



核磁共振波谱法在蛋白质三维结构解析中的应用*

尹 林 申峻丞 杨立群**

(中山大学化学学院, 聚合物复合材料及功能材料教育部重点实验室, 广东省高性能树脂基复合材料重点实验室, 广州 510275)

摘要 蛋白质特定的三维结构与其生物功能密切相关, 因此, 研究蛋白质的三维结构有助于揭示其生物功能机制。将核磁共振 (NMR) 波谱法应用于研究溶液状态下蛋白质的三维结构, 能够更加准确地揭示蛋白质结构与生物功能之间的关系。本文综述了 NMR 解析蛋白质三维结构的理论和技术方法, 以及 NMR 结合其他生物物理手段, 并辅以分子建模计算法研究蛋白质三维结构的研究进展和最新方法, 为精准解析蛋白质的三维结构提供思路及策略。

关键词 核磁共振波谱法, 蛋白质三维结构, 同位素标记蛋白质, 结构约束, 蛋白质分子激发态结构, 超大蛋白质复合体的三维结构

中图分类号 Q5, Q6, Q7

DOI: 10.16476/j.pibb.2021.0065

每一种蛋白质都有自己特殊的空间结构, 被称为三维结构或空间构象。蛋白质作为一种重要的生物大分子, 参与进行大多数生命活动 (如酶催化代谢、免疫调节、细胞信息传递、载体运输以及能量供应等), 所以, 研究蛋白质的三维结构对揭示其生物功能至关重要。这些理论有助于更深层次认识蛋白质在健康生理活动和疾病产生过程中发挥的作用和机制, 也将促进相关疾病的临床诊治和药物设计。

尽管 X 射线衍射法已被广泛用于解析蛋白质三维结构, 然而, 该方法仍然存在一些缺陷, 例如 X 射线衍射法只能用于研究蛋白质晶体的静态结构, 并且一些蛋白质因结构的无序性而难以形成晶体 (或形成的晶体存在缺陷而不足以 X 射线衍射法研究)。相比之下, 核磁共振 (nuclear magnetic resonance, NMR) 波谱法更适合用于研究溶液状态下蛋白质的三维结构, 能够更加准确地揭示蛋白质结构与生物功能之间的自然规律^[1]。

2D NMR 技术已广泛用于解析较小分子质量 (≤ 10 ku) 蛋白质的三维结构^[2]。虽然 3D 和 4D NMR 技术逐渐被用于解析分子质量较高的蛋白质的三维结构, 但是, NMR 测试技术仍然面临两大

难题而限制其在蛋白质三维结构研究中的应用: 第一, 随着蛋白质的氨基酸数目和分子质量的增加, 造成 NMR 共振信号峰严重重叠; 第二, 横向弛豫速率 (R_2) 的加快, 致使 NMR 信号峰增宽和分辨率变差。因此, 本文主要概述 NMR 解析蛋白质三维结构的理论和技术方法, 以及 NMR 结合其他生物物理手段, 并辅以分子建模计算法研究蛋白质三维结构的研究进展和最新方法, 为精准解析蛋白质的三维结构提供思路及策略。

1 核磁共振波谱法解析蛋白质三维结构的理论概述

自旋量子数不为零的原子核 (主要为 ^1H 、 ^{13}C 、 ^{15}N 、 ^{19}F 和 ^{31}P 等), 它们在外加静磁场中自旋运动时存在着不同的能级, 相邻两能级间的能量差 (ΔE) 为:

$$\Delta E = \gamma \hbar H_0 \quad (1)$$

式中, γ 为磁旋比, $\hbar = h/2\pi$ (h 为普朗克常数), H_0

* 广东省自然科学基金 (2017B030311007) 资助项目。

** 通讯联系人。

Tel: 020-39339796, E-mail: yanglq@mail.sysu.edu.cn

收稿日期: 2021-08-31, 接受日期: 2021-10-12

为静磁场强度。

在外加静磁场中,原子核绕其自旋轴旋转,自旋轴与静磁场保持某一夹角而围绕静磁场方向进动的频率(ω_0)为:

$$\omega_0 = \gamma H_0 \quad (2)$$

当用频率为 ω_0 的射频照射静磁场中自旋运动的原子核时,原子核才能有效地吸收射频辐射的能量,从低能级跃迁至高能级,实现核磁共振。所以,核磁共振波谱来源于原子核能级间的跃迁,即用一定射频的电磁波对样品进行照射,使特定结构环境中的原子核发生共振跃迁,记录发生核磁共振时的信号峰位置和强度,得到核磁共振波谱,其共振信号反映了官能团和构象等分子结构信息。通常将 ^1H 、 ^{13}C 和 ^{15}N 原子核的磁共振技术应用于蛋白质的结构解析,例如通过NMR实验获得化学位移、弛豫时间、耦合常数、核奥弗豪泽效应(nuclear Overhauser effect, NOE)等参数,用于研究蛋白质的三维结构及其分子动力学等^[3-7]。

NMR解析蛋白质三维结构的流程图如图1a所示。为了避免蛋白质共振信号峰的重叠,提高其分辨率,一般需要制备 ^{15}N 、 ^{13}C 或 ^2H 标记的蛋白质,采用 $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 三共振多维核磁技术采集实验数据,对共振信号进行分析、筛选及归属,主要相关的NMR技术包括HNCA、HN(CO)CA、HNCO、HN(CA)CO、HNCACB、HN(CO)CACB、

HNCACO、HNCOCA、 $\text{HNCO}_{i-1}\text{CA}_i$ 等^[8-20]。根据NMR软件结合NMR技术所获得的 $^1\text{H}^{\text{N}}$ 、 ^{15}N 、 ^{13}CO 、 $^{13}\text{C}^{\alpha}$ 和 $^{13}\text{C}^{\beta}$ 化学位移、耦合常数和质子间距离,推算蛋白质主链中氨基酸的序列及其二面角结构参数,得出蛋白质的二级结构。通过测定NOE、顺磁弛豫增强(paramagnetic relaxation enhancement, PRE)以及残余偶极耦合(residual dipolar coupling, RDC)等NMR参数获得蛋白质的三维结构约束参数(包括原子间空间距离、化学键空间取向等),进一步结合NMR计算软件解析蛋白质的三维结构。

以常用的3D $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 三共振HNCA和HN(CO)CANMR为例,蛋白质分子中的氨基酸通过磁化矢量传递(图1b①),HNCA可测定同一个氨基酸内的 $^1\text{H}^{\text{N}}$ 和 ^{15}N 原子与 $^{13}\text{C}^{\alpha}$ 原子的关联(以第*i*个氨基酸为例),以及第*i*个氨基酸的 $^1\text{H}^{\text{N}}$ 和 ^{15}N 原子与第*i*-1个氨基酸 $^{13}\text{C}^{\alpha}$ 原子的关联;HN(CO)CA通常只提供同一个氨基酸内的 $^1\text{H}^{\text{N}}$ 和 ^{15}N 原子与邻近氨基酸 $^{13}\text{C}^{\alpha}$ 原子之间的关联信息。在蛋白质样品3D NMR谱图中(图1b②),通过上述关联信息解析蛋白质主链中氨基酸的序列结构(图1b③),根据质子间距离和二面角结构参数(图1b④),并采用化学位移预测法解析二级结构。结合结构约束参数,通过NMR计算软件进行计算可获得蛋白质的三维结构。

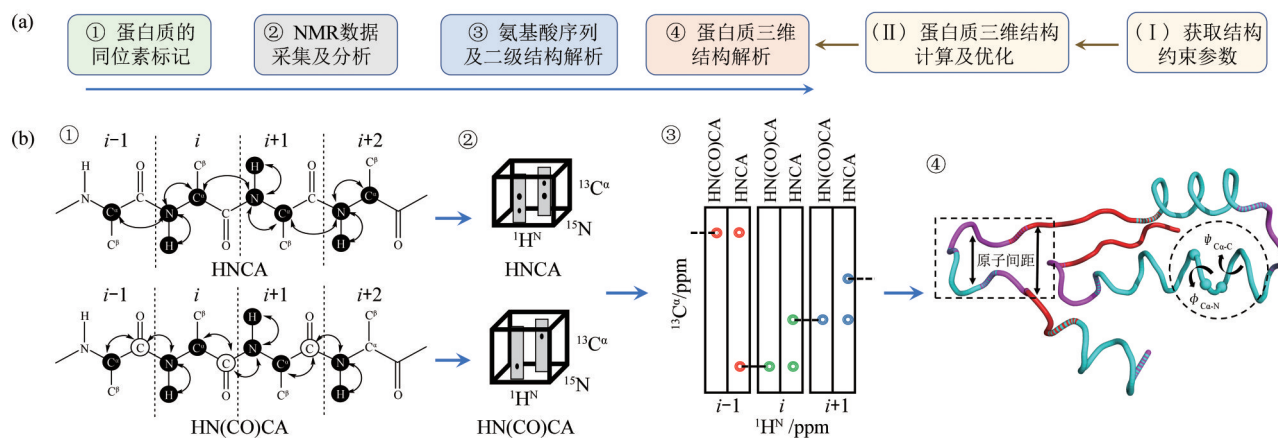


Fig. 1 Process and schemes of investigating protein three-dimensional structure

图1 解析蛋白质三维结构的流程及示意图

(a) NMR解析蛋白质三维结构的流程图。(b) 3D $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 三共振HNCA和HN(CO)CANMR解析蛋白质三维结构的示意图: ① ^1H 、 ^{13}C 和 ^{15}N 原子核相关自旋系统及其NMR数据采集及共振信号分析; ②蛋白质样品3D NMR谱图示意图; ③解析氨基酸序列结构; ④二面角($\phi_{\text{Ca-N}}$ 和 $\psi_{\text{Ca-C}}$)及质子间距离示意图。

2 研究蛋白质三维结构的核磁共振波谱法

2.1 蛋白质的同位素标记

蛋白质的同位素标记包括以下过程: 构建含有目标蛋白基因序列的质粒, 将其导入宿主生物体, 在含有同位素标记物的培养基中进行培养, 通过宿主生物表达得到同位素标记蛋白质, 采用生化技术进行分离纯化, 获得高纯度的同位素标记蛋白质。

2.1.1 标记方法

均匀同位素标记法被广泛用于蛋白质的同位素标记^[21]。它主要以¹⁵N标记的NH₄Cl或(NH₄)₂SO₄作为唯一氮源, ¹³C标记的葡萄糖、甘油、甲醇或丙酮酸作为唯一的碳源, 对蛋白质进行均匀的¹⁵N/¹³C双标记。为了提高NMR谱灵敏度 and 分辨率, 通常在含D₂O/H₂O的培养基中制备全氘代或部分氘代的¹⁵N/¹³C/²H三标记蛋白质样品, 进行三共振多维核磁测试。

为了克服大量质子被氘取代而干扰NOE法解析蛋白质结构的问题, 发展了对4种甲基氨基酸(Ala、Val、Leu、Ile)选择性地甲基质子化法^[22]。例如, 在含有¹³C/²H标记葡萄糖、¹⁵N标记NH₄Cl和D₂O的细菌培养基中, 加入3-²H,¹³C- α -酮异戊酸(α -ketoisovalerate)制备含有¹³CH₃-标记Val和Leu残基的¹⁵N/¹³C/²H蛋白质^[23], 加入3,3-²H,¹³C- α -酮丁酸(α -ketobutyrate)制备含有¹³CH₃-标记Ile残基的¹⁵N/¹³C/²H蛋白质^[24]。

