

文章编号: 1004-0374(2010)03-0207-05

· 专题: 2009年国际生化大会译稿 ·

## 蛋白质磁共振 —— 从结构生物学到结构基因组学

Kurt Wüthrich

(瑞士联邦技术研究所)

**摘要:** 在后基因组时代, 随着大量物种全基因组序列的获得, 结构生物学家面临着结构基因组学的新机遇和挑战。与传统的结构生物学不同的是, 结构基因组学的研究主要集中在结构和功能未知并且与从前研究的蛋白质相似性很小的蛋白质。准确的来讲, 结构基因组学通过高通量蛋白质表达、结构解析来完成所有蛋白质家族的结构表征, 从而能够通过结构预测功能。加州结构基因组学联合实验室发展了高度自动化的蛋白质合成、结晶、结构解析生产线。然而由于一些蛋白质不能被结晶, 要想覆盖所有蛋白质结构域还有很大困难。Wüthrich的研究小组通过一些高通量的目的蛋白质筛选和NMR结构解析的方法解决了这一难题。与X射线晶体学解析蛋白质结构相比, NMR技术由于能够解析更接近生理状态的溶液结构而具有互补性。通过获得溶液中的蛋白质稳定性、动力学特征和相互作用信息, 正如在朊蛋白和SARS相关蛋白的研究中所表现的那样, NMR技术从扩大已知的蛋白质结构数据库、新的蛋白质功能到化学生物学研究中都扮演着激动人心的角色。

**关键词:** 磁共振; 结构生物学; 结构基因组学; 自动结构解析

**中图分类号:** Q617; Q503 **文献标识码:** A

## NMR with proteins – from structural biology to structural genomics

Kurt Wüthrich

(Swiss Federal Institute of Technology, Switzerland)

**Abstract:** In post-genomic era, with the availability of the complete DNA sequences of a wide range of organisms, structural biologists are faced with new opportunities and challenges in structural genomics. In contrast to classical structural biology, research in structural genomics is focused on gene products with unknown structures, unknown functions, and minimal similarity to previously studied proteins. A precisely formulated goal of structural genomics is to determine representative three-dimensional structures for all protein families, which requires 'high-throughput' technology for protein production and structure determination, and the long-term outlook is to predict physiological protein functions from knowledge of new three-dimensional structures. The California-based Joint Center for Structure Genomics (JCSG) is one of the four large-scale consortia in the NIH-funded Protein Structure Initiative (PSI), which developed and operates an extensively automated high-throughput pipeline for protein production, crystallization and crystal structure determination. However, there remain gaps in the coverage of protein fold space that are due to certain proteins being not readily amenable to crystal structure determination. Wüthrich's research team works on filling such gaps with a high-output approach, which involves novel strategies of target selection as well as new technology for NMR structure determination. When compared to structure determination by X-ray crystallography, the NMR method is complementary by the fact that atomic resolution structure and other function-related data can be measured under solution conditions close to the physiological milieu in body fluids. By generating data on protein structure stability, dynamics and intermolecular interactions in solution, NMR has an exciting role also in the longer-term challenge leading from the expanding protein universe to new insights into protein functions and chemical biology. As illustrations, structural genomics traits of research on prion proteins and SARS-CoV will be discussed.

