

蛋白质复合物表达纯化技术的研究进展

李思葳^{1,2} 李佳² 臧奕^{1,3*}

(¹国科大杭州高等研究院药物科学与技术学院, 杭州 310012; ²中国科学院上海药物研究所, 上海 201203;

³临港实验室, 上海 200031)

摘要 蛋白质复合物作为细胞的功能分子, 在生物体信号传递及功能调控中发挥重要作用, 复合物结构的解析对于研究蛋白间相互作用的分子机制、预测蛋白功能、预测药物靶标以及药物设计都具有重要意义。目前蛋白质复合物的纯化方式主要有三种: 纯化获得复合物单一组分后在体外进行组装形成稳定复合体; 将复合物全部组分表达至同一宿主中, 使复合物在体内发生组装; 纯化内源蛋白质复合物。该文归纳综述了当前蛋白质复合物纯化的三种方式, 以期为蛋白复合物纯化技术的发展及其他重要蛋白复合物结构的研究提供参考。

关键词 蛋白表达纯化; 蛋白-蛋白相互作用; 蛋白质复合物

Research Progress in Protein Complex Expression and Purification Technology

LI Siwei^{1,2}, LI Jia², ZANG Yi^{1,3*}

(¹School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, Chinese Academy of Sciences, Hangzhou 310012, China; ²Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China;

³Lingang Laboratory, Shanghai 200031, China)

Abstract Protein complex, the functional molecule of cell, appears to play a central role in signal transduction and functional regulation of organism. The structure of protein complex is of great significance for studying the molecular mechanism of protein-protein interaction, predicting protein function, discovering potential drug targets and contributing to drug design. Three methods are mainly used for purifying protein complexes so far. One major technique is to obtain complex components separately and assemble *in vitro*; another is co-expression of protein components in the same host cell following by assembling *in vivo*; the third one is purification of endogenous protein complexes, notably through the use of knock-in method. Here, this study summarizes three major techniques for isolating protein complex and anticipates to provide reference for further researches on important protein complex structure.

Keywords protein expression and purification; protein-protein interaction; protein complex

蛋白质是生命的物质基础, 是细胞的功能分子。蛋白与蛋白间发生相互作用产生一系列级联反应, 实现对蛋白自身功能的调控以及下游信号的传递与反馈, 而这种相互作用的实现需要不同蛋白间发生位点特异性识别, 形成短暂或长期存在的功能复合体^[1]。由于蛋白质结构是功能的基础, 因此确定

蛋白质复合物的结构对于研究蛋白间相互作用的分子机制、蛋白功能预测、蛋白设计以及临床药物发现至关重要^[2-4]。其中, 高效的纯化技术和手段是复合物结构解析的基础和关键。

蛋白纯化的目的是将具有生物学活性和天然结构的蛋白从全细胞组分中分离出来, 进而深

收稿日期: 2023-06-26 接受日期: 2023-09-19

国家自然科学基金(重大研究计划项目)(批准号: 92253306)资助的课题

*通讯作者。Tel: 021-50800753-124, E-mail: yzang@simmm.ac.cn

Received: June 26, 2023

Accepted: September 19, 2023

This work was supported by the National Natural Science Foundation of China (Major Research Program) (Grant No.92253306)

*Corresponding author. Tel: +86-21-50800753-124, E-mail: yzang@simmm.ac.cn

入研究蛋白间相互作用活性位点、作用机制、结构与功能之间的联系。蛋白纯化的主要方式是体外纯化,即把目的蛋白在大肠杆菌、酵母细胞、昆虫细胞及哺乳动物细胞等非自体系统中异源过表达^[5],操作简单且能获得大量蛋白,但对于与宿主细胞组织相容性较差的部分蛋白^[6],内源纯化更为合适。相较于蛋白单体,蛋白复合物的纯化更为复杂,需要多个蛋白进行组装并形成稳定的复合物。目前科学家已经成功解析出多种蛋白质复合物的结构,如催化组蛋白H3K4甲基化的人源混合谱系白血病(mixed-lineage leukaemia, MLL)家族的甲基转移酶蛋白复合体^[7]、丙型肝炎病毒(hepatitis C virus, HCV)的结构蛋白E1E2复合体^[8]、T细胞表面受体TCR/pMHC复合物^[9]以及染色质重塑复合物PBAF(polybromo-associated BRG1-associated factor, PBAF)复合物^[10]等,这些蛋白复合物结构的解析使我们能够在原子水平上观察到蛋白间相互作用的新机制,为临床疾病的研究以及后续药物的研发提供思路。蛋白复合物的纯化技术大致可以分为以下三种:(1) 分别纯化蛋白复合物的组分,在体外适宜条件下进行组装,孵育形成稳定的蛋白复合物;(2) 将蛋白复合物的全部组分在细胞内共表达,蛋白的组装发生在细胞内,形成稳定复合物后被体外纯化;(3) 纯化内源蛋白复合物,如利用CRISPR/Cas基因敲入(knock-in)技术^[11-13],对细胞或动物基因敲入纯化标签用以蛋白纯化,或直接从全细胞组分中分离目的蛋白复合体。本文在此对蛋白质复合物的表达纯化方式加以综述,以期后续其他重要蛋白复合物结构的研究提供参考。