片段同位素标记法有助于降低NMR信号峰的重叠程度, 用于无序蛋白质(disordered protein)结构的测定、高分子质量蛋白质的结构解析以及蛋白质构象变化的研究等^[25-28]。该方法通过对两个或以上的蛋白质片段进行分别表达, 对不同的组分采用不同的同位素标记(¹⁵N/¹³C/²H), 再将这些蛋白质片段连接起来, 组成具有同位素标记片段的完整目标蛋白, 主要有表达蛋白连接法(expressed protein ligation, EPL)、蛋白反式剪接法(protein trans-splicing, PTS)和天然化学连接法(native chemical ligation, NCL)等^[29-31]。

2.1.2 表达系统

通常采用的表达系统为原核生物表达系统、真核生物表达系统以及无细胞表达系统。大肠杆菌(*E. coli*)是原核生物表达系统中代表性的表达宿主, 它能在各种同位素标记环境中表达蛋白质, 成本较低。所以, 在NMR研究中大肠杆菌表达系统最为常用, 主要以pET为表达载体, 异丙基- β -D-

硫代半乳糖苷(IPTG)为诱导剂。目前研究工作主要集中在通过改造该系统中的质粒来提高蛋白质的溶解性及产量^[32]。

由于大肠杆菌表达系统对于一些含有特殊结构的蛋白质(如含二硫键的蛋白质等)普适性较差, 所以发展起来了真核生物表达系统(包括酵母细胞、昆虫细胞和哺乳动物细胞表达系统)^[33]。尽管毕赤酵母(*Pichia pastoris*)常用于表达¹⁵N/¹³C/²H标记的蛋白质, 但是, 仍然存在一些缺陷, 例如在氘代介质中表达时间较长以及选择性地甲基质子化效率较低等^[34-35]。昆虫细胞和哺乳动物细胞表达系统需要使用含同位素标记氨基酸的培养基, 导致制备同位素标记蛋白质的成本较为昂贵^[36-37]。一些对宿主细胞具有毒性的蛋白质, 可采用无细胞表达系统, 其特点是具有较高的蛋白质表达效率^[38-39]。

2.1.3 纯化技术

采用同位素标记法及在各种表达系统中得到的菌体, 需要利用生化技术进一步分离提纯才能获得高纯度的同位素标记蛋白样品, 主要包括菌体内含蛋白的释放、总蛋白的富集、层析色谱法分离提纯、蛋白质的浓缩以及纯度鉴定等过程。其中层析色谱分离是最重要的一步, 它直接决定同位素标记蛋白样品的纯度和产率。根据同位素标记蛋白不同的物理性质(如等电点、溶解度、电荷性质等), 可采用多种层析色谱分离技术, 例如亲和层析色谱法(如Ni-NTA和Glutathione Sepharose色谱法)、离子交换色谱法(如DEAE色谱法)、体积排阻色谱法(如Superdex、Sephacryl、Sephacryl和Sephadex色谱法)。

2.2 核磁共振实验数据的采集及其解析的软件

2.2.1 核磁共振实验数据采集

近年来针对核磁共振技术采集时间长和信噪比较低等问题, 发展了一些新的NMR实验数据采集技术, 如非均匀采样(non-uniform sampling)^[40]、SOFAST(band-selective optimized-flip-angle short-transient)^[41-42]、ADAPT-NMR(assignment-directed data collection algorithm utilizing a probabilistic toolkit in NMR)^[43]、BEST(band-selective excitation short-transient)^[44-45]、GFT(G-matrix Fourier transform)^[46]、投影重建(projection reconstruction)^[47]等。这些技术能实现快速采样, 信噪比较高, 更有利于研究溶液状态下不稳定的蛋白质三维结构。

非均匀采样技术不仅保持了较高的信噪比，而且将实验数据采集时间缩短为原来的1/2~1/6，已得到了广泛应用^[48-49]。它主要是在间接维度(indirect dimension)中对NMR数据进行随机选择性采样，结合特定采样限制因素(例如在每个间接维度下每个采集时间应至少测量1次等)，将得到的低分辨率数据经过重建算法(reconstruction algorithm)处理，在一定程度上得到与均匀采样技术相当的NMR数据^[50]。数据重建算法包括多维分解法(multidimensional decomposition)^[48]、最大熵重构法(maximum entropy reconstruction)^[49]、非等间距数据的傅立叶变换算法(Fourier transform algorithms for non-equispaced data)^[51]、机器学习法(machine learning)^[52]、稀疏多维迭代线形增强法(sparse multidimensional iterative lineshape-enhanced, SMILE)^[53]等。

2.2.2 解析核磁共振信号的软件

由NMR实验数据得到的谱图，需要对共振信号进行归属，经过结构计算和评估才能获得较为准确的蛋白质三维结构。然而，对于一些分子质量较高的蛋白质，由于其信号峰的数量显著增多，导致NMR谱图解析较为困难。目前已出现了多种化学位移自动归属软件，用于分析蛋白质主链和侧链的结构以及NOESY信号，包括CYANAY^[54-56]、PINE-SPARKY.2^[57]、APSY^[58-60]、PONDEROSA^[61]、NMRNet^[62]、ADAPT-NMR^[43]等。化学位移自动归属的在线服务器主要为I-PINE web server (<http://i-pine.nmrfam.wisc.edu/>)^[63]。Pritisanac等^[64]改进了CYANAY中的FLYA算法，能自动归属超大分子质量蛋白质甲基的NOESY信号。为了获得蛋白质的三维结构，需要根据¹H、¹³C、¹⁵N化学位移及NOE等结构约束参数、采用计算软件计算蛋白质的二级结构和三维结构(表1)。

Table 1 Calculation software of protein secondary structure unit and three-dimensional structure

表1 蛋白质二级结构单元和三维结构的计算软件

软件	功能简介	网址	参考文献
TALOS+	结构模拟软件，依据化学位移归属经验性预测蛋白质主链的二面角(即C _α —N之间的旋转角φ和C _α —C之间的旋转角ψ)	https://spin.niddk.nih.gov/NMRPipe/talos/	[65]
CYANA	结构计算软件，在已知氨基酸序列的前提下，根据NMR实验获取的约束条件(二面角约束、原子距离约束等)进行蛋白质结构计算，附带自动归属NOESY信号的功能	http://www.cyana.org/wiki/index.php/Tutorials	[56]
CNS	结构计算软件，根据NOE、耦合常数、化学位移和偶极耦合等约束条件进行蛋白质结构计算	http://cns-online.org/v1.3/	[66]
Amber	生物分子模拟和分析程序，根据NOE、化学位移、残基偶极耦合常数和扭转角等约束条件对蛋白质结构进行优化	http://ambermd.org/index.php	[67]
Aria	NOE信号自动归属和蛋白质结构计算软件，采用迭代结构计算方法提高NOE信号归属的效率	http://aria.pasteur.fr/	[68]
PROCHECK	蛋白质结构质量评估在线软件，评估目的蛋白质结构与高分辨率的蛋白质晶体结构之间各结构参数的差异(包括键长、二面角及原子间距离等)	https://servicesn.mbi.ucla.edu/PROCHECK/	[69]

2.3 核磁共振波谱法解析蛋白质的氨基酸序列和二级结构单元

获得¹⁵N、¹³C或²H同位素标记的高纯度蛋白质样品后，在解析三维结构之前需要进行NMR实验的可行性论证，包括选取适合的缓冲溶液体系和测试温度以保证在测试过程中蛋白质结构和性质的稳定性，通过考察NMR线宽的浓度依赖性来判断蛋白质是否发生聚集等。如果NMR谱线太宽(或NMR峰重叠严重)，则需要改变实验条件进行改善。在蛋白质¹H NMR谱图中，甲基及脂肪族氨基

酸侧链质子信号主要分布在0~4 ppm，主链及侧链的酰胺质子主要分布在7~10 ppm，芳香环上的质子信号主要分布在6~8 ppm。

2.3.1 氨基酸残基自旋系统的辨识及蛋白质的氨基酸序列解析

常见的20种氨基酸残基中有15种残基即为一个自旋系统，其余5种氨基酸残基则含有多个自旋系统。根据蛋白质分子质量的大小，通常采用以下两种方法进行氨基酸残基自旋系统的辨识以及氨基酸序列结构的解析。

分子质量低于 14 ku 的蛋白质一般不需要进行同位素标记, 其氨基酸自旋系统可用 2D ^1H NMR 技术 (如 COSY、DQF-COSY、RCT-COSY 和 TOCSY) 进行识别, 这些技术与 NOESY 联用可识别氨基酸的序列结构。其中 COSY 类技术提供同一氨基酸的酰胺质子与 H^α 质子 ($\text{NH}_i\text{-}^\alpha\text{H}_i$) 的关联信息, NOESY 提供氨基酸内酰胺质子与邻近氨基酸 H^α 质子 ($\text{NH}_{i+1}\text{-}^\alpha\text{H}_i$) 的关联信息。由于 NOESY 和 COSY 谱图都具有对角对称性, 因此, 将这两种谱图的各一半沿着对角线叠合组成一张 NOESY-COSY 谱^[70], 在 COSY 谱中从 $\text{NH}_i\text{-}^\alpha\text{H}_i$ 信号峰的氨基酸残基自旋系统出发, 分别作水平线和垂直线, 得到 NH_i 和 $^\alpha\text{H}_i$ 化学位移, 以 $^\alpha\text{H}_i$ 原子的化学位移作为起点做水平线, 从 NOESY 谱中找到 NH_{i+1} 的氨基酸残基自旋系统, 得出各氨基酸残基自旋系统的顺序。

对于较高分子质量的蛋白质, 因为自旋系统信号峰数目增加而容易产生信号峰的重叠, 所以, 需要制备 $^{15}\text{N}/^{13}\text{C}$ 或 $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ 标记的蛋白质样品, 采用 $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 异核多维 NMR 进行氨基酸自旋系统的识别, 进一步解析氨基酸的序列。自 Bax 等^[8] 将 $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 三共振实验应用于解析钙调蛋白 (16.7 ku) 的结构以来, 出现了更多的三共振脉冲技术: a. 分子质量约为 10~30 ku 的单链蛋白质, 常用的三共振实验技术主要有 HNCA、HN(CO)CA、HNCO、HN(CA)CO、CBCA(CO)NH、CBCANH、HNCACB 和 HN(CO)CACB 等; b. 分子质量高于 30 ku 的蛋白质, 则需在上述基础上联用 TROSY 以及 4D、5D 和 6D NMR^[71-75]。

在实际应用中, 可将 $^{15}\text{N}/^{13}\text{C}/^1\text{H}$ 三共振实验按照关联类型以两个为一组的形式分成 4 组三共振 NMR 谱图: a. HNCA (残基内谱图) 和 HN(CO)CA (残基间谱图); b. HNCO (残基间谱图) 和 HN(CA)CO (残基内谱图); c. CBCANH (残基内谱图) 和 CBCA(CO)NH (残基间谱图); d. HNCACB (残基内谱图) 和 HN(CO)CACB (残基间谱图)。

2.3.2 蛋白质的二级结构单元解析

通过以下 NMR 技术可测定蛋白质的二级结构单元 (α 螺旋、 β 转角、 β 折叠和无规线团) 以及结构特征参数 (质子间距离、二面角和氢键相互作用): a. NOESY 用于识别具有质子间距离 $\leq 5 \text{ \AA}$ 结构特征的二级结构单元; b. DQF-COSY 用于测定 $\text{C}_\alpha\text{-N-H}$ 三键的耦合常数 ($^3J_{\text{HNH}}$), 进一步确定二级

