) [20]。其中留一交叉检验是交叉检验中的一种特殊状况, 其原理就是训练集中的每一个蛋白质轮流作为待测蛋白, 将剩余蛋白质作为训练样本进行测试, 是当前比较客观的检验方法之一[21]。由于在检测蛋白过程当中会出现较高的假阳性, 因此在预测过程中要尽可能的预测出正确的界面残基, 以减少假阳性和假阴性。并分别用敏感性(Sensitivity, Sn)、特异性(Specificity, Sp)、预测成功率(Accuracy, Acc)、马修相关系数(Matthews's Correlation Coefficient, MCC)指标进行度量[22], 预测结果与真实数据的相关性。

### 2.5. 软件和 Web 网站的建立

当预测算法通过预测结果评估被证实具有较高可信度时, 可借助互联网建立一个在线预测网站, 以方便更多的研究人员在线获得待测蛋白的预测结果。目前已有大量的在线预测网站应用于叶绿体蛋白质亚细胞定位研究中(见表 1)。

## 3. 实验定位的方法

在蛋白质亚细胞定位研究中生物信息学预测结果具有一定的参考价值, 但并不能完全确定蛋白质的亚细胞定位, 最后还需要通过实验确定蛋白质的具体定位。实验定位法包含免疫胶体金标记、免疫荧光技术和融合报告基因定位法。

**Table 1.** Main prediction websites for subcellular localization of chloroplast proteins  
**表 1.** 主要叶绿体蛋白质亚细胞定位预测网站

名称	网址	预测依据
Ipsort	<a href="http://ipsort.hgc.jp/">http://ipsort.hgc.jp/</a>	根据 N 末端分选信号
TargetP	<a href="http://www.cbs.dtu.dk/services/TargetP/">http://www.cbs.dtu.dk/services/TargetP/</a>	根据叶绿体信号肽
ChloroP	<a href="http://www.cbs.dtu.dk/services/ChloroP/">http://www.cbs.dtu.dk/services/ChloroP/</a>	根据叶绿体信号肽
WoLF PSORT	<a href="https://wolfsort.hgc.jp/">https://wolfsort.hgc.jp/</a>	根据氨基酸组成
SubLoc	<a href="https://bigd.big.ac.cn/databasecommons/database/id/5768">https://bigd.big.ac.cn/databasecommons/database/id/5768</a>	根据氨基酸组成
Cell-PLoc	<a href="http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc/">http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc/</a>	根据氨基酸组成
Euk-mPLoc 2.0	<a href="http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/">http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/</a>	根据氨基酸组成
CELLO	<a href="http://cello.life.nctu.edu.tw/">http://cello.life.nctu.edu.tw/</a>	根据氨基酸组成
PCLR	<a href="http://www.andrewschein.com/cgi-bin/pclr/pclr.cgi">http://www.andrewschein.com/cgi-bin/pclr/pclr.cgi</a>	根据叶绿体信号肽
SLP-Local	<a href="https://sunflower.kuicr.kyoto-u.ac.jp/~smatsuda/spllocal.html">https://sunflower.kuicr.kyoto-u.ac.jp/~smatsuda/spllocal.html</a>	根据氨基酸组成
Plant-mPLoc	<a href="http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/">http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/</a>	根据蛋白质功能域和氨基酸组成

### 3.1. 融合报告基因定位法

对 2016 年以来植物叶绿体蛋白的亚细胞定位实验方法进行搜集整理发现, 近年来, 以绿色荧光蛋白基因作为报告基因的融合报告基因定位法被广泛使用, 利用瞬时表达技术使融合基因在植物细胞中瞬时表达, 利用激光共聚焦显微镜和超分辨率显微镜观察进行活体跟踪检测, 确定细胞中蛋白质分子的定位以及代谢途径[23]。

绿色荧光蛋白(GFP)可以自发荧光, 易于检测、性质稳定并且对宿主细胞无毒害作用, 因此可以作为报告蛋白在植物、动物、微生物中广泛使用。由于叶绿体自带红色荧光, 因此常用绿色荧光蛋白基因以及其突变体黄色荧光蛋白(YFP)基因作为融合报告基因进行亚细胞定位实验, 也可以使用 EGFP、EYFP 等增强型绿色荧光蛋白(荧光强度比 GFP 高 6 倍以上更强)的 GFP 的突变体, 改善检测效果[24]。