1 体外组装,纯化蛋白复合物

蛋白质复合物的体外重构是研究分子间相互作用、探究蛋白功能的重要手段^[14]。蛋白单体体外纯化的实现使得复合物的纯化成为可能,最为简单直接的方式便是在蛋白复合物各组分已知的前提下,分别异源过表达组分蛋白,获取到正确折叠的活性单体蛋白后,在体外适宜缓冲液及一定强度刺激信号下实现功能复合物的组装^[15-17]。组装成稳定复合体后,清除体系内其他多余单体以提高目的样本纯度,在此过程中各组分蛋白的比例、组装条件与纯化方式的选择都至关重要。

当选用体外组装的方式纯化蛋白复合物时,首先要对蛋白复合物各组分有充分了解,如各组分的

组成,其蛋白表达偏好性,适用的宿主细胞种类等,且蛋白的组装发生在体外,因而要了解复合物各蛋白结合比例与组装条件。2018年HUGHES等^[18]研究者将全长瞬时受体电位离子通道5(transient receptor potential vanilloid 5, TRPV5)在酵母细胞中表达,在大肠杆菌中扩增表达钙调蛋白(calmodulin, CaM),分别纯化后将两种蛋白按照1:20的摩尔比孵育,并添加10 mmol/L CaCl₂促进TRPV5-CaM复合物组装,最终获得高纯度目的蛋白复合物。IVIC等^[19]研究者分别纯化蛋白复合物组分核输入蛋白7(importin 7, Imp7)、核输入蛋白 β (importin β , Imp β)和组蛋白1.0(histone 1.0, H1.0),并将这三种蛋白按摩尔比为1:1:1.2的比例混合,并置于冰上孵育,随后加入终浓度为0.1%的戊二醛使复合物交联辅助稳定蛋白复合物。2016年GONG等^[20]将N-端带有MBP标签的固醇调节元件结合蛋白1的C-端调控域(C-terminal regulatory domain of Sre1, Sre1-CTD)质粒表达在大肠杆菌中,将N-端带有10 $\times$ His标签的固醇调节元件结合蛋白裂解激活蛋白C-端WD40结构域(C-terminal WD40 domain of Scp1, Scp1-WD40)的质粒表达在昆虫细胞中, Sre1-CTD和Scp1-WD40按照1:1.2的摩尔比在体外进行组装,形成稳定复合物。

蛋白复合物的体外重建是分子生物学研究中的重要部分^[21],分别纯化蛋白复合物组分,在体外适宜条件下组装形成稳定配合物这一纯化方式使诸多蛋白间相互作用的分子机制被发现^[22-23]。这一简单高效的纯化手段推动了蛋白纯化领域的发展,但此种方法应用的前提是对复合物各组分已有充足的研究,除此以外也有诸多限制,如蛋白复合物的单一组分独自表达在异源环境中时,可能会因为脱离稳定的功能基团而发生自身蛋白稳定性的改变,产生蛋白高聚或自身异常折叠的现象,以及复合物组分纯化后难以组装形成稳定功能基团的现象^[24]。同时,对于复合物组分复杂的大分子来讲,纯化步骤过于繁杂,难以使全部蛋白组分的纯化进程保持同步,实验周期较长且费时费力。

2 体内组装,纯化蛋白复合物

纯化蛋白质复合物的另一种主要方式是共表达策略,即将蛋白复合物的各组分同时表达在同一宿主细胞中,此时蛋白各组分在宿主内组装形成稳定复合物,而后通过亲和层析及凝胶过滤色谱法进

一步分离纯化目的蛋白复合物^[25-26]。共表达策略结合了体内与体外蛋白纯化的优势,既保留了体外纯化的简便高效,又能够利用宿主细胞的内源环境为目的蛋白复合物的组装提供场所和必需物质,极大提高了复合物组装的效率和稳定性^[27]。