结构单元的二面角。例如, 右手 α 螺旋 $^3J_{\text{HNH}} = 3.9 \text{ Hz}$ ($\phi = -57^\circ$), 3_{10} 螺旋 $^3J_{\text{HNH}} = 4.2 \text{ Hz}$ ($\phi = -60^\circ$), 反平行 β 折叠 $^3J_{\text{HNH}} = 8.9 \text{ Hz}$ ($\phi = -139^\circ$), 平行 β 折叠 $^3J_{\text{HNH}} = 9.7 \text{ Hz}$ ($\phi = -119^\circ$)^[76]; c. 具有慢交换速率的质子易形成氢键, 通过 NOESY 测定蛋白质样品在 H_2O 和 D_2O 中的质子交换速率来研究氢键相互作用^[77]。

基于氨基酸 ^{15}N 、 ^{13}C 和 ^1H 化学位移数据库的化学位移二级结构预测法也常用于解析蛋白质的二级结构单元。然而, 由于蛋白质的化学位移容易受到原子核局部磁环境影响, 导致其化学位移包含多种信息 (二级结构单元、侧链构象、氢键、动力学、溶剂化作用等)^[78]。为了更加准确地获得蛋白质二级结构的单一信息, 建立了化学位移标志识别法 (chemical shift index, CSI), 即根据其数据库建立各种氨基酸自旋系统中 ^{15}N 、 ^{13}C 和 ^1H 的化学位移标志参考值, 用于解析蛋白质的二级结构单元^[79-80]。相关软件主要有 NMRPipe^[81], 在线分析服务器主要为 TALOS+ Web Server (<https://spin.niddk.nih.gov/bax/nmrserver/talos/>) 和 CSI 2.0 Web Server (<http://csi.wishartlab.com/>)^[65, 82]。

2.4 蛋白质的三维结构解析

2.4.1 核磁共振波谱法解析蛋白质三维结构

在蛋白质氨基酸序列和二级结构单元的解析基础上, 通过 NMR 实验进一步测定结构约束参数 (包括原子间空间距离、化学键空间取向等), 结构计算软件计算蛋白质三维结构并评估其可靠性和准确性。NOE 和 PRE 实验可获得原子空间距离约束参数, 化学键空间取向则通过 RDC 实验来获取 (图 2)。

a. NOE

NOE 信号强度与两个偶极相互作用的质子之间距离 ($r_{\text{H-H}}$, 1.8~5 $\text{\AA}$) 的六次方成反比^[83]。在蛋白质三维结构研究中, 主要根据 NOESY 谱图的 NOE 信号强度获得短程距离的质子间距离约束参数 (图 2a)。蛋白质三维结构解析的准确性依赖于 NOE 信号的获取数量和正确归属。然而, 伴随着蛋白质分子质量的增加, NOESY 谱易出现信号峰简并和重叠严重的问题。近期出现的异核多维 NOESY 技术, 即根据与 ^1H 键合的异核 (^{13}C 或 ^{15}N) 化学位移来分辨 NOE 交叉信号峰, 能够通过增加维度的方式来减小信号峰简并和重叠的问题。对于分子质量较小的蛋白质 (< 30 ku), 常用 2D $^1\text{H}\text{-}^1\text{H}$ NOESY、3D $^{15}\text{N}/^{13}\text{C}$ -edited-NOESY、3D $^{13}\text{C}(^{15}\text{N})$

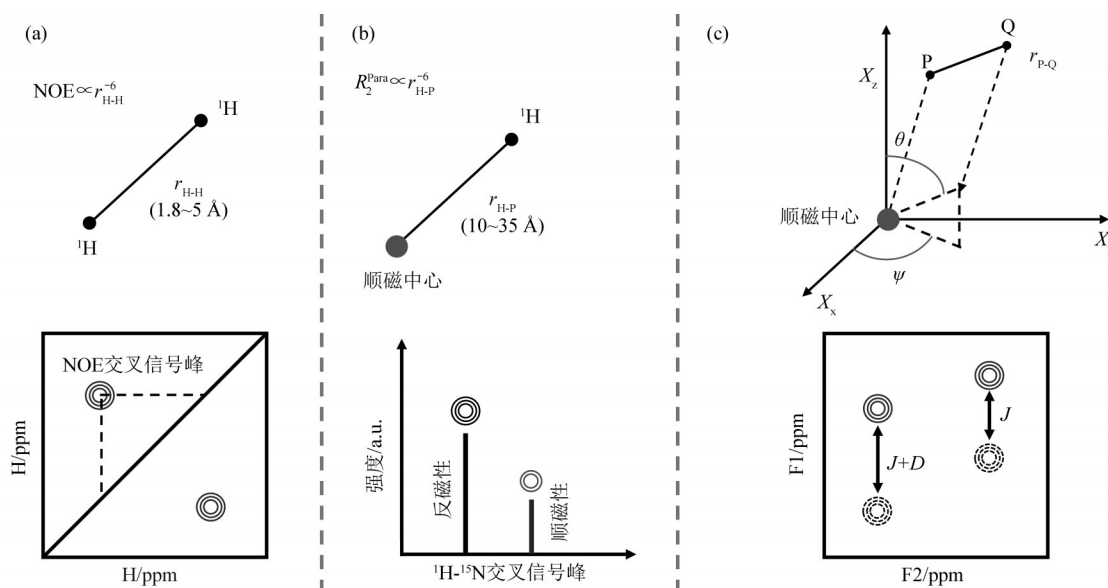


Fig. 2 NMR experimental methods for acquiring protein structural constraint parameters

图2 测定蛋白质结构约束参数的NMR实验方法

(a) NOE; (b) PRE; (c) RDC。

^1H -NOESY-HSQC^[84-86]。对于分子质量较大的蛋白质，则采用3D ^{15}N - ^1H - ^{15}N HSQC-NOESY-HSQC、3D ^{13}C - ^1H - ^{13}C HMQC-NOESY-HMQC、3D ^{15}N - ^1H TROSY-NOESY、4D $^{15}\text{N}/^{15}\text{N}$ HSQC-NOESY-HSQC、4D $^{13}\text{C}/^{15}\text{N}$ HMQC-NOESY-HSQC 和 4D $^{13}\text{C}/^{13}\text{C}$ HMQC-NOESY-HSQC^[87-90]。在全氘代蛋白质的结构研究中，由于未能获得质子的NOE信号，所以，通常将全氘代蛋白质进行选择性质子化，例如酰胺基团和甲基的质子化^[91-92]。

b. PRE

PRE技术用于测定蛋白质结构中远程空间距离的结构约束参数，原理如图2b所示。未成对单电子的顺磁性对其周围的NMR自旋原子核产生顺磁效应，进一步产生原子间的偶极-偶极相互作用，加快一定空间范围内（10~35 Å）的原子核弛豫速率（即PRE速率， $R_{\text{Para}} 2$ ）^[93]。 $R_{\text{Para}} 2$ 与原子核到顺磁中心距离（ $r_{\text{H-P}}$ ）的六次方成反比（公式（3）），主要通过 ^1H , ^{15}N -TROSY或 ^1H , ^{15}N -HSQC等技术进行测定^[94]。

$$R_{\text{Para}}^2 = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_i^2 g_e^2 \mu_B^2 S(S+1)}{r_{\text{H-P}}^6} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right) \quad (3)$$

式中， $r_{\text{H-P}}$ 指 ^1H 原子与未配对单电子（顺磁中心）P之间的距离（Å）， g_e 是电子的 g 因子， μ_B 是玻尔磁子， S 是总电子自旋量子数， μ_0 是真空磁导率， ω_H 是 ^1H 原子的拉摩尔频率（Hz）， τ_c 是分子相关时间（ns）， $\tau_c = (\tau_r^{-1} + \tau_s^{-1})^{-1}$ ，其中 τ_r 是蛋白质分子的运动相关时间， τ_s 是电子弛豫时间。

通常采用以下两种方法制备顺磁性的蛋白质测试体系：a. 化学修饰法，常用的方法为半胱氨酸巯基衍生化法，即将蛋白质分子链中半胱氨酸的巯基与顺磁性氮氧化物自由基标签通过化学键合，得到含有顺磁性标签的衍生化蛋白质，顺磁性氮氧化物自由基标签主要有1-氧自由基-2,2,5,5-四甲基-D-吡咯啉-3-甲基-甲硫基磺酸盐（MTSL）、2,2,6,6-四甲

基哌啶-1-氧自由基（TEMPO）等^[95-96]，顺磁性离子螯合剂（乙二胺四乙酸（EDTA）、二乙基三胺五乙酸（DTPA）、1,4,7,10-四氮杂环十二烷-1,4,7,10-四羧酸（DOTA）等^[96-98]，其中MTSL标签最为常用；b. 融合蛋白表达法，即在蛋白质分子中通过基因表达引入可螯合顺磁性金属离子的肽链^[99]。然而，这些方法均在蛋白质分子链中引入了顺磁性物质，需要考虑它们对蛋白质自身结构的影响。

c. RDC

在较高分子质量蛋白质的结构解析中，常常存在结构约束数据获取较少或较难获得的问题，导致缺乏足够的约束信息用于精确计算蛋白质的三维结构，因而，需要采用RDC技术获取蛋白质分子中

化学键空间取向的约束参数, 得到蛋白质中不同结构域之间相对取向的约束信息^[100]。RDC是指原子核偶极矩间的偶极-偶极相互作用产生偶极-偶极弛豫效应, 导致NMR信号裂分(偶极耦合裂分)的现象。如图2c所示, 对于蛋白质分子中存在偶极耦合相互作用的原子核P和Q, 它们受到顺磁性介

$$D = -\frac{1}{4\pi} \frac{B_0^2}{15k_B T} \frac{\gamma_P \gamma_Q h}{4\pi^2 r_{P-Q}^3} [\Delta\chi_{ax}(3\cos^2\theta - 1) + 1.5\Delta\chi_{rh}\sin^2\theta\cos 2\varphi] \quad (4)$$

式中, D 是残余偶极耦合常数(Hz), B_0 是磁感应强度(T), k_B 是玻尔兹曼常数, T 是测试温度(K), h 是普朗克常数, r_{P-Q} 为存在偶极耦合相互作用的原子核P和Q之间的距离(Å), γ_P 和 γ_Q 分别是发生偶极耦合相互作用原子核P和Q的旋磁比, $\Delta\chi_{ax}$ 和 $\Delta\chi_{rh}$ 分别是磁感应张量的轴向分量和正交分量; θ 和 φ 分别是键矢量和排列张量坐标轴的取向。

蛋白质分子在溶液中处于各向同性的快速翻滚运动状态, 这种各向同性运动带来的平均效应使得其偶极耦合裂分的净值为零^[102]。所以, 需要采取一些方法限制蛋白质的快速翻滚运动, 促使蛋白质产生一定的各向异性, 使得偶极耦合裂分的净值不为零。目前文献报道大致有以下两种方法: a. 一些天然蛋白质含有各向异性磁化率的顺磁性金属离子(如 Fe^{3+} 等), 其固有磁化率可以使蛋白质产生弱取向, 进而产生RDC信号^[103]; b. 利用外部取向介质诱导蛋白质分子在溶液中定向排列, 出现的各向异性可产生10倍以上的RDC信号^[104]。常用的取向介质包括顺磁性金属(如 Fe^{3+} 和钬系金属)、烷基-聚乙二醇分子液晶、DNA纳米管液晶、取向聚丙烯酰胺凝胶以及取向胶原蛋白凝胶等^[105-110]。如果在蛋白质测试体系中加入两种以上定取向介质, 促使蛋白质分子沿不同方向定向排列, 则可获得多个 D 值, 计算出更加准确的蛋白质三维结构^[111]。需要注意的是, 由于外部取向介质法的诱导机制归功于蛋白质与介质的空间和电荷相互作用, 因此, 必须考虑取向介质对蛋白质自身的结构影响^[112]。