**Key words:** NMR; structural biology; structural genomics; automated structure determination

## 1 用磁共振解析蛋白质结构方法的建立

通过磁共振(NMR)来解析蛋白质在溶液中原子分辨率的三维结构具有许多的挑战,1982年,Ernst教授提出的二维NMR实验,如COSY(correlated spectroscopy)、NOESY(nuclear overhauser enhancement spectroscopy)、SECSY(spin-echo correlated spectroscopy)、FOCSY(fold over-corrected spectroscopy)等。COSY是同核化学位移相关谱,谱中对角线外的峰称为交叉峰或相关峰,主要反映相距三键氢(邻碳氢)或两键氢(同碳氢)的耦合关系。NOESY是通过NOE效应来建立核与核之间的联系,空间距离小于 $5\text{\AA}$ 的二个质子之间会产生NOE效应从而在NOESY谱上得到交叉峰,由交叉峰强度可得出核与核之间的距离。SECSY和FOCSY也是磁共振二维谱中的典型代表。从这些谱中我们可以得到许多信息,包括耦合信息和距离信息,这些信息可以用来解析蛋白质结构。图1就是一个 $^1\text{H}$ - $^1\text{H}$ 二维谱的典型代表。

1984年,Wuthrich等建立了磁共振解析蛋白质结构的草案。基本程序如下:(1)蛋白质水溶液的制备。蛋白质浓度在 $1\text{ mmol/L}$ 左右,并且保持均一和稳定。(2)用磁共振谱仪采集蛋白质样品的二维氢谱。(3)按顺序归属每个氢原子的化学位移数值。(4)收集NOE距离约束的数据。(5)通过距离几何来计算蛋白质的三维结构。(6)通过能量最低化来优化蛋白质结构。他们花费了几年时间来收集数据计算球形蛋白质BUSI的结构(图2)。它是首次用NMR方法解析的一个蛋白结构。

生物分子复合体的结构以及相互作用的分子机制也可以用NMR技术解决。环孢素结合亲环蛋白A(cyclophilin A),从而造成的影响可以通过其化学

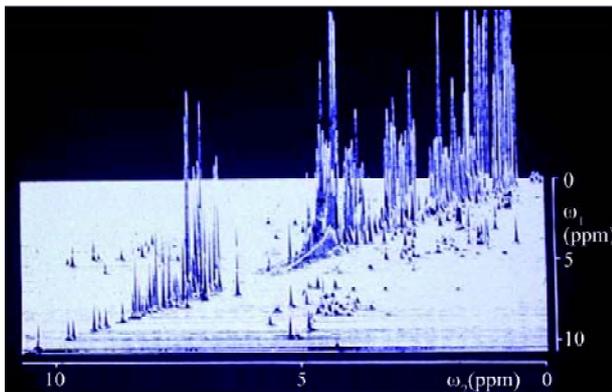


图1  $^1\text{H}$ - $^1\text{H}$ 二维谱的立体展示

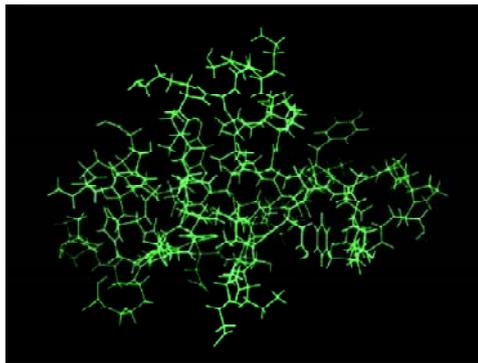


图2 蛋白质BUSI的结构

位移的改变显示出来。亲环蛋白A是免疫抑制剂环孢素A(cyclosporin A)的主要体内蛋白受体,它的细胞内功能主要是肽脯氨酰顺反异构酶(PPIase)活性、介导CsA的免疫抑制、诱导细胞凋亡,在细胞外主要参与炎症应答和氧化应激应答。环孢素A是一种免疫抑制药物,它与亲环蛋白A可以1:1形成复合体。从其复合体结构来看,环孢素A嵌入亲环蛋白A的表面凹槽主要由疏水相互作用介导。其相互作用表面可以非常明确的显示这一点。另外,环孢素A参与结合的位点也比较明确的显示出来(图3)。

## 2 NMR 在结构生物学中的应用

现在有两种方法可以解析原子分辨率的蛋白质的三维结构:蛋白质晶体的X-射线衍射和蛋白质在溶液中的磁共振结构。

血红蛋白的晶体结构是X-射线衍射方法首次解析原子分辨率的蛋白质结构(图4)。血红蛋白是高等生物体内负责运载氧的一种蛋白质。人体内的血红蛋白由四个亚基构成,分别为两个 $\alpha$ 亚基和两个