目前,共表达纯化的方式主要有两种,第一种是将蛋白复合物各组分分别克隆至不同的表达载体上,将这些带有目的基因的载体质粒同时转染至同一宿主细胞内,各组分蛋白在细胞内转录翻译并在细胞内发生相互作用,形成目的蛋白复合物^[28-29],这一方法是蛋白共表达纯化的常用策略。当多种表达载体在同一宿主内共同表达时,不同载体间的启动子互不干扰,载体携带的目的基因可以相对独立地正确完成转录和翻译过程。但这种方法对选取的载体也有一定要求,第一是共表达所用载体的抗生素标记要有所不同^[30],当不同抗性的质粒同时转染时,多种抗性的存在可以提高转染的阳性率。2021年WANG等^[31]将复合物组分Epo1的C-端部分残基和Bem3的N-端部分残基的cDNA序列分别构建至pET28b和pET21b两种带有不同抗性的载体上,同时转染至大肠杆菌中,进而获得Bem3-Epo1蛋白复合物。第二是选用载体的拷贝数需相似,尽量减少因载体拷贝数而引起的各蛋白组分表达量间的差异^[29];2019年,DONG等^[32]将TCR组分克隆至高表达载体pCAG中,同时将C-端带有SBP和Flag标签的CD3亚基全长序列也构建至高表达载体pcDNA3.4中,以人胚胎肾细胞HEK 293F为异源表达宿主,将含有CD3和TCR复合物的两种质粒同时转染到细胞中使蛋白共表达。第三是当选用杆状病毒感染昆虫细胞进行蛋白表达时,则无需考虑质粒载体的抗性与拷贝数^[33]。ZHUANG等^[34]将目的蛋白复合物组分D1多巴胺受体(D1-dopamine receptors, D1R)、刺激性G蛋白DN_Gαs、Gβ1和Gγ2的杆状病毒按照1:1:1:1的比例同时在SF9昆虫细胞中共表达,在5 μmol/L SKF配体和丙酮酶存在的条件下,D1R-Gs复合物组分在细胞膜上组装并形成稳定的功能复合物。

第二种常用方法是将目的蛋白所有组分的cDNA构建至同一个表达载体中,随后将带有多种蛋白基因序列的质粒转染至宿主细胞内,各组分蛋白在细胞内转录翻译,并实现复合物的组装,形成稳定的目的蛋白复合物^[30,35-36]。此种方法所用的表达载体分为两种,即多转录单元多基因表达载体与单

转录单元多基因表达载体。前者具有多个独立的转录单元,每个转录单元有自己的启动子,蛋白间的表达相对独立,不受其他蛋白的限制,常用的多转录单元多基因表达载体有pETDuet-1、pFBDM、pFL、pKDM、pUCDM等^[37-38]。2023年,ZAHM等^[39]选取人核分裂周期蛋白80(nuclear division cycle 80, Ndc80)复合物的两端氨基酸残基进行密码子优化,并将其导入具有多个启动子的pETDuet-1载体中,同时在两端引入半胱氨酸残基促进二聚体形成,最终得到稳定的Ndc80c loop异源二聚体。但多转录单元多基因表达载体也有部分缺点,如各个转录单元内蛋白的表达量不均一,蛋白表达数量有限,转录单元间的启动子可能会彼此间产生干扰,影响蛋白的正确转录与翻译等。

单转录单元多基因表达载体是能够容纳多种蛋白的cDNA序列,但启动子只有一个的载体。载体中所有蛋白的转录翻译由同一启动子调控,最终形成融合蛋白,常用的单转录单元多基因表达载体有PSTCm9、pETM-14、pRS42X-LNK等^[40]。2020年,WU等^[41]解析了羟甲基戊二酰辅酶A还原酶降解蛋白1(hydroxymethylglutaryl-goA reductase degradation protein 1, Hrd1)泛素连接酶复合物介导蛋白降解的复合物结构。研究者为了获得稳定的Hrd1~Usa1-Der1-Hrd3复合物,在能够表达多种蛋白的质粒载体Prs42X中插入全长Der1,融合Hrd1-Usa1及Hrd3基因序列,这四种不同蛋白的转录翻译由GAL1启动子统一操纵,质粒转染入酵母细胞INVScl后开始表达,并在细胞内组装形成稳定复合物。同年,GRABARCZYK^[25]将Tof1和Csm3两种蛋白的基因序列克隆至带有多个克隆位点的pETM-14载体中,形成pETM14-Tof1-Csm3或其他截短质粒,转染入大肠杆菌后转录翻译形成目的蛋白复合物(表1)。

上面两段简述了蛋白共表达纯化的两种主要方式,在实际应用中这两种方式并非完全独立,而是相辅相成的,许多研究者将这两种方法结合起来^[42],用以提高蛋白复合物的表达量及稳定性。与蛋白的体外组装相比较,多种蛋白质粒的构建或共转染的操作更易实现。体内组装能够提供复合物组分相对自然的环境,原核或真核细胞内的固有成分和营养物质也能够为复合物的组装提供能量和介质。载体种类与宿主细胞的多种选择也可满足不同蛋白的表达偏好性。但也有一些困难存在,如同时过表达多

表1 蛋白复合物组分体内共表达载体总结表

Table 1 Summary of *in vivo* co-expression vectors of protein complex components

载体种类 Type of vectors	载体名称 Vectors	优点 Advantages	限制 Limitations	成功案例 Successful cases
Vectors for co-expression	pET28b, pET21b, pFastBac etc	Promoters function exclusively; the target genes can be correctly and independently translated	Different antibiotic resistance; similar vector copy number	[29-30] [32-34]
Vectors with multiple transcription units	pETDuet-1, pFBDM, pFL, pKDM, pUCDM etc	Multiple transcription units with distinct promoters; independent expression of proteins	Uneven protein expression within different transcription units; limited diversity of proteins expression; interference among various promoters	[38-39]
Vectors with single transcription unit	PSTCm9, pETM-14, pRS42X-LNK etc	Same promoters; harmonized protein transcription-translation levels	The order of protein expression may affect the expression level of proteins	[41]