2.4.2 其他生物物理技术辅助解析蛋白质三维结构

近期研究表明, 一些生物物理技术有助于NMR解析蛋白质的三维结构^[113], 例如荧光共振能量转移法(Förster/fluorescence resonance energy transfer, FRET)、化学交联结合质谱技术(chemical cross-linking coupled with mass spectrometry, CXMS)、小角X射线散射技术法(small angle X-ray scattering, SAXS)和冷冻电镜法(Cryo-electron microscopy, Cryo-EM)。

质的取向作用前的耦合常数为 J , 受到顺磁性介质作用产生取向后的耦合常数则为 J 和残余偶极耦合常数(D)之和。 D 与原子核P和Q之间的键矢量(θ)以及排列张量坐标轴取向(φ)的关系列于公式(4)^[101], 根据计算软件进一步计算蛋白质二级结构域的空间取向信息。

a. FRET

与PRE技术获得距离约束参数的原理类似, FRET法也是基于偶极-偶极相互作用原理, 即如果蛋白质分子内或分子间的两个荧光生色团相距一定距离(约20~80 Å), 并且荧光供体基团的荧光发射波长与荧光受体基团的吸收波长有一定的重叠, 那么, 荧光供体基团将把能量转移给荧光受体基团, 使受体基团被激发, 产生红移的次级荧光^[113-114]。上述两个波长重叠的程度越大, FRET能量转移的效率越高。因此, 根据FRET能量转移效率, 可计算它们之间的距离(公式(5))^[113]。

$$\langle R_{DA} \rangle = R_0 (\langle E^{-1} \rangle - 1)^{1/6} \quad (5)$$

式中, R_{DA} 为供体和受体荧光基团之间的距离(Å), E 为FRET能量转移效率, R_0 为Förster距离(即某一给定供体-受体对的临界转移距离)。常用的荧光供体单体有丹磺酰氯、AlexaFluor类荧光分子、 Tb^{3+} 荧光离子和四甲基罗丹明(TMR), 荧光受体单体为异硫氰酸荧光素(FITC)、Cy3和Cy5荧光分子, 丹磺酰-FITC、AlexaFluor488-Cy5和 Tb^{3+} -Cy3在FRET能量转移效率达到50%时, R_0 分别为38、52和62 Å^[113]。

在解析蛋白质三维结构方面, 单分子FRET(single-molecule FRET, smFRET)根据得到的距离约束参数并结合分子动力学模拟算法, 可用于研究运动时间尺度在毫秒范围的蛋白质构象动态变化^[113-114]。例如, Ha等^[115]将荧光供体(TMR)和和荧光受体(Cy5)分别修饰到葡萄球菌核酸酶(staphylococcal nuclease, SNase)的两个特定位点, 通过smFRET测定单个SNase分子中TMR和Cy5基团的FRET能量转移效率和它们之间的距离, 发现SNase分子之间存在蛋白骨架结构或侧链在毫秒级时间范围内的波动。膜蛋白具有丰富的生命功能, 与许多疾病有关, 已成为主要的药物靶点, 近期研究报道了采用smFRET研究膜蛋白折叠结构动力学, 例如研究自插入螺旋膜蛋白和细菌毒素的组装和折叠动力学以及突变引起的膜蛋白错误折叠

等, 这些工作将有利于更加了解膜蛋白的折叠结构及机制, 以便设计更加高效的药物^[116]。

b. CXMS

CXMS 法是一种获得松散距离约束 (sparse distance restraints) 的技术。该方法采用特定长度的化学交联剂 (长度约 10~30 Å), 将单个蛋白质或不同蛋白质之间的官能团通过共价键连接起来, 再对交联蛋白质样品进行酶切, 采用质谱法分析小分子产物的化学结构, 利用专业的交联数据处理软件得出交联位点的氨基酸残基及其距离约束参数, 结合 NMR 解析蛋白质的三维结构^[113, 117]。CXMS 技术测试的蛋白质样品不受分子质量的限制, 所以, 对于 NMR 难以解析的高分子质量蛋白质的三维结构, CXMS 技术显示出明显的优势。Gong 等^[117] 最近报道一种改进的 CXMS 技术可将距离约束参数的测量值降至约 6 Å 或更小, 能够更精确地解析蛋白质的结构和动力学。Peng 等^[118] 采用双 (磺基琥珀酰亚胺) 硫酸盐作为交联剂与蛋白质二硫键异构酶 (protein disulfide-isomerase, PDI) 的伯氨或仲氨反应产生稳定的酰胺或亚酰胺键, 经过胰蛋白酶水解以及质谱分析, 采用 Crosswork 软件解析得到 PDI 的二聚体结构。

c. SAXS

SAXS 法通过 X 射线散射实验 (X 射线波长为 0.05~0.5 nm、散射角为 0.1~5°) 获得蛋白质溶液散射强度与角度之间关系^[119], 计算出蛋白质的刚性尺寸、分子质量、聚集体状态、整体形状等信息^[120-121]。SAXS 能提供蛋白质的全局距离约束条件^[122-123], 作为 NMR 的补充约束数据, 所以, SAXS 与 NMR 相结合可提高蛋白质三维结构解析的准确性^[124-125], 常使用的软件主要有 XPLOR-NIH 和 CNS 等^[122-124, 126]。Grishaev 等^[127] 将苹果酸合酶 G (malate synthase G, 81.4 ku) 的 RDC、NOE 和 *J* 耦合常数等 NMR 约束参数与 SAXS 测定的全局距离信息结合, 分析得出更加精准的三维结构。Schiavina 等^[128] 通过 NMR 实验得到了尼帕病毒磷蛋白 N 末端结构域 (Nipah virus phosphoprotein N-terminal region, NiV-PNT, 406 个氨基酸) 的初始结构, 将集合优化方法 (ensemble optimization method, EOM) 与 SAXS 散射数据进行匹配分析, 优化出高分辨率的 NiV-PNT 三维结构。

d. Cryo-EM

Cryo-EM 法通过对蛋白质低温冷冻成玻璃态的

样品进行电镜成像, 经过傅里叶变化及重建算法构建出初始蛋白质三维电子密度图 (Cryo-map), 得出高分辨的蛋白质三维结构图^[129-130]。与 SAXS 相类似, Cryo-EM 主要提供全局距离约束条件, 但分辨率更高。Cryo-EM 技术不受蛋白质分子质量的限制, 与 NMR 联合使用可得到高分辨率的蛋白质三维结构^[131-134]。Gauto 等^[135] 将 Cryo-EM 与 NMR 联用, 解析得到较高准确度和精确性 (1 Å 以下) 的十二聚集体 TET2 氨肽酶 (468 ku) 的三维结构。Bardiaux 等^[136] 通过 NOE 和 RDC 等 NMR 结构约束技术计算得到纤维状菌毛 (Pilus) 分子的 PpdD 亚基结构, 结合 Cryo-EM 电子密度图 (分辨率为 8 Å) 构建出纤维状菌毛的三维结构。

2.5 蛋白质分子激发态结构的解析

蛋白质在生命活动中总是处于一种不断运动的状态, 通过构象变化执行着各种生物功能^[137-138], 包括主链运动、侧链芳环翻转、结构域相对运动、寡聚体结构域交换等, 覆盖了时间尺度从皮秒到秒的变化^[4, 104]。这种构象交换过程中, 不同构象状态之间具有不同的含量和寿命, 各状态的含量取决于它们的相对自由能, 而寿命则很大程度上取决于不同状态之间转化的自由能垒大小^[139]。其中, 含量较大的称为主要状态或基态 (major state or ground state), 而含量较低 (约 5%) 的瞬时状态被称为次要状态或激发态 (minor state or excited state), 由于其含量较低且寿命较短, 只能凭借它与基态之间的化学交换才能被检测, 所以也被称为不可见状态 (invisible state)^[140-141]。基态与激发态结构之间的构象交换过程可发生在较宽的时间尺度下 (图 3)^[104, 142], 主要 NMR 技术包括 PRE、弛豫扩散^[143]、CPMG RD (cPMG 弛豫扩散, Carr-Purcell-Meiboom-Gill relaxation dispersion)、CEST (化学交换饱和和转移, chemical exchange saturation transfer)、DEST (暗态交换饱和和转移, dark state excitation saturation transfer)^[144]、H/D 交换^[145] 和 ZZ 交换^[146] 等。其中与蛋白质功能联系最为密切的构象交换时间尺度是微秒至毫秒, 例如配体的结合与释放、蛋白质折叠、变构效应和酶催化等^[104, 147]。近些年来发展起来的 CPMG RD 和 CEST 能更好地研究尺度为微秒至毫秒的蛋白质激发态结构 (图 3a, b)。

2.5.1 CPMG RD

CPMG RD 基于自旋回波 (spin-echo) 原理, 以不同 CPMG 频率施加 π 脉冲, 根据发生的化学交

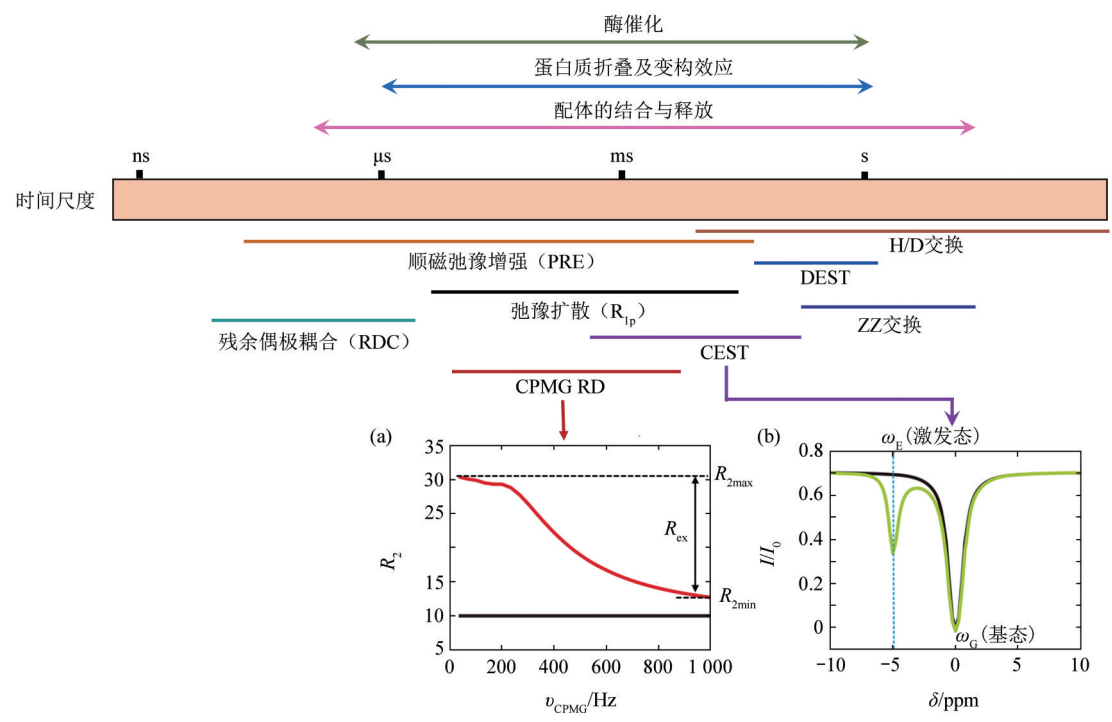


Fig. 3 Timescales of protein motion and main NMR techniques in the study of excited state structure [1, 4, 113]

图3 蛋白质的运动时间尺度以及激发态结构研究中的主要NMR技术 [1, 4, 113]