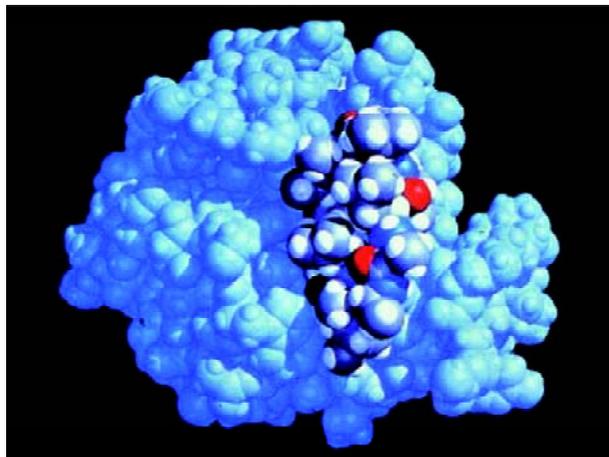


图3 环孢素A与其受体亲环蛋白A的复合体结构

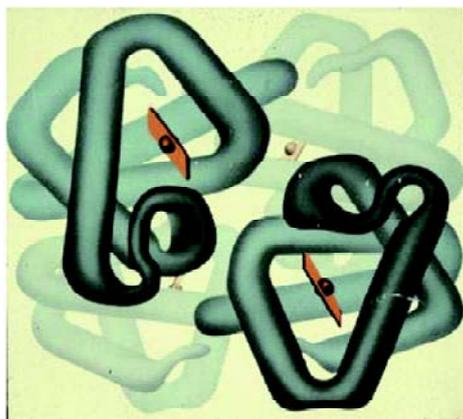


图4 血红蛋白的晶体结构

$\beta$  亚基, 在与人体环境相似的电解质溶液中, 血红蛋白的四个亚基可以自动组装成  $\alpha_2\beta_2$  的形态。血红蛋白的每个亚基由一条肽链和一个血红素分子构成, 肽链在生理条件下会盘绕折叠成球形, 把血红素分子包在里面。血红素分子是一个具有卟啉结构的小分子, 在卟啉分子中心, 由卟啉中四个吡咯环上的氮原子与一个亚铁离子配位结合, 珠蛋白肽链中第八位的一个组氨酸残基中的咪唑侧链上的氮原子从卟啉分子平面的上方与亚铁离子配位结合, 当血红蛋白不与氧结合的时候, 有一个水分子从卟啉环下方与亚铁离子配位结合, 而当血红蛋白载氧的时候, 就由氧分子顶替水的位置。血红蛋白与氧分子的结合具有协同效应, 一个氧分子与血红蛋白第一个亚基的血红素结合可以导致整个血红蛋白分子的构象变化, 从而使得另外的氧分子与这个血红蛋白的结合更加容易。

1970年, 细胞色素C的中心活性区域结构解析出来。细胞色素C是呼吸链的一个重要组成成份, 是一种含铁卟啉基团的蛋白质, 它可以在呼吸链复合酶III(细胞色素还原酶)和IV(细胞色素氧化酶)之间传递电子。细胞色素C中, 血红素就是以其乙烯基与酶蛋白中Cys14、Cys12的巯基形成硫醚键而结合的。中心的铁离子结合和释放电子是细胞色素C发挥其电子转移功能的核心。周边蛋白质为电子转移提供了良好的环境, 并使血红素分子紧密的结合在细胞色素C上(图5)。

细胞色素C分子中含赖氨酸较高, 所以等电点偏碱, 其pH为10.8, 相对分子质量为12 000~13 000。它分为氧化型和还原型两种, 因为还原型较稳定并易于保存, 一般都将细胞色素C制成还原型的, 氧化型细胞色素C在408 nm、530 nm有最大吸收峰,