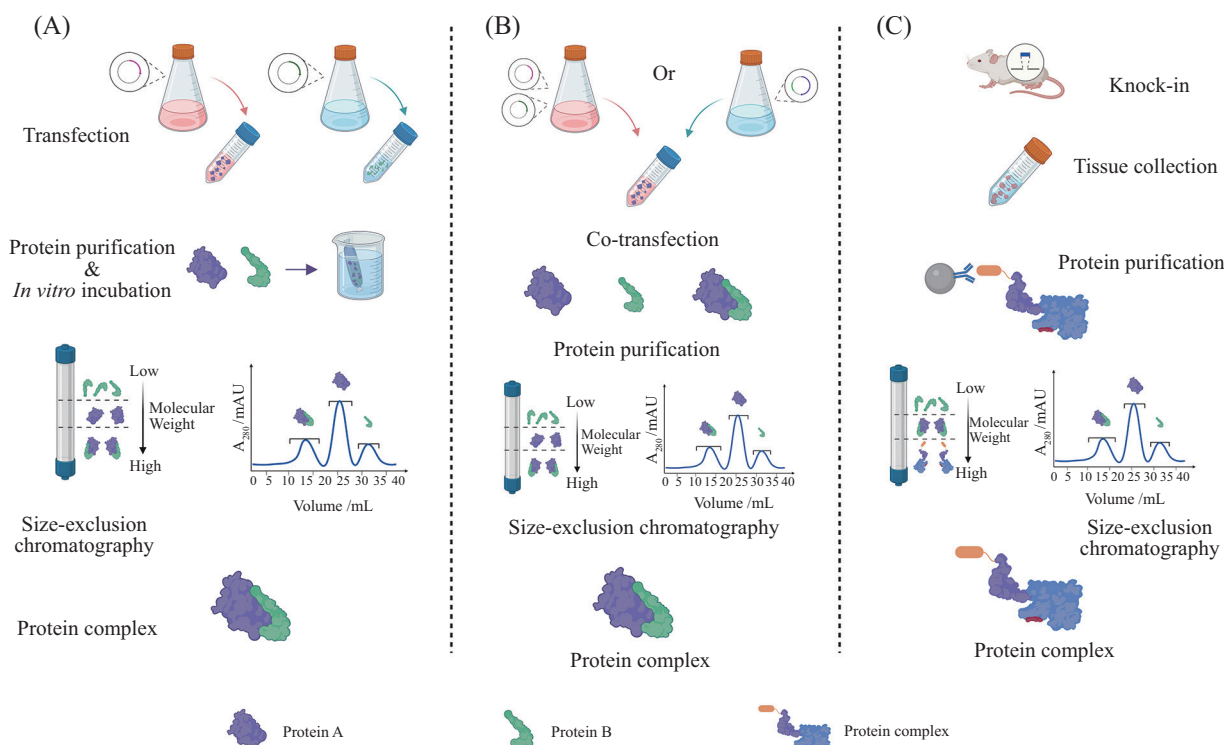
种复合物组分蛋白时, 每种蛋白表达量的差异未知; 某一组分表达量较低, 导致复合物无法形成或形成的复合物并不完整, 无法获得目的蛋白^[43]; 蛋白复合物形成过程中需要特定的条件而条件未知; 纯化的蛋白中可能带有杂蛋白等。

3 纯化内源蛋白复合物

以上两种纯化方式均属于重组蛋白的异源过表达, 适用于蛋白单体或复合物组分稳定且明确的蛋白。而当某些蛋白复合物稳定性差, 或蛋白复合物组分未知时, 难以通过异源过表达的方式获得稳定的天然构象蛋白质, 而且在体外异源过表达某种或某些蛋白时, 也可能会产生非天然的蛋白间相互作用^[44], 影响实验结果的准确性。此时可以通过内源纯化的方式, 从细胞中获取天然蛋白, 如对于质量和体积较大的蛋白复合物如核孔复合物(nuclear pore complex, NPC), 可以直接从细胞中进行分离纯化。如施一公团队^[45]从蛙卵巢中分选第VI期卵母细胞, 在卵母细胞细胞膜一侧打孔将其细胞核吸入到适宜缓冲液中, 清洗核内容物后成功获得了非洲爪蟾的核孔复合物, 发现了构成细胞质环(cytoplasmic ring, CR)的关键组分, 揭示了NPC分子结构的精细设计和精密组装^[46]。除此以外, 随着技术的发展, 研究者们也可以利用CRISPR/Cas9-knock-in技术向动物或细胞中内源的目的蛋白基因上添加用于蛋白纯化的标签, 在不影响蛋白生理功能的情况下, 通过传统的蛋白纯化手段获得具有天然构象的目的蛋白以及与目的蛋白稳定结合的功能基团。用这种方式获得的蛋白质构象稳定, 也可以辅助发现与目的蛋