(a) CPMG RD曲线 (R_{2max} : 存在化学交换的横向弛豫速率; R_{2min} : 无化学交换的横向弛豫速率); (b) CEST曲线 (I_0 : 未施加射频场或无化学交换时某一氨基酸残基的原子核共振信号强度; I : 施加强射频场, 存在基态和激发态化学交换的某一氨基酸残基的原子核共振信号强度; 绿色曲线: 存在化学交换; 黑色曲线: 无化学交换)。

换测定不同氨基酸残基的弛豫速率和化学位移, 分析 0.3~10 ms 时间尺度下的蛋白质激发态结构 (含量大于 0.5%) [147-149]。例如采用 ^1H - ^{15}N HSQC、Methyl-HSQC 或 TROSY 等测试技术, CPMG RD 可测定出蛋白质主链 $^1\text{H}^N$ 、 ^{15}N 、 $^{13}\text{C}^\alpha$ 和 ^{13}CO 以及侧链 $^{13}\text{C}^\beta$ 和甲基的横向弛豫速率 (R_2) 随 CPMG 频率 (ν_{CPMG}) 变化的关系曲线 (图 3a)。交换速率 R_{ex} (R_{2max} 与 R_{2min} 之差) 不仅能够提供各氨基酸残基发生化学交换的情况及它们在激发态的相对含量, 而且可以获得激发态氨基酸残基中不同类型原子核的化学位移, 用于计算较高准确性的激发态蛋白质三维结构 [104, 139, 148-152]。

Bouvignies 等 [153] 采用 CPMG RD 技术分别获得了溶酶体激发态结构中主链 ^1H 、 ^{13}C 和 ^{15}N 以及侧链甲基的化学位移, 结合 CS-Rosetta 分子模型构建技术获得了原子分辨率下溶菌酶突变体的激发态结构, 阐明了不同突变体位点对基态与激发态溶菌酶结构的影响。Neudecker 等 [154] 通过 CPMG RD 技术测定了 Fyn SH3 结构域在激发态的 $^1\text{H}^N$ 、 ^{15}N 、

$^{13}\text{C}_\alpha$ 、 $^{13}\text{C}^\alpha$ 和 $^1\text{H}^\alpha$ 化学位移值, 根据 RDC 获得的 $^1\text{H}^N$ - ^{15}N 化学键取向信息等结构约束参数, 结合 Xplor-NIH 等软件进行计算, 发现激发态 Fyn SH3 结构域的 C 端为无序状态, 暴露出易聚集的末端 β 折叠链, 这可能是导致淀粉样原纤维形成的原因, 以至于诱发神经退行性疾病。

2.5.2 CEST

CEST 用于研究构象交换发生在时间尺度为 2~100 ms、寿命为 5~50 ms 的蛋白质激发态结构 [104, 145]。为了检测蛋白质在激发态的构象, CEST 将低丰度的激发态构象信息通过化学交换饱和和转移的方式传递给高丰度的基态构象并加以放大。如图 3b 所示, 如果蛋白质存在激发态和基态两个相互交换的构象状态, 那么, 在 CEST 分析结果中将出现一对向下的信号峰 (绿色曲线), 较大的信号峰对应于基态, 较小的信号峰则对应于激发态 [1]。应用 ChemEx 软件包 (<https://github.com/gbouvignies/chemex>) 对信号峰强度进行拟合计算, 可以获得化学交换速率常数、激发态相对含量以及

激发态与基态化学位移的差值等信息^[155-156]。通过¹H-¹⁵N HSQC或TROSY实验,CEST可检测到的激发态蛋白质主链¹H、¹⁵N和¹³C以及侧链甲基的化学位移信息^[157-161]。

CEST与CPMG RD不同之处主要在于:

a. CEST更适合研究构象交换时间较慢的蛋白质激发态结构;b. CEST中基态与激发态的化学位移差异更显著且更准确^[155, 162]。将CEST与RDC、PRE以及分子对接技术等相结合,可获得原子分辨率下的激发态三维结构^[162]。例如Sekhar等^[163]将CEST与CPMG RD、分子对接技术相结合,研究发现超氧化物歧化酶的双聚体界面和静电环区域是导致其聚集的关键区域,有望成为治疗肌萎缩性脊髓侧索硬化(amyotrophic lateral sclerosis)的药物靶点。

3 核磁共振波谱法解析高分子质量单链蛋白及超大蛋白质复合体的三维结构

3.1 高分子质量单链蛋白的三维结构解析

当分子质量高于10 ku时,蛋白质在溶液中的运动速度随着分子质量升高而减缓,导致其NMR峰的谱线加宽且共振信号简并和重叠,增加了三维结构解析的难度。对于分子质量约为30~80 ku的单链蛋白质,常采用TROSY联用的3D(或4D)三共振技术研究蛋白质的三维结构^[164]。例如Tugarinov等报道^[91],TROSY联用4D-HNCACO、HNCOCA和HNCO_iCA_i三共振NMR,以及TROSY联用3D-HNCO、HN(CA)CO、HNCACB和HN(CO)CACB三共振NMR,结合3D-HN(CO)CACB和3D-HNCACB、2D-¹⁵N-¹H-TROSY-HSQC和4D-¹⁵N,¹⁵N-NOESY,可以指认苹果酸合酶G 95%以上的共振信号(主链¹H^N、¹⁵N、¹³C^α、¹³C^β和侧链¹³C^β),得出了主链的氨基酸序列结构。Bertelsen等^[165]报道采用类似的NMR技术,解析了DnaK蛋白(bacterial heat-shock protein 70, 70 ku)主链的氨基酸序列结构,结合RDC研究了DnaK蛋白各结构域之间的相对运动和结构取向。

3.2 超大蛋白质复合体的三维结构解析

生命体中存在以聚集体方式形成的超大蛋白质复合体,它们发挥着重要的生物功能,所以,NMR研究超大蛋白质复合体的三维结构对解释其生物功能具有重要意义。然而,超大蛋白质复合体过高的分子质量导致严重的NMR信号重叠和简并。为了突破分子质量的限制,目前在NMR技术

和同位素标记方面已有突破性的进展。在NMR技术方面,最为突出的是Fiaux等^[166]将¹H-¹⁵N CRIPT-TROSY应用于研究超大蛋白质复合体(900 ku),通过比较十四聚体细菌分子伴侣GroEL(bacterial chaperonin 60, 800 ku)与七聚体辅分子伴侣GroES(bacterial cochaperonin 60, 72 ku)复合前后NMR谱图的差异,分析了复合体的结构及其形成机理。

含有甲基和芳香基团的氨基酸残基占蛋白质氨基酸残基总数的45%~55%,通常位于蛋白质分子的疏水结构域及与配体相互作用的界面中^[167]。这使得选择性甲基质子化标记法被成功应用到超大动态蛋白质复合体的结构研究中。该方法首先对蛋白质进行全氘代标记,然后选择性地对脂肪族氨基酸残基(Ala、Ile、Leu、Met、Thr和Val)的侧链甲基进行质子化。为了获得更多的分子间NOE距离约束参数用于解析超大蛋白质复合体的三维结构,也会对芳香族氨基酸残基(Phe、Tyr和Trp)进行选择性质子化。

氨基酸残基侧链甲基的动力学信息可提供毫秒尺度下的蛋白质构象变化信息,有助于更全面地认识蛋白质结构与功能的关系^[168]。Huang等^[169]对SecB分子伴侣(70 ku)的脂肪族氨基酸残基的甲基以及芳香族氨基酸残基进行了选择性质子化标记,以麦芽糖结合蛋白(MBP, 396个氨基酸)和碱性磷酸酶(PhoA, 471个氨基酸)为SecB分子伴侣的结合底物,通过三共振NMR和3D ¹³C/¹⁵N-edited NOESY对上述质子化标记基团的共振信号进行归属,分析了SecB/MBP与SecB/PhoA复合物中分子间NOE信号等参数,发现在SecB分子伴侣长而连续的疏水空腔中存在5个与PhoA结合的位点和7个与MBP结合的位点,揭示了SecB分子伴侣识别MBP和PhoA底物的机理及其抗折叠的生物活性功能。

Saio等^[170]对全氘代的触发因子(trigger factor, TF)进行甲基选择性标记及芳香族残基选择性质子化,经过NMR实验获得了丰富的化学位移数据和分子间NOE距离信息,解析出TF/PhoA复合物(200 ku)的三维结构。这些选择性质子化标记法还可用于解析其他超大蛋白质复合体的三维结构,例如DnaJ/PhoA复合物(111.5 ku)^[85]、20S蛋白酶体核心颗粒(20S CP, 1 000 ku)^[171]和SecA分子伴侣(204 ku)^[172]等。

4 核磁共振波谱法结合分子建模算法研究蛋白质的三维结构

4.1 蛋白质初始结构模型的构建

4.1.1 基于数据库模板的建模法

基于数据库模板的建模法通过对比目标蛋白质和数据库中已知结构蛋白质的氨基酸序列或二级结构(称为模板),经过计算模拟得到目标蛋白初始三维结构。检索氨基酸序列的方法主要有HHBlits和LOMETS等^[173-174],计算模拟法包括I-TASSER和SWISS-MODEL等^[175-176]。该建模法主要有同源建模法和折叠识别法。同源建模法将目标蛋白质与模板的氨基酸序列进行比对,筛选出与目标蛋白质在进化上同源的模板,获得最优化模板的三维结构氨基酸的坐标或从模板结构中收集结构约束参数,结合计算模拟法构建目标蛋白质的初始结构模型,分为模板识别、模板选择、模型构建和模型质量评估4个步骤。但是,由于来自不同进化起源的蛋白质可能具有相似的结构,在同源建模法应用中可能会产生错误的结构信息。针对该问题可采用折叠识别法,即利用穿线法(threading)来识别数据库中与目标蛋白质具有相似氨基酸结构或序列的模板,根据它们的相似程度进行排序,优化出目标蛋白质的初始结构模型^[175]。

4.1.2 基于无数据库模板的建模法

当在数据库中找不到与目标蛋白质结构匹配的模板时,一般采用从头建模法来构建目标蛋白质的初始结构模型,即基于无数据库模板的建模法。该方法将目标蛋白质切割为小片段序列,通过在数据库中搜索相似结构的片段获取目标蛋白质片段的结构参数信息,修订至目标蛋白质的序列结构中,采用计算模拟的方法组合这些片段结构,得到目标蛋白质的初始结构模型,代表性的算法有Rosetta(CS-Rosetta)^[177]。然而,随着蛋白质分子质量或氨基酸数目的增加,需要检索的构象相空间(phase space)数量也急剧增加,这使得较大蛋白质的从头计算建模变得极其困难。近些年来随着从头建模法不断的优化和发展,分子质量的限制影响开始变得越来越小,由以前只能建立小于150个氨基酸的蛋白质初始结构模型^[178]。到现在最高可构建含有770个氨基酸的蛋白质初始结构模型^[179]。

4.2 蛋白质结构优化

将建模法建立的蛋白质初始结构与实验获得的结构约束参数相结合,经过分子动力学模拟和能量

最小化处理,可得到目标蛋白质的三维结构。结构约束参数主要来源于NMR和小角X射线散射实验,包括NOESY测定的质子间距、PRE获得的长程距离、RDC获得的空间取向、骨架自旋系统的化学位移提供二面角以及酰胺氢/氘交换速率提供相互作用界面等信息^[3, 126, 180-183]。这些结构约束参数还可作为初步优化的目标蛋白质的结构评估指标,经过多次迭代优化,直至得到一个(或一组)与结构约束参数相一致的目标蛋白质的三维结构^[184]。