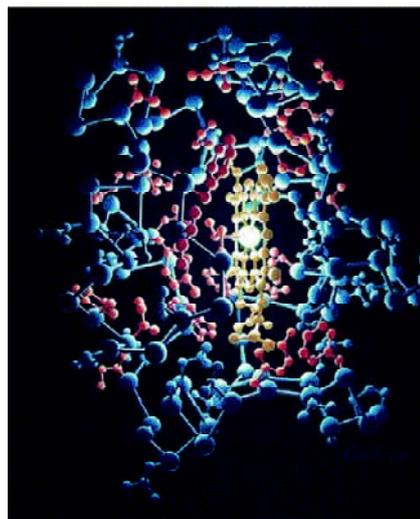


图5 细胞色素C核心功能区域的结构

还原型细胞色素C的最大吸收峰为415 nm、520 nm和550 nm, 这一特性可用于细胞色素C的含量测定。

线粒体细胞色素C释放在细胞凋亡过程中起重要作用。细胞色素C释放到胞质后可引发caspase活化级联反应, 导致细胞凋亡。细胞色素C的释放是线粒体外膜通透性增高的结果。Bcl 2蛋白家族具有调控细胞色素C释放的功能。除细胞色素C外, 线粒体膜间隙的凋亡诱导因子(AIF)在凋亡过程中也释放至胞质, 这两种途径充分保证了细胞死亡程序的有效执行。

除了运载氧, 血红蛋白还可以与二氧化碳、一氧化碳、氰离子结合, 结合的方式也与氧完全一样, 所不同的只是结合的牢固程度, 一氧化碳、氰离子一旦和血红蛋白结合就很难离开, 这就是煤气中毒和氰化物中毒的原理。通过氰化正铁血红蛋白的NMR谱(图6), 我们可以分析其结构上的细微差异, 从而解释氰离子与血红蛋白结合的动力学性质。

### 3 结构基因组学

由多个单位共同协作来进行的结构基因组学, 其目的是用结构生物学的方法研究整个生物体或整个基因组中所有蛋白质和蛋白质相关复合物的结构。自然界中存在着600多万种蛋白质, 已经解析结构的只有5万多种, 通过同源模拟结构数目可以进一步增加。蛋白质可以被分为很多亚家族, 很多亚家族中都只有一个成员, 为每一个亚家族提供至少一个代表成员的结构, 从而覆盖自然界中绝大多数可能的蛋白质结构组成形式, 这是一项很艰巨的