瞬时表达的宿主细胞主要有拟南芥原生质体、烟草叶片细胞以及洋葱表皮细胞。但通过瞬时表达技术进入细胞的外源 DNA 并不与宿主细胞染色体 DNA 进行整合, 而是直接随载体进入细胞, 迅速表达, 在 2~4 d 内就能检测到表达产物[25]。瞬时表达技术能够表达多种外源基因并且不会受基因位置以及基因沉默等方面的影响, 具有简单、快捷、准确、效率高等优点, 是一种快速研究基因表达、蛋白质亚细胞定位以及基因互作的重要手段[26]。因此利用瞬时表达技术, 使目的基因与含有绿色荧光蛋白基因的表达载体连接, 成功表达出的融合产物就能散发出绿色荧光, 最后利用特殊的显微镜对融合产物进行跟踪观察检测。

### 3.2. 利用激光共聚焦显微镜观察

蛋白质的功能虽然与蛋白质的定位相关, 但蛋白质功能的发挥是一个动态的过程, 涉及到蛋白质之间的互动和修饰, 需要保证实验材料必须是活细胞。20 世纪 80 年代激光共聚焦显微镜的出现, 实现了细胞的活体 3D 成像[27] [28]。激光共聚焦显微镜是在荧光显微镜的基础之上配置激光器和扫描装置, 利用计算机技术对观察样品进行处理, 以获得清晰图像。激光共聚焦显微镜结合了显微镜技术、光电技术和计算机技术, 解决了传统显微镜观察方法在观察之前需要对样品进行脱水、脱蜡、染色等处理, 会对细胞产生损伤的问题, 激光共聚焦显微镜可以在不损伤细胞的前提下, 对活体细胞进行观察和跟踪, 观

察过的样品还可以用于后续实验[28]。传统的光学显微镜使用的是场光源, 样品需要在场光源下一次性成像, 样品成像会受到附近的物体的衍射光和散射光的影响, 会形成色差[29]。激光共聚焦显微镜的光源为激光, 加上采用了共聚焦技术, 减小了色差, 样品在成像过程中被分解成二维甚至三维的多个点, 逐点扫描成像, 最终导入计算机中形成一个完整的图像, 彻底消除了色差[28], 并且表现更出色的成像清晰度和精确度。

虽然激光共聚焦显微镜解决了活细胞的成像问题, 但受分辨率的限制, 还不能精准的定位到植物亚细胞器内的具体位置。

### 3.3. 利用超分辨率显微镜观察

融合报告基因定位法广泛应用促进了一系列以提高分辨率为目的的显微镜成像技术的发展[30]。继激光共聚焦显微镜之后, 超分辨率显微镜也因其强大的分辨率, 精确的定位, 突破了显微镜成像技术的一大难关。

超分辨率显微镜能对亚细胞器中的蛋白质进行精确定位, 可在植物和动物细胞中广泛使用。2011年, Zheng 等人[31]利用超分辨率显微镜将海马脊柱的突出细胞骨架蛋白精确定位于突触和神经元胞质中。2015年, 刘玥等人[32]利用 3D-SIM 超分辨率显微镜确定植物蔗糖合成酶 ZmSUS-SH1 定位于烟草表皮细胞叶绿体的基粒周围。激光共聚焦显微镜能将融合蛋白精确于具体的细胞器, 而超分辨率显微镜能将融合蛋白精确于细胞器中的具体位置。

超分辨率成像技术提高了显微镜的分辨率, 成功克服了融合蛋白荧光不易区分以及叶绿素自发荧光等缺点, 实现了叶绿体蛋白质在细胞器内的精确定位[32]。同时利用计算机软件可以对样品进行 3D 重构, 并对样品的 3D 模型进行任意角度的旋转和切割, 甚至还可以导出 3D 模型, 实现蛋白质在亚细胞器内的可视化精确定位。

## 4. 结论

综上, 叶绿体蛋白质亚细胞定位的两种方法各有千秋, 实验定位方法测量结果精准, 但耗时长、成本高, 仪器设备昂贵, 而且实验结果需要人为的经验判断, 具有一定的主观性。生物信息学预测方法具有利用大数据提供预先参考信息的优势, 因而对传统的实验方法发挥良好的辅助作用, 也节省了时间成本。