白产生相互作用的其他未知蛋白^[47-48]。

对于内源纯化来讲, 敲入标签的选择至关重要。标签的选用一般会根据纯化蛋白的宿主细胞不断优化, 使得选用的标签在宿主细胞中高效特异, 且宿主细胞内不会包含这些标签的蛋白识别点, 进而最大限度减少宿主自身蛋白对纯化结果的影响, 同时敲入的标签不应影响目的蛋白的结构与功能。当选用细菌类宿主细胞如嗜热性球菌时, 因其胞内组氨酸的蛋白含量及比例相对较低, 因而可选择技术成熟、应用广泛、成本低效率高、纯化条件温和而步骤简单的His融合标签。2018年, YU等^[47]在嗜热性球菌MW0414菌株中向氢气进化膜结合氢化酶(Hydrogen gas-evolving membrane-bound hydrogenase, MBH)复合物的MbhJ亚基N-端一步敲入His9亲和标签, 利用镍柱纯化获得内源性MBH蛋白复合物。当选用较细菌更为高等的真菌类细胞作为宿主如酵母真菌时, 为了更好地维持融合标签的特异性, 纯化获得良好的蛋白, 一般选用串联亲和纯化(tandem affinity purification, TAP)^[50-51]方法, 即在目的蛋白处knock-in两种不同的融合标签, 通过两次亲和层析实现蛋白的分离纯化。2020年, WANG等^[50]通过向酵母细胞菌株CB010的Spt20蛋白C-端敲入TAP标签而纯化得到内源性转录辅助激活复合物SAGA复合物。2017年, KARUPPASAMY等^[53]在Bit2缺陷的酵母菌株中对雷帕霉素(target of rapamycin, TOR)编码基因*Bit61*进行TAP标签标记, 最终获得TORC2蛋白复合物的结构数据。除TAP标签外, Flag标签也常被用于酵母细胞内源标记蛋白^[54-55], WANG等^[56]向复合物组分之一的SNF6蛋白C-端内源敲入3-Flag



A: 蛋白复合物各组分分别在不同宿主细胞内表达纯化, 并在体外孵育组装形成稳定复合物; B: 蛋白复合物各组分基因序列克隆至多个质粒或单一质粒中, 将这些质粒共转染入同一宿主细胞内, 使各组分蛋白在体内组装并通过尺寸排阻色谱法分离获得高纯度稳定复合物; C: 利用 CRISPR-Cas9 基因编辑技术向小鼠或细胞内特定位置敲入融合标签, 收集小鼠组织或细胞, 裂解后根据标签种类选择适宜亲和层析柱, 纯化获得稳定内源蛋白复合物。

A: plasmids containing the components of the protein complex are transfected into distinct host cells separately for expression and purification. These protein components are then incubated and assembled under suitable conditions *in vitro* to form stable protein complexes; B: multiple plasmids or a single plasmid cloned with protein complex components are co-transfected into the same host cell, enabling the assembly of proteins *in vivo*. The monomeric proteins are subsequently removed by size-exclusion chromatography to obtain pure protein complexes; C: the fusion tags knock-in mice or mammalian cells are generated by CRISPR-Cas9 gene editing, by which a specific tag is inserted at either the C or N terminus of the protein complexes. Animal tissues or cells are collected and lysed, followed by purification of the protein complexes of interest using appropriate affinity chromatography columns based on the protein tag.

图1 蛋白质复合物的三种纯化技术

Fig.1 Three major purification techniques for protein complexes

标签, 获得内源性酵母SWI/SNF复合物, 并利用低温冷冻电镜解析蛋白结构。

对于真核细胞及哺乳动物的内源蛋白纯化, 则更多地选用GFP标签、HA标签、Flag标签和GST标签等, 一般同时选用两种融合标签, 通过两步纯化获得单一纯净的目的蛋白复合物。一篇在2022年发表于*Nature*上的文章^[57]报道其利用CRISPR-knock-in技术, 向雄鼠精子阳离子通道(the cation channel of sperm, CatSper)复合物的主要组分CATSPER1蛋白的N-端插入3-FLAG-EGFP-TEV标签, 从而获得内源性CatSper复合物, 并通过冷冻电镜确定复合物的高分辨三维结构(图1)。

综上, 通过CRISPR-knock-in技术向内源目的蛋白中添加融合标签进行蛋白纯化这一方法推进了内

源蛋白质复合物纯化的进程, 这一技术使得不稳定、易降解、组分未知及异源表达相容性差的蛋白复合物的纯化成为可能, 由于融合标签被插入至天然蛋白中, 因而纯化出的蛋白更接近生理状态下的天然构象, 最大限度地保留了蛋白在细胞内的初始状态。但通过knock-in技术内源敲入标签纯化蛋白也有限制, 如获得knock-in动物或细胞周期较长、目的蛋白在正常细胞或组织中的表达水平较低, 纯化时往往需要大量扩增, 纯化成本高昂以及真核细胞内源蛋白较多, 容易在纯化中造成干扰等(表2)。