5 结论与展望

蛋白质的三维结构与其生物功能密切相关,研究蛋白质的三维结构有助于揭示其生物功能机制。NMR是一种重要的解析蛋白质三维结构的方法,通常采用蛋白质同位素(¹⁵N/¹³C/²H)标记法,以获得出现较高强度核磁信号峰的高分辨率NMR谱图。将¹⁵N/¹³C/²H三共振技术与NMR计算软件相结合可推算蛋白质主链中氨基酸的序列及其二面角结构参数,得出蛋白质的二级结构。NOE、PRE及RDC技术可获得蛋白质的三维结构约束参数(原子间空间距离和化学键空间取向),经过NMR软件进行结构计算以及结构评估,解析得出蛋白质的三维结构。NMR与其他生物物理手段(如FRET、CXMS、SAXS和Cryo-EM)相结合,有助于获得更加精准的蛋白质三维结构。NMR与分子建模算法相结合,能更加快速、低成本地解析蛋白质的三维结构。将NMR用于研究蛋白质分子激发态结构以及对超大蛋白质复合物三维结构的解析,对于阐明蛋白质在生命活动中发挥的生物功能有着重要的研究价值和意义。

尽管目前已发展了多种NMR技术解析蛋白质的三维结构,但是,在更加精准解析蛋白质三维结构方面仍然存在需要进一步研究的问题。例如,NMR在研究高分子质量蛋白质及多聚体蛋白质的三维结构仍然存在核磁信号峰重叠严重、谱线较宽、分辨率较低等问题。这些问题的研究将能更加提升NMR在解析蛋白质三维结构的应用价值,揭示更多蛋白质的生物功能机制。

参 考 文 献

- [1] Sekhar A, Kay L E. An NMR view of protein dynamics in health and disease. *Annu Rev Biophys*, 2019, **48**: 297-319
- [2] Korolkova Y, Maleeva E, Mikov A, *et al.* New insectotoxin from

- tibellus oblongus spider venom presents novel adaptation of ICK fold. *Toxins* (Basel), 2021, **13**(1): 29
- [3] Nerli S, Mcshan A C, Sgourakis N G. Chemical shift-based methods in NMR structure determination. *Prog Nucl Magn Reson Spectrosc*, 2018, **106-107**: 1-25
- [4] Lisi G P, Loria J P. Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function. *Prog Nucl Magn Reson Spectrosc*, 2016, **92-93**: 1-17
- [5] Frueh D P. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc*, 2014, **78**: 47-75
- [6] Sugase K, Dyson H J, Wright P E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 2007, **447**(7147): 1021-1025
- [7] Axe J M, Boehr D D. Long-range interactions in the alpha subunit of tryptophan synthase help to coordinate ligand binding, catalysis, and substrate channeling. *J Mol Biol*, 2013, **425**(9): 1527-1545
- [8] Kay L E, Ikura M, Tschudin R, *et al.* Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson*, 2011, **213**(2): 423-441
- [9] Farmer B T, Venters R A, Spicer L D, *et al.* A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J Biomol NMR*, 1992, **2**(2): 195-202
- [10] Grzesiek S, Bax A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson*, 1992, **96**(2): 432-440
- [11] Bax A, Ikura M. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α -carbon of the preceding residue in uniformly $^{15}\text{N}/^{13}\text{C}$ enriched proteins. *J Biomol NMR*, 1991, **1**(1): 99-104
- [12] Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc*, 1999, **34**(2): 93-158
- [13] Muhandiram D R, Kay L E. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J Magn Reson*, 1994, **103**(3): 203-216
- [14] Clubb R T, Thanabal V, Wagner G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^1\text{H}^{\text{N}}$, ^{15}N , and $^{13}\text{C}'$ chemical shifts in ^{15}N - ^{13}C -labelled proteins. *J Magn Reson*, 1992, **97**(1): 213-217
- [15] Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*, 1992, **114**(16): 6291-6293
- [16] Grzesiek S, Bax A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Reson*, 1992, **99**(1): 201-207
- [17] Wittekind M, Mueller L. HNCACB, a high-Sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α - and β -carbon resonances in proteins. *J Magn Reson*, 1993, **101**(2): 201-205
- [18] Yamazaki T, Lee W, Arrowsmith C H, *et al.* A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J Am Chem Soc*, 1994, **116**(26): 11655-11666
- [19] Yang D, Kay L E. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J Am Chem Soc*, 1999, **121**(11): 2571-2575
- [20] Konrat R, Yang D, Kay L E. A 4D TROSY-based pulse scheme for correlating $^1\text{H}^{\text{N}}$, ^{15}N , ^{13}Ca i, $^{13}\text{C}'$ i-1 chemical shifts in high molecular weight, ^{15}N , ^{13}C , ^2H labeled proteins. *J Biomol NMR*, 1999, **15**(4): 309-313
- [21] Pritchard R B, Hansen D F. Characterising side chains in large proteins by protonless ^{13}C -detected NMR spectroscopy. *Nat Commun*, 2019, **10**(1): 1747
- [22] Kerfah R, Hamelin O, Boissbouvier J, *et al.* CH_3 -specific NMR assignment of alanine, isoleucine, leucine and valine methyl groups in high molecular weight proteins using a single sample. *J Biomol NMR*, 2015, **63**(4): 389-402
- [23] Siemons L, Mackenzie H W, Shukla V K, *et al.* Intra-residue methyl-methyl correlations for valine and leucine residues in large proteins from a 3D-HMBC-HMQC experiment. *J Biomol NMR*, 2019, **73**(12): 749-757
- [24] Mishra S H, Frueh D P. Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps. *J Biomol NMR*, 2015, **62**(3): 281-290
- [25] Michel E, Skrisovska L, Wuthrich K, *et al.* Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. *ChemBioChem*, 2013, **14**(4): 457-466
- [26] Nabeshima Y, Mizuguchi M, Kajiyama A, *et al.* Segmental isotope-labeling of the intrinsically disordered protein PQBP1. *FEBS Lett*, 2014, **588**(24): 4583-4589
- [27] Yagi H, Tsujimoto T, Yamazaki T, *et al.* Conformational change of H^+ -ATPase β monomer revealed on segmental isotope labeling NMR spectroscopy. *J Am Chem Soc*, 2004, **126**(50): 16632-16638
- [28] Castaneda C A, Spasser L, Bavikar S N, *et al.* Segmental isotopic labeling of ubiquitin chains to unravel monomer-specific molecular behavior. *Angew Chem Int Ed Engl*, 2011, **50**(47): 11210-11214
- [29] Wang Z A, Cole P A. Methods and applications of expressed protein ligation. *Methods Mol Biol*, 2020, **2133**: 1-13
- [30] Sarmiento C, Camarero J A. Biotechnological applications of protein splicing. *Curr Protein Pept Sci*, 2019, **20**(5): 408-424
- [31] Conibear A C, Watson E E, Payne R J, *et al.* Native chemical ligation in protein synthesis and semi-synthesis. *Chem Soc Rev*, 2018, **47**(24): 9046-9068
- [32] Hayashi K, Kojima C. Efficient protein production method for NMR using soluble protein tags with cold shock expression vector. *J Biomol NMR*, 2010, **48**(3): 147-155
- [33] White C E, Kemp N M, Komives E A. Expression of highly disulfide-bonded proteins in *Pichia pastoris*. *Structure*, 1994, **2**(11): 1003-1005
- [34] Clark L, Dikyi I, Rosenbaum D M, *et al.* On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies. *J*

- Biomol NMR, 2018, **71**(4): 203-211
- [35] Clark L, Zahm J A, Ali R, *et al.* Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. J Biomol NMR, 2015, **62**(3): 239-245
- [36] Franke B, Opitz C, Isogai S, *et al.* Production of isotope-labeled proteins in insect cells for NMR. J Biomol NMR, 2018, **71**(3): 173-184
- [37] Egorova-Zachernyuk T A, Bosman G J, Degrip W J. Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. Appl Microbiol Biotechnol, 2011, **89**(2): 397-406
- [38] Hoffmann B, Lohr F, Laguerre A, *et al.* Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. Prog Nucl Magn Reson Spectrosc, 2018, **105**: 1-22
- [39] Sobhanifar S, Reckel S, Junge F, *et al.* Cell-free expression and stable isotope labelling strategies for membrane proteins. J Biomol NMR, 2010, **46**(1): 33-43
- [40] Hyberts S G, Arthanari H, Robson S A, *et al.* Perspectives in magnetic resonance: NMR in the post-FFT era. J Magn Reson, 2014, **241**: 60-73
- [41] Schanda P, Brutscher B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc, 2005, **127**(22): 8014-8015
- [42] Schanda P, Kupce E, Brutscher B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR, 2005, **33**(4): 199-211
- [43] Lee W, Hu K, Tonelli M, *et al.* Fast automated protein NMR data collection and assignment by ADAPT-NMR on Bruker spectrometers. J Magn Reson, 2013, **236**: 83-88
- [44] Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc, 2006, **128**(28): 9042-9043
- [45] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson, 2007, **187**(1): 163-169
- [46] Kim S, Szyperski T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc, 2003, **125**(5): 1385-1393
- [47] Kupce E, Freeman R. Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. J Am Chem Soc, 2004, **126**(20): 6429-6440
- [48] Luan T, Jaravine V, Yee A, *et al.* Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. J Biomol NMR, 2005, **33**(1): 1-14
- [49] Rovnyak D, Hoch J C, Stern A S, *et al.* Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. J Biomol NMR, 2004, **30**(1): 1-10
- [50] Billeter M. Non-uniform sampling in biomolecular NMR. J Biomol NMR, 2017, **68**(2): 65-66
- [51] Marion D. Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. J Biomol NMR, 2005, **32**(2): 141-150
- [52] Hansen D F. Using deep neural networks to reconstruct non-uniformly sampled NMR spectra. J Biomol NMR, 2019, **73**(10-11): 577-585
- [53] Ying J, Delaglio F, Torchia D A, *et al.* Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. J Biomol NMR, 2017, **68**(2): 101-118
- [54] Wurz J M, Kazemi S, Schmidt E, *et al.* NMR-based automated protein structure determination. Arch Biochem Biophys, 2017, **628**: 24-32
- [55] Tikole S, Jaravine V, Rogov V V, *et al.* Fast automated NMR spectroscopy of short-lived biological samples. ChemBioChem, 2012, **13**(7): 964-967
- [56] Guntert P, Buchner L. Combined automated NOE assignment and structure calculation with CYANA. J Biomol NMR, 2015, **62**(4): 453-471
- [57] Lee W, Markley J L. PINE-SPARKY.2 for automated NMR-based protein structure research. Bioinformatics, 2018, **34**(9): 1586-1588
- [58] Hiller S, Fiorito F, Wuthrich K, *et al.* Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA, 2005, **102**(31): 10876-10881
- [59] Dutta S K, Serrano P, Proudfoot A, *et al.* APSY-NMR for protein backbone assignment in high-throughput structural biology. J Biomol NMR, 2015, **61**(1): 47-53
- [60] Gossert A D, Hiller S, Fernandez C. Automated NMR resonance assignment of large proteins for protein-ligand interaction studies. J Am Chem Soc, 2011, **133**(2): 210-213
- [61] Lee W, Kim J H, Westler W M, *et al.* PONDEROSA, an automated 3D-NOESY peak picking program, enables automated protein structure determination. Bioinformatics, 2011, **27**(12): 1727-1728
- [62] Klukowski P, Augoff M, Zieba M, *et al.* NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. Bioinformatics, 2018, **34**(15): 2590-2597
- [63] Lee W, Bahrami A, Dashti H T, *et al.* I-PINE web server: an integrative probabilistic NMR assignment system for proteins. J Biomol NMR, 2019, **73**(5): 213-222
- [64] Pritisanac I, Wurz J M, Alderson T R, *et al.* Automatic structure-based NMR methyl resonance assignment in large proteins. Nat Commun, 2019, **10**(1): 4922
- [65] Shen Y, Delaglio F, Cornilescu G, *et al.* TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR, 2009, **44**(4): 213-223
- [66] Brunger A T. Version 1.2 of the crystallography and NMR system. Nat Protoc, 2007, **2**(11): 2728-2733
- [67] Maier J A, Martinez C, Kasavajhala K, *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput, 2015, **11**(8): 3696-3713
- [68] Rieping W, Habeck M, Bardiaux B, *et al.* ARIA2: automated NOE assignment and data integration in NMR structure calculation.