叶绿体蛋白质的功能与蛋白质的亚细胞定位联系紧密, 具体功能的发挥需要由蛋白质之间相互作用、修饰等动态变化来体现。多种的蛋白质亚细胞定位实验方法的出现、发展以及使用, 为叶绿体蛋白质数据库积累了大量的数据。计算机软件技术的崛起, 生物信息学技术的发展, 使互联网在线预测蛋白质亚细胞定位成为可能, 也为高效率分析叶绿体蛋白亚细胞行为提供了有益参考。

## 参考文献

- [1] Van Wijk, K.J. (2000) Proteomics of the Chloroplast: Experimentation and Prediction. *Trends in Plant Science*, **5**, 420-425. [https://doi.org/10.1016/S1360-1385\(00\)01737-4](https://doi.org/10.1016/S1360-1385(00)01737-4)
- [2] 齐欣, 崔继哲, 朱宏. 叶绿体的蛋白质组学研究进展[J]. 现代生物医学进展, 2007, 7(3): 440-442.
- [3] Ouyang, M., Li, X., Ma, J., et al. (2011) LTD Is a Protein Required for Sorting Light-Harvesting Chlorophyll-Binding Proteins to the Chloroplast SRP Pathway. *Nature Communications*, **2**, 277. <https://doi.org/10.1038/ncomms1278>
- [4] 王金辉, 李铁女, 刘惠芬, 等. 叶绿体蛋白质组学研究进展[J]. 生物技术通报, 2011(2): 1-6.
- [5] Nair, R. and Rost, B. (2008) Protein Subcellular Localization Prediction Using Artificial Intelligence Technology. *Methods in Molecular Biology*, **484**, 435-463. [https://doi.org/10.1007/978-1-59745-398-1\\_27](https://doi.org/10.1007/978-1-59745-398-1_27)
- [6] Fujiwara, Y. and Asogawa, M. (2001) Prediction of Subcellular Localizations Using Amino Acid Composition and Order. *Genome Informatics*, **12**, 103-112.