4 讨论与展望

在目前的结构生物学研究中, 蛋白纯化仍是非常基础且重要的一环, 稳定均一蛋白质的体外获得

表2 三种纯化方式对比
Table 2 Comparison of three purification methods

方法 Methods	优点 Advantages	限制 Limitations
Assembly <i>in vitro</i>	Easy and efficient; mature technology; stabilized protein complex components	Requires adequate knowledge of the target protein complexes; difficulty in harmonizing experimental processes; longer experimental cycles for complicated protein complexes
Assembly <i>in vivo</i>	A relatively natural environment; catering to different protein expression preferences; multiple proteins can be expressed simultaneously; low cost; easy operation	Protein expression level may be different for respective components; uncertainty about the stability of protein complexes; possibility of impurity
Endogenous purification	Applicable to protein complexes with unknown components; proteins obtained are closer to the natural conformation	High cost; long experimental cycles; low expression of protein complexes

以及蛋白结构解析技术的发展辅助推进了科学家们从原子水平探究蛋白间相互作用机制、解析生物体复杂信号网络的进程。高分辨率的蛋白结构一直是研究者们目标,而蛋白结构的清晰度与蛋白样本的质量密切相关。当前,蛋白质复合物的纯化存在许多难点,包括复合物各组分稳定性未知、组分间相互作用力较弱或结合时间较短、复合物组分及结合条件未知等,这些因素使得完整蛋白复合物的获得异常困难。

外源过表达获取蛋白复合物的方式主要有体外组装与体内组装两种,体外组装的方法简单高效,适合于简单的二元或组分已知且易表达的多元复合物的体外纯化,但因复合物的组装发生在体外,因而对于组装条件的要求十分严格。将复合物各组分共转染或形成融合蛋白转染入宿主细胞内,使组装发生在内源环境中这一方法是对体外组装的改进,宿主细胞能够为目的蛋白复合物的形成提供相对天然的环境与必需的物质,且复合物的纯化发生在组装后,这一顺序也提高了最终获得蛋白复合物的稳定性。但在蛋白表达时,不同组分在同一宿主中的表达量有所差异,复合物组装的信号通路部分未知,这些因素都会影响复合物的质量。为了减少这些因素的干扰,科学家们提出了另一种蛋白纯化方法——内源纯化,直接将内源蛋白复合物从其本身的固有位置纯化至体外,由此得到具有天然构象的蛋白复合物,科学家们利用该技术解析核孔复合物等组分未知的大分子复合物蛋白结构,推动了生命科学研究的发展。但这种方法对纯化技术要求较高,纯化条件有诸多限制,纯化成本也较高。

蛋白纯化技术发展至今,已经帮助科学家们纯化解析了多种蛋白质复合物的结构,极大地促进了结构生物学领域的发展和进步。从原子水平观察蛋白间相互作用的分子机制也为蛋白功能探索和信号通路的研究提供基础,为后续药物靶标的研发提供理论依据。蛋白纯化技术目前尚存在一些局限,还有待完善,相信在科学工作者们的共同努力下,我们定能突破重重难关,取得突破性的成果。

参考文献 (References)

- [1] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. *Nature*, 2021, 596(7873): 583-9.
- [2] ALBER F, FORSTER F, KORKIN D, et al. Integrating diverse data for structure determination of macromolecular assemblies [J]. *Annu Rev Biochem*, 2008, 77: 443-77.
- [3] BRABERG H, ECHEVERRIA I, BOHN S, et al. Genetic interaction mapping informs integrative structure determination of protein complexes [J]. *Science*, 2020, 370(6522): eaaz4910.
- [4] YANG Z, ZENG X, ZHAO Y, et al. AlphaFold2 and its applications in the fields of biology and medicine [J]. *Signal Transduct Target Ther*, 2023, 8(1): 115.
- [5] RUCKOVA E, MULLER P, VOJTESEK B. Protein expression and purification [J]. *Klin Onkol*, 2014, 27(Suppl 1): S92-7.
- [6] PANDEY A, SHIN K, PATTERSON R E, et al. Current strategies for protein production and purification enabling membrane protein structural biology [J]. *Biochem Cell Biol*, 2016, 94(6): 507-27.
- [7] XUE H, YAO T, CAO M, et al. Structural basis of nucleosome recognition and modification by MLL methyltransferases [J]. *Nature*, 2019, 573(7774): 445-9.
- [8] TORRENTS DE LA PEÑA A, SLIEPEN K, ESHUN-WILSON L, et al. Structure of the hepatitis C virus E1E2 glycoprotein complex [J]. *Science*, 2022, 378(6617): 263-9.
- [9] SUSAC L, VUONG M T, THOMAS C, et al. Structure of a fully assembled tumor-specific T cell receptor ligated by pMHC [J].