- Bioinformatics, 2007, **23**(3): 381-382
- [69] Laskowski RA, Rullmannn JA, Macarthur MW, *et al.* AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, 1996, **8**(4): 477-486
- [70] Wuthrich K. NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew Chem Int Ed Engl*, 2003, **42**(29): 3340-3363
- [71] Orton H W, Stanek J, Schubeis T, *et al.* Protein NMR resonance assignment without spectral analysis: 5D solid-state automated projection spectroscopy (SO-APSY). *Angew Chem Int Ed Engl*, 2020, **59**(6): 2380-2384
- [72] Zawadzka-Kazimierczuk A, Kozminski W, Sanderova H, *et al.* High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. *J Biomol NMR*, 2012, **52**(4): 329-337
- [73] Bermel W, Felli I C, Gonnelli L, *et al.* High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR*, 2013, **57**(4): 353-361
- [74] Bermel W, Bertini I, Felli I C, *et al.* Speeding up sequence specific assignment of IDPs. *J Biomol NMR*, 2012, **53**(4): 293-301
- [75] Novacek J, Zawadzka-Kazimierczuk A, Papouskova V, *et al.* 5D ^{13}C -detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR*, 2011, **50**(1): 1-11
- [76] Pardi A, Billeter M, Wüthrich K. Calibration of the angular dependence of the amide proton- C^{α} proton coupling constants, $^3J_{\text{HN}\alpha}$, in a globular protein. *J Mol Biol*, 1984, **180**(3): 741-751
- [77] Raschle T, Rios Flores P, Opitz C, *et al.* Monitoring backbone hydrogen-bond formation in β -barrel membrane protein folding. *Angew Chem Int Ed Engl*, 2016, **55**(20): 5952-5955
- [78] Wishart D S. Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc*, 2011, **58**(1-2): 62-87
- [79] Di Lello P, Rouge L, Pan B, *et al.* ^1H , ^{13}C and ^{15}N backbone resonance assignment for the 40.5 kDa catalytic domain of ubiquitin specific protease 7 (USP7). *Biomol NMR Assign*, 2016, **10**(2): 345-349
- [80] Zhang H, Neal S, Wishart D S. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR*, 2003, **25**(3): 173-195
- [81] Delaglio F, Grzesiek S, Vuister G W, *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 1995, **6**(3): 277-293
- [82] Hafsa N E, Wishart D S. CSI 2.0: a significantly improved version of the Chemical Shift Index. *J Biomol NMR*, 2014, **60**(2-3): 131-146
- [83] Vogeli B. The nuclear Overhauser effect from a quantitative perspective. *Prog Nucl Magn Reson Spectrosc*, 2014, **78**: 1-46
- [84] McMahon K L, Tran H N T, Deuis J R, *et al.* Discovery, pharmacological characterisation and NMR structure of the novel micro-conotoxin SxIIIC, a potent and irreversible NaV channel inhibitor. *Biomedicines*, 2020, **8**(10): 391
- [85] Jiang Y, Rossi P, Kalodimos C G. Structural basis for client recognition and activity of Hsp40 chaperones. *Science*, 2019, **365**(6459): 1313-1319
- [86] Barthe P, Mukamolova G V, Roumestand C, *et al.* The structure of PknB extracellular PASTA domain from *mycobacterium tuberculosis* suggests a ligand-dependent kinase activation. *Structure*, 2010, **18**(5): 606-615
- [87] Lange O F, Rossi P, Sgourakis N G, *et al.* Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci USA*, 2012, **109**(27): 10873-10878
- [88] Saio T, Kawagoe S, Ishimori K, *et al.* Oligomerization of a molecular chaperone modulates its activity. *Elife*, 2018, **7**: 35731
- [89] Sanganna Gari R R, Seelheim P, Liang B, *et al.* Quiet quater membrane protein G (OmpG) nanopore for biosensing. *ACS Sens*, 2019, **4**(5): 1230-1235
- [90] Gaponenko V, Sarma S P, Altieri A S, *et al.* Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints. *J Biomol NMR*, 2004, **28**(3): 205-212
- [91] Tugarinov V, Muhandiram R, Ayed A, *et al.* Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc*, 2002, **124**(34): 10025-10035
- [92] Tugarinov V, Kay L E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc*, 2003, **125**(45): 13868-13878
- [93] Clore G M, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev*, 2009, **109**(9): 4108-4139
- [94] Keizers P H, Ubbink M. Paramagnetic tagging for protein structure and dynamics analysis. *Prog Nucl Magn Reson Spectrosc*, 2011, **58**(1-2): 88-96
- [95] Furuita K, Kataoka S, Sugiki T, *et al.* Utilization of paramagnetic relaxation enhancements for high-resolution NMR structure determination of a soluble loop-rich protein with sparse NOE distance restraints. *J Biomol NMR*, 2015, **61**(1): 55-64
- [96] Lee K Y, Fang Z, Enomoto M, *et al.* Two distinct structures of membrane-associated homodimers of GTP- and GDP-bound KRAS4B revealed by paramagnetic relaxation enhancement. *Angew Chem Int Ed Engl*, 2020, **59**(27): 11037-11045
- [97] Leonov A, Voigt B, Rodriguez-Castañeda F, *et al.* Convenient synthesis of multifunctional EDTA-based chiral metal chelates substituted with anS-mesylcysteine. *Chemistry*, 2005, **11**(11): 3342-3348
- [98] Huang F, Pei Y Y, Zuo H H, *et al.* Bioconjugation of proteins with a paramagnetic NMR and fluorescent tag. *Chemistry (Easton)*, 2013, **19**(50): 17141-17149
- [99] Donaldson L W, Skrynnikov N R, Choy W Y, *et al.* Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. *J Am*

- Chem Soc, 2001, **123**(40): 9843-9847
- [100] Jain N U. Use of residual dipolar couplings in structural analysis of protein-ligand complexes by solution NMR spectroscopy. *Methods Mol Biol*, 2009, **544**: 231-252
- [101] Bertini I, Luchinat C, Piccioli M. Paramagnetic probes in metalloproteins. *Methods in Enzymology*, 2001, **339**: 314-340
- [102] Cavanagh J, Fairbrother W, Palmer A, *et al.* Protein NMR spectroscopy principles and practice. 2nd ed. Netherlands: Elsevier, 2007: 665-673
- [103] Banci L, Bertini I, Huber J G, *et al.* Partial orientation of oxidized and reduced cytochrome b_5 at high magnetic fields: magnetic susceptibility anisotropy contributions and consequences for protein solution structure determination. *J Am Chem Soc*, 1998, **120**(49): 12903-12909
- [104] Chiliveri S C, Deshmukh M V. Recent excitements in protein NMR: large proteins and biologically relevant dynamics. *J Biosci*, 2016, **41**(4): 787-803
- [105] Hus J C, Marion D, Blackledge M. *De novo* determination of protein structure by NMR using orientational and long-range order restraints. *J Mol Biol*, 2000, **298**(5): 927-936
- [106] Gay G L, Lindhout D A, Sykes B D. Using lanthanide ions to align troponin complexes in solution: order of lanthanide occupancy in cardiac troponin C. *Protein Sci*, 2004, **13**(3): 640-651
- [107] Rückert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc*, 2000, **122**(32): 7793-7797
- [108] Douglas S M, Chou J J, Shih W M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc Natl Acad Sci USA*, 2007, **104**(16): 6644-6648
- [109] Tycko R, Blanco F J, Ishii Y. Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J Am Chem Soc*, 2000, **122**(38): 9340-9341
- [110] Ma J, Goldberg G I, Tjandra N. Weak alignment of biomacromolecules in collagen gels: an alternative way to yield residual dipolar couplings for NMR measurements. *J Am Chem Soc*, 2008, **130**(48): 16148-16149
- [111] Sgourakis N G, Lange O F, Dimaio F, *et al.* Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. *J Am Chem Soc*, 2011, **133**(16): 6288-6298
- [112] Newby F N, De Simone A, Yagi-Utsumi M, *et al.* Structure-free validation of residual dipolar coupling and paramagnetic relaxation enhancement measurements of disordered proteins. *Biochemistry*, 2015, **54**(46): 6876-6886
- [113] Tang C, Gong Z. Integrating non-NMR distance restraints to augment NMR depiction of protein structure and dynamics. *J Mol Biol*, 2020, **432**(9): 2913-2929
- [114] Okamoto K, Sako Y. Recent advances in FRET for the study of protein interactions and dynamics. *Curr Opin Struct Biol*, 2017, **46**: 16-23
- [115] Ha T, Ting A Y, Liang J, *et al.* Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc Natl Acad Sci USA*, 1999, **96**(3): 893-898
- [116] Krainer G, Keller S, Schlierf M. Structural dynamics of membrane-protein folding from single-molecule FRET. *Curr Opin Struct Biol*, 2019, **58**: 124-137
- [117] Gong Z, Ye S X, Tang C. Tightening the crosslinking distance restraints for better resolution of protein structure and dynamics. *Structure*, 2020, **28**: 1160-1167
- [118] Peng L, Rasmussen M I, Chailyan A, *et al.* Probing the structure of human protein disulfide isomerase by chemical cross-linking combined with mass spectrometry. *J Proteomics*, 2014, **108**: 1-16
- [119] Mertens H D T, Svergun D I. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys*, 2017, **628**: 33-41
- [120] Tompa P, Han K-H. Intrinsically disordered proteins. *Phys Today*, 2012, **65**(8): 64-65
- [121] Korasick D A, Tanner J J. Determination of protein oligomeric structure from small-angle X-ray scattering. *Protein Sci*, 2018, **27**(4): 814-824
- [122] Rambo R P, Tainer J A. Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu Rev Biophys*, 2013, **42**: 415-441
- [123] Mertens H D, Svergun D I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol*, 2010, **172**(1): 128-141
- [124] Yadav D K, Lukavsky P J. NMR solution structure determination of large RNA-protein complexes. *Prog Nucl Magn Reson Spectrosc*, 2016, **97**: 57-81
- [125] Petoukhov M V, Svergun D I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J*, 2005, **89**(2): 1237-1250
- [126] Rodriguez-Zamora P. Conjugation of NMR and SAXS for flexible and multidomain protein structure determination: from sample preparation to model refinement. *Prog Biophys Mol Biol*, 2020, **150**: 140-144
- [127] Grishaev A, Tugarinov V, Kay L E, *et al.* Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR*, 2008, **40**(2): 95-106
- [128] Schiavina M, Salladini E, Murrall M G, *et al.* Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS. *Sci Rep*, 2020, **10**(1): 19574
- [129] Geraets J A, Pothula K R, Schroder G F. Integrating Cryo-EM and NMR data. *Curr Opin Struct Biol*, 2020, **61**: 173-181
- [130] Amador F J, Kimlicka L, Stathopoulos P B, *et al.* Type 2 ryanodine receptor domain A contains a unique and dynamic α -helix that transitions to a β -strand in a mutant linked with a heritable cardiomyopathy. *J Mol Biol*, 2013, **425**(21): 4034-4046
- [131] Newcomer R L, Schrad J R, Gilcrease E B, *et al.* The phage L capsid decoration protein has a novel OB-fold and an unusual capsid binding strategy. *Elife*, 2019, **8**: 45345