- [7] 张松, 黄波, 夏学峰, 等. 蛋白质亚细胞定位的生物信息学研究[J]. 生物化学与生物物理进展, 2007, 34(6): 573-579.
- [8] Du, P., Li, T. and Wang, X. (2011) Recent Progress in Predicting Protein Sub-Subcellular Locations. *Expert Review of Proteomics*, **8**, 391-404. <https://doi.org/10.1586/epr.11.20>
- [9] Chou K C, Shen H B. (2007) Recent Progress in Protein Subcellular Location Prediction. *Analytical Biochemistry*, **370**, 1-16. <https://doi.org/10.1016/j.ab.2007.07.006>
- [10] 项新媛. 基于融合信息的蛋白质亚叶绿体定位预测研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2014.
- [11] 邢浩然, 刘丽娟, 刘国振. 植物蛋白质的亚细胞定位研究进展[J]. 华北农学报, 2006, 21(s2): 1-6.
- [12] Garavelli, J.S. (2010) The RESID Database of Protein Modifications as a Resource and Annotation Tool. *Proteomics*, **4**, 1527-1533. <https://doi.org/10.1002/pmic.200300777>
- [13] Du, P., Cao, S. and Li, Y. (2009) SubChlo: Predicting Protein Subchloroplast Locations with Pseudo-Amino Acid Composition and the Evidence-Theoretic K-Nearest Neighbor (ET-KNN) Algorithm. *Journal of Theoretical Biology*, **261**, 330-335. <https://doi.org/10.1016/j.jtbi.2009.08.004>
- [14] Du, P. and Li, Y. (2006) Prediction of Protein Submitochondria Locations by Hybridizing Pseudo-Amino Acid Composition with Various Physicochemical Features of Segmented Sequence. *BMC Bioinformatics*, **7**, 518. <https://doi.org/10.1186/1471-2105-7-518>
- [15] Huang, C. and Yuan, J.Q. (2013) Predicting Protein Subchloroplast Locations with Both Single and Multiple Sites via Three Different Modes of Chou's Pseudo Amino Acid Compositions. *Journal of Theoretical Biology*, **335**, 205-212. <https://doi.org/10.1016/j.jtbi.2013.06.034>
- [16] Nakai, K. and Horton, P. (1999) PSORT: A Program for Detecting Sorting Signals in Proteins and Predicting Their Subcellular Localization. *Trends in Biochemical Sciences*, **24**, 34-36. [https://doi.org/10.1016/S0968-0004\(98\)01336-X](https://doi.org/10.1016/S0968-0004(98)01336-X)
- [17] Emanuelsson, O., Nielsen, H., Brunak, S., et al. (2000) Predicting Subcellular Localization of Proteins Based on Their N-Terminal Amino Acid Sequence. *Journal of Molecular Biology*, **300**, 1005-1016. <https://doi.org/10.1006/jmbi.2000.3903>
- [18] Chou, K.C. (2000) Prediction of Protein Subcellular Locations by Incorporating Quasi-Sequence-Order Effect. *Biochemical and Biophysical Research Communications*, **278**, 477-483. <https://doi.org/10.1006/bbrc.2000.3815>
- [19] 王瑞雪, 李前忠, 闫振河, 等. 基于多信息融合的亚叶绿体定位预测研究[J]. 内蒙古大学学报(自然版), 2017(1): 80-90.
- [20] 项新媛, 李前忠. 蛋白质的进化与结构信息对亚叶绿体蛋白定位的预测[J]. 内蒙古大学学报(自然版), 2013(6): 589-596.
- [21] Shi, W., Punta, M., Bohon, J., et al. (2011) Characterization of Metalloproteins by High-Throughput X-Ray Absorption Spectroscopy. *Genome Research*, **21**, 898. <https://doi.org/10.1101/gr.115097.110>
- [22] 沈金城, 张松, 孙之荣. 一种基于 GO 分子功能信息的序列编码辅助预测亚细胞定位方法[J]. 科学通报, 2007, 52(13): 1536-1541. <https://doi.org/10.1360/CSB2007-52-13-1536>
- [23] Smith, M.C.M. (1991) Molecular Biological Methods for Bacillus. *FEBS Letters*, **287**, 227. [https://doi.org/10.1016/0014-5793\(91\)80059-C](https://doi.org/10.1016/0014-5793(91)80059-C)
- [24] 孙琳琳, 佟少明, 侯和胜. 拟南芥叶肉原生质体分离条件的优化[J]. 植物生理学报, 2009, 45(3): 282-284.
- [25] 于一帆, 朱小彬, 葛会敏, 等. 基于绿色荧光蛋白瞬时表达的植物亚细胞定位方法[J]. 江苏农业科学, 2014, 42(12): 58-61.
- [26] Stülke, J. and Hillen, W. (2000) Regulation of Carbon Catabolism in Bacillus Species. *Annual Review of Microbiology*, **54**, 849-880. <https://doi.org/10.1146/annurev.micro.54.1.849>
- [27] 李楠, 王黎明. 激光共聚焦显微镜的原理和应用[J]. 军医进修学院学报, 1996(3): 232-234.
- [28] 赵启韬, 苗俊英. 激光共聚焦显微镜在生物医学研究中的应用[J]. 北京生物医学工程, 2003, 22(1): 52-54. <https://doi.org/10.1111/j.1741-6612.2003.tb00465.x>
- [29] 霍霞, 吕建勋, 杨仁东, 等. 激光共聚焦显微镜与光学显微镜之比较[J]. 激光生物学报, 2001, 10(1): 76-79.
- [30] 刘玥, 尹悦佳, 梁重阳, 等. 3D-SIM 结构照明超分辨率显微镜实现蛋白质在植物亚细胞器内的定位[J]. 植物学报, 2015, 50(4): 495-503.
- [31] Zheng, C.Y., Wang, Y.X., Kachar, B., et al. (2011) Differential Localization of SAP102 and PSD-95 Is Revealed in Hippocampal Spines Using Super-Resolution Light Microscopy. *Communicative & Integrative Biology*, **4**, 104-105. <https://doi.org/10.4161/cib.14172>
- [32] Tanos, B.E., Yang, H.J., Soni, R., et al. (2013) Centriole Distal Appendages Promote Membrane Docking, Leading to Cilia Initiation. *Genes & Development*, **27**, 163-168. <https://doi.org/10.1101/gad.207043.112>