- Cell, 2022, 185(17): 3201-13.e19.
- [10] YUAN J, CHEN K, ZHANG W, et al. Structure of human chromatin-remodelling PBAF complex bound to a nucleosome [J]. Nature, 2022, 605(7908): 166-71.
- [11] ALBADRI S, DEL BENE F, REVENU C. Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish [J]. Methods, 2017, doi: 10.1016/j.ymeth.2017.03.005.
- [12] BANAN M. Recent advances in CRISPR/Cas9-mediated knock-ins in mammalian cells [J]. J Biotechnol, 2020, 308: 1-9.
- [13] HSU P D, LANDER E S, ZHANG F. Development and applications of CRISPR-Cas9 for genome engineering [J]. Cell, 2014, 157(6): 1262-78.
- [14] EDELMANN F T, NIEDNER A, NIESSING D. Production of pure and functional RNA for *in vitro* reconstitution experiments [J]. Methods, 2014, 65(3): 333-41.
- [15] SHARIF H, WANG L, WANG W L, et al. Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome [J]. Nature, 2019, 570(7761): 338-43.
- [16] LU J Y, BERGMAN N, SADRI N, et al. Assembly of AUF1 with eIF4G-poly(A) binding protein complex suggests a translation function in AU-rich mRNA decay [J]. RNA, 2006, 12(5): 883-93.
- [17] ALTMAN J D, MOSS P A, GOULDER P J, et al. Phenotypic analysis of antigen-specific T lymphocytes [J]. Science, 1996, 274(5284): 94-6.
- [18] HUGHES T E T, PUMROY R A, YAZICI A T, et al. Structural insights on TRPV5 gating by endogenous modulators [J]. Nat Commun, 2018, 9(1): 4198.
- [19] IVIC N, POTOENJAK M, SOLIS-MEZARINO V, et al. Fuzzy interactions form and shape the histone transport complex [J]. Mol Cell, 2019, 73(6): 1191-203.e6.
- [20] GONG X, QIAN H, SHAO W, et al. Complex structure of the fission yeast SREBP-SCAP binding domains reveals an oligomeric organization [J]. Cell Res, 2016, 26(11): 1197-211.
- [21] VASSYLYEVA M N, KLYUYEV S, VASSYLYEV A D, et al. Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins [J]. Proc Natl Acad Sci USA, 2017, 114(26): E5138-E47.
- [22] CAO T, SUN L, JIANG Y, et al. Crystal structure of a nuclear actin ternary complex [J]. Proc Natl Acad Sci USA, 2016, 113(32): 8985-90.
- [23] PAN D, NAKATSU T, KATO H. Crystal structure of peroxisomal targeting signal-2 bound to its receptor complex Pex7p-Pex21p [J]. Nat Struct Mol Biol, 2013, 20(8): 987-93.
- [24] ZHOU L, ZHOU Y, HANG J, et al. Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex [J]. Cell Res, 2014, 24(4): 497-500.
- [25] GRABARCZYK D B. Crystal structure and interactions of the Tof1-Csm3 (Timeless-Tipin) fork protection complex [J]. Nucleic Acids Res, 2020, 48(12): 6996-7004.
- [26] BLOK N B, TAN D, WANG R Y, et al. Unique double-ring structure of the peroxisomal Pex1/Pex6 ATPase complex revealed by cryo-electron microscopy [J]. Proc Natl Acad Sci USA, 2015, 112(30): E4017-25.
- [27] PERRAKIS A, ROMIER C. Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview [J]. Methods Mol Biol, 2008, 426: 247-56.
- [28] MATEJA A, PADUCH M, CHANG H Y, et al. Structure of the Get3 targeting factor in complex with its membrane protein cargo [J]. Science, 2015, 347(6226): 1152-5.
- [29] LETOQUART J, VAN TRAN N, CAROLINE V, et al. Insights into molecular plasticity in protein complexes from Trm9-Trm112 tRNA modifying enzyme crystal structure [J]. Nucleic Acids Res, 2015, 43(22): 10989-1002.
- [30] ARORA K, CORBETT K D. The conserved XPF:ERCC1-like Zip2:Spo16 complex controls meiotic crossover formation through structure-specific DNA binding [J]. Nucleic Acids Res, 2019, 47(5): 2365-76.
- [31] WANG J, LI L, MING Z, et al. Crystal structure of the Epo1-Bem3 complex for bud growth [J]. Int J Mol Sci, 2021, 22(8): 3812.
- [32] DONG D, ZHENG L, LIN J, et al. Structural basis of assembly of the human T cell receptor-CD3 complex [J]. Nature, 2019, 573(7775): 546-52.
- [33] XIE Y, LORD C L, CLARKE B P, et al. Structure and activation mechanism of the yeast RNA Pol II CTD kinase CTDK-1 complex [J]. Proc Natl Acad Sci USA, 2021, 118(3): e2019163118.
- [34] ZHUANG Y, XU P, MAO C, et al. Structural insights into the human D1 and D2 dopamine receptor signaling complexes [J]. Cell, 2021, 184(4): 931-42.e18.
- [35] REN D, WANG F, HE X, et al. Construction of bioactive chimeric MHC class I tetramer by expression and purification of human-murine chimeric MHC heavy chain and beta(2)m as a fusion protein in *Escherichia coli* [J]. Protein Expr Purif, 2006, 50(2): 171-8.
- [36] BROHAWN S G, SCHWARTZ T U. Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice [J]. Nat Struct Mol Biol, 2009, 16(11): 1173-7.
- [37] FITZGERALD D J, BERGER P, SCHAFFITZEL C, et al. Protein complex expression by using multigene baculoviral vectors [J]. Nat Methods, 2006, 3(12): 1021-32.
- [38] GERHOLD C B, WINKLER D D, LAKOMEK K, et al. Structure of Actin-related protein 8 and its contribution to nucleosome binding [J]. Nucleic Acids Res, 2012, 40(21): 11036-46.
- [39] ZAHM J A, JENNI S, HARRISON S C. Structure of the Ndc80 complex and its interactions at the yeast kinetochore-microtubule interface [J]. Open Biol, 2023, 13(3): 220378.
- [40] STANGIER M M, KUMAR A, CHEN X, et al. Structure-function relationship of the Bik1-Bim1 complex [J]. Structure, 2018, 26(4): 607-18.e4.
- [41] WU X, SIGGEL M, OVCHINNIKOV S, et al. Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex [J]. Science, 2020, 368(6489): eaaz2449.
- [42] POWIS K, ZHANG T, PANCHAUD N, et al. Crystal structure of the Ego1-Ego2-Ego3 complex and its role in promoting Rag GTPase-dependent TORC1 signaling [J]. Cell Res, 2015, 25(9): 1043-59.
- [43] GU Y, LI H, DONG H, et al. Structural basis of outer membrane protein insertion by the BAM complex [J]. Nature, 2016, 531(7592): 64-9.
- [44] LUDTKE T H, KLEPPA M J, RIVERA-REYES R, et al. Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18 protein in 293 and A549 cells [J]. Biochem J, 2022, 479(1): 91-109.
- [45] ZHANG Y, LI S, ZENG C, et al. Molecular architecture of the