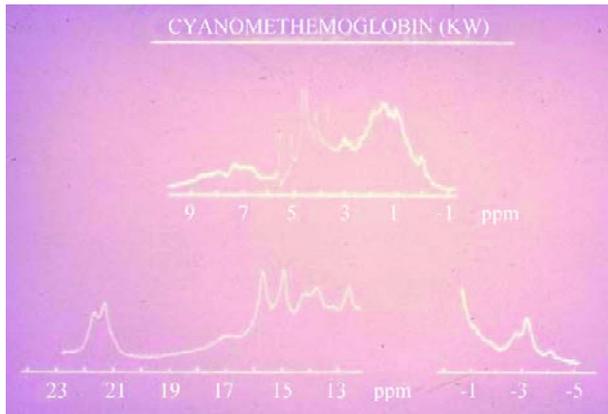


图 6 氰化正铁血红蛋白的 NMR 谱

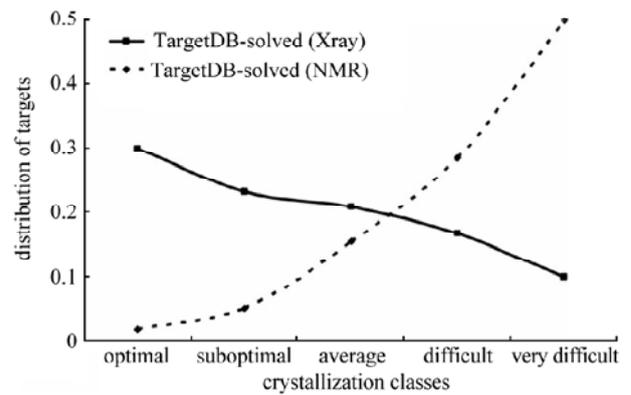


图 7 X 射线晶体技术和 NMR 技术高度互补

任务。但是工作的进行还是相当不错的，结构基因组研究小组在 2008 年 9 月 21 日至 12 月 15 日不到 3 个月的时间里确定了 320 个蛋白质目标，最终完成了 52 个结构。在结构基因组学的研究中 X 射线晶体和 NMR 技术互补得很好，NMR 技术为那些很难结晶的蛋白提供了很好的解决办法(图 7)。

SARS 病毒结构基因组的研究就是这两种技术很好地结合的一个典型，将 SARS 病毒的蛋白质一个个分选出来，对于那些具有无规则结构或者无法长出晶体的蛋白质可用 NMR 技术来解析。

#### 4 蛋白质溶液结构的自动解析

NMR 技术传统上是比较耗时的，结构基因组学研究要求高效的结构解析。近些年来发展出了一整套结构自动解析的方案(图 8)。首先，样品准备好后利用一维 NMR 谱进行初步筛选，合适的蛋白进一步进行标记，检测二维谱图，从而找到比较合适的目标蛋白；其次，双标记之后，采集结构解析所需要的三维谱，利用软件进行主链和侧链的归属，在不同的三维谱之间进行相互确认；最后，经过计算和优化得到精确的三维溶液结构，递交到

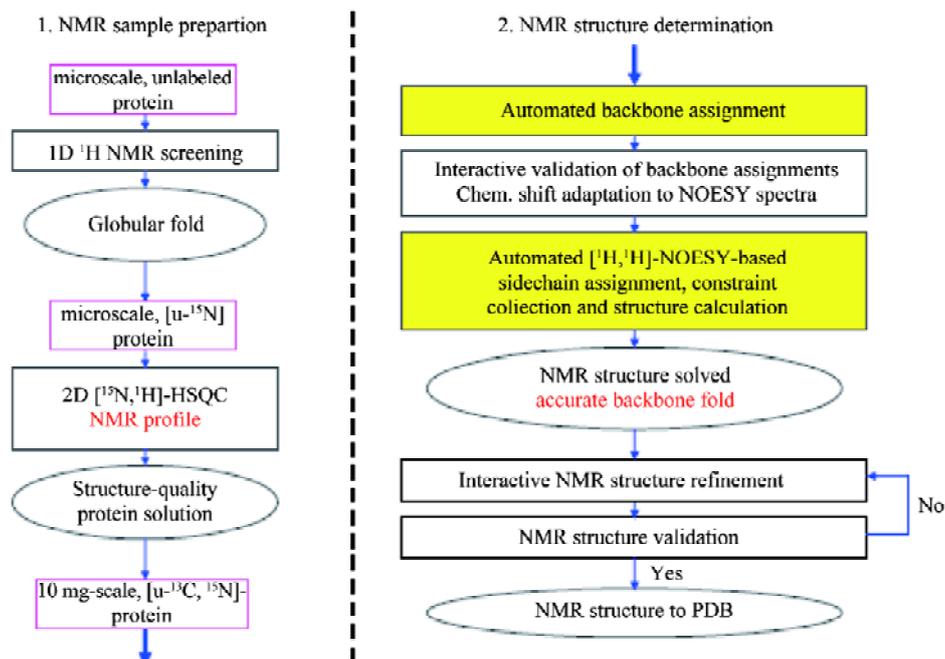


图 8 NMR 结构自动解析流程