- [132] Leelananda S P, Lindert S. Using NMR chemical shifts and Cryo-EM density restraints in iterative Rosetta-MD protein structure refinement. *J Chem Inf Model*, 2020, **60**(5): 2522-2532
- [133] Shimada I, Ueda T, Kofuku Y, *et al.* GPCR drug discovery: integrating solution NMR data with crystal and Cryo-EM structures. *Nat Rev Drug Discov*, 2019, **18**(1): 59-82
- [134] Zhang K, Keane S C, Su Z, *et al.* Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid Cryo-EM, NMR, and molecular dynamics approach. *Structure*, 2018, **26**(3): 490-498
- [135] Gauto D F, Estrozi L F, Schwieters C D, *et al.* Integrated NMR and Cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat Commun*, 2019, **10**(1): 2697
- [136] Bardiaux B, De Amorim G C, Luna Rico A, *et al.* Structure and assembly of the enterohemorrhagic *Escherichia coli* type 4 pilus. *Structure*, 2019, **27**(7): 1082-1093
- [137] Alderson T R, Kay L E. Unveiling invisible protein states with NMR spectroscopy. *Curr Opin Struct Biol*, 2020, **60**: 39-49
- [138] Usher E T, Showalter S A. Mapping invisible epitopes by NMR spectroscopy. *J Biol Chem*, 2020, **295**(51): 17411-17412
- [139] Gopalan A B, Hansen D F, Vallurupalli P. CPMG experiments for protein minor conformer structure determination. *Methods Mol Biol*, 2018, **1688**: 223-242
- [140] Gopalan A B, Vallurupalli P. Measuring the signs of the methyl ^1H chemical shift differences between major and "Invisible" minor protein conformational states using methyl ^1H multi-quantum spectroscopy. *J Biomol NMR*, 2018, **70**(3): 187-202
- [141] Hansen D F, Neudecker P, Vallurupalli P, *et al.* Determination of Leu side-chain conformations in excited protein states by NMR relaxation dispersion. *J Am Chem Soc*, 2010, **132**(1): 42-43
- [142] Sekhar A, Kay L E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA*, 2013, **110**(32): 12867-12874
- [143] Walinda E, Morimoto D, Sugase K. Resolving biomolecular motion and interactions by R_2 and $R_{1\rho}$ relaxation dispersion NMR. *Methods*, 2018, **148**: 28-38
- [144] Fawzi N L, Ying J, Torchia D A, *et al.* Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat Protoc*, 2012, **7**(8): 1523-1533
- [145] Tiwari V P, Pandit S, Vallurupalli P. Exchangeable deuterons introduce artifacts in amide ^{15}N CEST experiments used to study protein conformational exchange. *J Biomol NMR*, 2019, **73**(1-2): 43-48
- [146] Zhao B, Hansen A L, Zhang Q. Characterizing slow chemical exchange in nucleic acids by carbon CEST and low spin-lock field $R_{1\rho}$ NMR spectroscopy. *J Am Chem Soc*, 2014, **136**(1): 20-23
- [147] Kleckner I R, Foster M P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta*, 2011, **1814**(8): 942-968
- [148] Baldwin A J, Hansen D F, Vallurupalli P, *et al.* Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion NMR spectroscopy. *J Am Chem Soc*, 2009, **131**(33): 11939-11948
- [149] Vallurupalli P, Hansen D F, Lundstrom P, *et al.* CPMG relaxation dispersion NMR experiments measuring glycine $^1\text{H}^a$ and $^{13}\text{C}^a$ chemical shifts in the "Invisible" excited states of proteins. *J Biomol NMR*, 2009, **45**(1-2): 45-55
- [150] Korzhnev D M, Religa T L, Banachewicz W, *et al.* A transient and low-populated protein-folding intermediate at atomic resolution. *Science*, 2010, **329**(5997): 1312-1316
- [151] Vallurupalli P, Hansen D F, Kay L E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci USA*, 2008, **105**(33): 11766-11771
- [152] Hansen D F, Vallurupalli P, Kay L E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR*, 2008, **41**(3): 113-120
- [153] Bouvignies G, Vallurupalli P, Hansen D F, *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*, 2011, **477**(7362): 111-114
- [154] Neudecker P, Robustelli P, Cavalli A, *et al.* Structure of an intermediate state in protein folding and aggregation. *Science*, 2012, **336**(6079): 362-366
- [155] Vallurupalli P, Bouvignies G, Kay L E. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc*, 2012, **134**(19): 8148-8161
- [156] Tiwari V P, Vallurupalli P. A CEST NMR experiment to obtain glycine $^1\text{H}^a$ chemical shifts in "Invisible" minor states of proteins. *J Biomol NMR*, 2020, **74**(8-9): 443-455
- [157] Yuwen T, Sekhar A, Kay L E. Separating dipolar and chemical exchange magnetization transfer processes in ^1H -CEST. *Angew Chem Int Ed Engl*, 2017, **56**(22): 6122-6125
- [158] Fizil A, Gaspari Z, Barna T, *et al.* "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chemistry (Easton)*, 2015, **21**(13): 5136-5144
- [159] Zhou Y, Yang D. $^{13}\text{C}^a$ CEST experiment on uniformly ^{13}C -labeled proteins. *J Biomol NMR*, 2015, **61**(2): 89-94
- [160] Rennella E, Huang R, Velyvis A, *et al.* $^{13}\text{CHD}_2$ -CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins. *J Biomol NMR*, 2015, **63**(2): 187-199
- [161] Bouvignies G, Kay L E. A 2D ^{13}C -CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding. *J Biomol NMR*, 2012, **53**(4): 303-310
- [162] Vallurupalli P, Sekhar A, Yuwen T, *et al.* Probing conformational dynamics in biomolecules *via* chemical exchange saturation transfer: a primer. *J Biomol NMR*, 2017, **67**(4): 243-271
- [163] Sekhar A, Rumfeldt J a O, Broom H R, *et al.* Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *eLife*, 2015, **4**: e07296

- [164] Platzer G, Zerko S, Saxena S, *et al.* ^1H , ^{15}N , ^{13}C resonance assignment of human osteopontin. *Biomol NMR Assign*, 2015, **9**(2): 289-292
- [165] Bertelsen E B, Chang L, Gestwicki J E, *et al.* Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc Natl Acad Sci USA*, 2009, **106**(21): 8471-8476
- [166] Fiaux J, Bertelsen E B, Horwich A L, *et al.* NMR analysis of a 900K GroEL GroES complex. *Nature*, 2002, **418**(6894): 207-211
- [167] Miyanoi Y, Takeda M, Terauchi T, *et al.* Recent developments in isotope-aided NMR methods for supramolecular protein complexes -SAIL aromatic TROSY. *Biochim Biophys Acta Gen Subj*, 2020, **1864**(2): 129439
- [168] Rosenzweig R, Kay L E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu Rev Biochem*, 2014, **83**: 291-315
- [169] Huang C, Rossi P, Saio T, *et al.* Structural basis for the antifolding activity of a molecular chaperone. *Nature*, 2016, **537**(7619): 202-206
- [170] Saio T, Guan X, Rossi P, *et al.* Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science*, 2014, **344**(6184): 1250494
- [171] Ruschak A M, Kay L E. Proteasome allostery as a population shift between interchanging conformers. *Proc Natl Acad Sci USA*, 2012, **109**(50): E3454-E3462
- [172] Gelis I, Bonvin A M, Keramisanou D, *et al.* Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell*, 2007, **131**(4): 756-769
- [173] Remmert M, Biegert A, Hauser A, *et al.* HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*, 2011, **9**(2): 173-175
- [174] Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res*, 2007, **35**(10): 3375-3382
- [175] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 2010, **5**(4): 725-738
- [176] Biasini M, Bienert S, Waterhouse A, *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 2014, **42**(Web Server issue): W252-W258
- [177] Nerli S, Sgourakis N G. Cs-Rosetta. *Methods Enzymol*, 2019, **614**: 321-362
- [178] Zhang Y. Progress and challenges in protein structure prediction. *Curr Opin Struct Biol*, 2008, **18**(3): 342-348
- [179] Bhattacharya D, Cao R, Cheng J. UniCon3D: *de novo* protein structure prediction using united-residue conformational search via stepwise, probabilistic sampling. *Bioinformatics*, 2016, **32**(18): 2791-2799
- [180] Khoo Y, Singer A, Cowburn D. Integrating NOE and RDC using sum-of-squares relaxation for protein structure determination. *J Biomol NMR*, 2017, **68**(3): 163-185
- [181] Shen Y, Bax A. Homology modeling of larger proteins guided by chemical shifts. *Nat Methods*, 2015, **12**(8): 747-750
- [182] Tang C, Schwieters C D, Clore G M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, 2007, **449**(7165): 1078-1082
- [183] Takahashi H, Nakanishi T, Kami K, *et al.* A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat Struct Biol*, 2000, **7**(3): 220-223
- [184] Wishart D. NMR spectroscopy and protein structure determination: applications to drug discovery and development. *Curr Pharm Biotechnol*, 2005, **6**(2): 105-120

Application of Nuclear Magnetic Resonance Spectroscopy for Studying Protein Three-dimensional Structure*

YIN Lin, SHEN Jun-Cheng, YANG Li-Qun**

(School of Chemistry, Key Laboratory for Polymeric Composite and Functional Materials of Ministry of Education, Guangdong Provincial Key Laboratory for High Performance Polymer-based Composites, Sun Yat-sen University, Guangzhou 510275, China)

Abstract The unique three-dimensional structure of protein is closely related to its biological function. Therefore, investigating the three-dimensional structure of protein is helpful to reveal its biological function mechanism. The study of protein three-dimensional structure in the solution state using nuclear magnetic resonance (NMR) spectroscopy can accurately reveal the relationship between protein structure and biological function. The aim of this article is to provide an effective strategy for accurate analysis of protein three-dimensional structure using NMR and combination with other biophysical means such as molecular modeling and computation methods through reviewing the research progress and latest technology in these fields. Firstly, we summarize the theory of NMR for studying protein three-dimensional structure. Secondly, we deeply review the theory and technology of NMR analysis of protein three-dimensional structure, including isotope labeling of proteins (labeling methods, expression systems, and purification techniques), NMR data acquisition and analysis software, analysis of amino acid sequence, secondary structural unit and three-dimensional structure of protein using NMR and combination with other biophysical means (Förster/fluorescence resonance energy transfer (FRET), chemical cross-linking coupled with mass spectrometry (CXMS), small angle X-ray scattering (SAXS), and cryo-electron microscopy (Cryo-EM)), and analysis of excited state structure of protein molecules using Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG RD) and chemical exchange saturation transfer (CEST) techniques. Thirdly, we overview recent researches about application of NMR for analysis of three-dimensional structure of high molecular mass single chain protein and supramolecular protein complex. Fourthly, we elaborate the latest progress in the field of NMR combined with molecular modeling and computation methods. Lastly, we summarize challenges and prospects of application of NMR for studying protein three-dimensional structure in the future.

Key words nuclear magnetic resonance spectroscopy, protein three-dimensional structure, isotope-labeled protein, structural constraint, excited state structure of protein molecule, three-dimensional structure of supramolecular protein complex

DOI: 10.16476/j.pibb.2021.0065

* This work was supported by a grant from the Key Project of Guangdong Natural Science Foundation of China (2017B030311007).

** Corresponding author.

Tel: 86-20-39339796, E-mail: yanglq@mail.sysu.edu.cn

Received: August 31, 2021 Accepted: October 12, 2021