- luminal ring of the *Xenopus laevis* nuclear pore complex [J]. *Cell Res*, 2020, 30(6): 532-40.
- [46] ZHU X, HUANG G, ZENG C, et al. Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex [J]. *Science*, 2022, 376(6598): eabl8280.
- [47] MACLEOD G, VARMUZA S. Tandem affinity purification in transgenic mouse embryonic stem cells identifies DDOST as a novel PPP1CC2 interacting protein [J]. *Biochemistry*, 2012, 51(48): 9678-88.
- [48] TIRARD M, BROSE N. Systematic localization and identification of SUMOylation substrates in knock-in mice expressing affinity-tagged SUMO1 [J]. *Methods Mol Biol*, 2016, 1475: 291-301.
- [49] YU H, WU C H, SCHUT G J, et al. Structure of an ancient respiratory system [J]. *Cell*, 2018, 173(7): 1636-49, e16.
- [50] HONEY S, SCHNEIDER B L, SCHIELTZ D M, et al. A novel multiple affinity purification tag and its use in identification of proteins associated with a cyclin-CDK complex [J]. *Nucleic Acids Res*, 2001, 29(4): E24.
- [51] RIGAUT G, SHEVCHENKO A, RUTZ B, et al. A generic protein purification method for protein complex characterization and proteome exploration [J]. *Nat Biotechnol*, 1999, 17(10): 1030-2.
- [52] WANG H, DIENEMANN C, STUTZER A, et al. Structure of the transcription coactivator SAGA [J]. *Nature*, 2020, 577(7792): 717-20.
- [53] KARUPPASAMY M, KUSMIDER B, OLIVEIRA T M, et al. Cryo-EM structure of *Saccharomyces cerevisiae* target of rapamycin complex 2 [J]. *Nat Commun*, 2017, 8(1): 1729.
- [54] LI N, ZHAI Y, ZHANG Y, et al. Structure of the eukaryotic MCM complex at 3.8 Å [J]. *Nature*, 2015, 524(7564): 186-91.
- [55] LI Z, CHEN S, ZHAO L, et al. Near-atomic structure of the inner ring of the *Saccharomyces cerevisiae* nuclear pore complex [J]. *Cell Res*, 2022, 32(5): 437-50.
- [56] WANG C, GUO Z, ZHAN X, et al. Structure of the yeast Swi/Snf complex in a nucleosome free state [J]. *Nat Commun*, 2020, 11(1): 3398.
- [57] LIN S, KE M, ZHANG Y, et al. Structure of a mammalian spermatogenesis channel complex [J]. *Nature*, 2021, 595(7869): 746-50.