

1 结构生物学——生命科学的前沿

1.1 结构生物学诞生的科学背景

20 世纪初，自然科学中各学科都有明确的学科界限，分科也很细。比如在生命科学中就分有病理学、生理学、微生物学、解剖学、药理学及药剂学等等，而且各学科间界限分明。而自从 20 世纪中叶以来，现代科学技术出现了一种新的趋势：在高度分化基础上的综合化、整体化发展的趋势。由于学科间方法上的相互运用，理论上的彼此借鉴，使学科界限模糊了，不同领域的科学家也开始有了共同语言。特别是生物学进入分子水平，彼此交叉、渗透也越来越深入，逐渐形成了一系列新兴的学科，如：以分子生物学的出现为龙头的分子免疫学、分子病毒学、分子药理学及生物药剂学等等。区别于以上由二级学科交叉所诞生的新学科，结构生物学是一级学科深层次交叉的结果。它不是结构学，而是生命科学。它所面临的是生命科学中的问题，从生命科学出发，又回到生命科学中去。结构生物学的诞生是与生命科学的发展趋势和特征密切相关的。生命科学是从现象到本质研究生命的科学，它的核心是生物学，所以我们从生物学的发展、变化来认识生命科学的发展趋势，进而了解结构生物学诞生的科学背景。

从 20 世纪下半期，生物学就已成为自然科学中引人注目的一颗明星。许多有远见的自然科学家都预言 21 世纪将是生物学的世纪。那么生物学是怎样一步步展现出其蓬勃发展的势头的呢？让我们先看看 19 世纪后半叶人们对生命中遗传的认识过程。说到遗传，我们不能不提到曾是奥地利的一个僧侣的孟德尔（George H. Medel），他是生命科学现代遗传学的奠基人。孟德尔通过对果蝇的形貌与染色体各部分关系的研究，发现杂交能改变遗传，这一发现在遗传学上的伟大意义在几十年后才被人们所认识。当时虽然提出了基因在遗传中的重要性，但其中的基本问题还是不清楚，即：作为基因载体的染色体究竟是 DNA 还是蛋白质？到了 1940 年，基因与酶、核酸与蛋白的关系已开始受到重视，但由于当时人们的错误观念（在结构上认为 DNA 是四链环形分子），还是没能认识基因的本质，更无法指出其与 DNA 的关系。直到 1944 年 Avery 等在研究肺炎双球菌的抗原性和毒性时，发现它们的遗传特性可因加入 DNA 而特异性地改变，从而阐明了 DNA 在遗传中的作用。同时他们的研究也证实了作为基因载体的染色体实际上是 DNA，而不是蛋白质。但由于当时并不了解 DNA 真正的分子结构，所以无法认识到它对遗传学发展的意义。然而，谁也没有想到使这一工作再向前推进一步的原因是一项化学分离技术的发明，这就是马丁（A. J. P. Martin）和辛格（R. L. M. Synge）的纸上层析技术，它使分离组成蛋白质的 20 种氨基酸和组成核酸的四种核苷酸成为可能，从而使蛋白质和核酸组成的容量分析工作又向前发展了。在这些基础上，查加夫（E. Chargaff）运用这项新技术在研究核酸中的四种核苷酸组成的比例关系中，发现 DNA 中脱氧胸腺嘧啶（dT）与脱氧腺嘌呤（dA）等分子数，而脱氧胞嘧啶（dC）与脱氧鸟嘌呤（dG）等分子数。这一发现后来成为沃森（James Watson）和克里克（Francis Crick）所提出的 DNA 双螺旋理论的重要依据。

当时的几位物理学家也看到了生物学未来的发展前景，对生物学发生了极大的兴趣。如量子力学领域的开拓者玻尔（Neils Bohr）于 1932 年发表了《生命和光》（Life and Light）的讲演，他指出生物学在使用了新的概念和一些物理学方法后将能发展到一个新的水平。他的学生德尔布留克（Max Delbrück）决定亲自参与生物学研究。在他们研究噬菌体的过程中证明了 DNA 是遗传信息的载体。另一位量子力学家薛定谔（Erwin Schrodinger）在 1945 年写了一本颇具影响的书《生命是什么》（What is life），书中在论述什么是生命时指出生命的特征在于生命系统能不断减少自己的熵，主张新的物理学要在研究生物学中形成，与玻尔的观点遥相呼应。

从以上对遗传的认识及生物学科发展的历史演变过程我们可以看到生物学是在各个学科的参与、推动下发展的；在与其他学科的交叉、渗透自身获得发展的同时，也为其他学科开拓了新的领域。从上述历史事实，我们完全可以预测生命科学势必发展成为自然科学的一个带头学科。

那么，结构生物学这一生命科学的前沿学科，又是怎样应运而生的呢？面对当时遗传学、微生物学的研究所揭示的大量事实，人们感到生物学科要想进一步发展，必须找到一种合理的遗传物质的结构来解释遗传的功能，来说明遗传过程是如何进行的，来解释为什么形形色色的生物有着令人惊奇的共性：它们都含有相似的化合物，具有相似的组成，并且执行着相似的功能。生物体的基本单位是细胞，而构成生命不同形态的细胞却有着极其相似的分子设计。要想揭开这一个个谜，要想更多地揭示生命的共性和本质，一定要在生物学领域内发生一场深刻的变革和跨时代的飞跃。其标志就是生物学研究从宏观到微观的过渡，把认识的层次从细胞深入到分子。在这一时期生物学领域里的几大突破，如发现 DNA 是遗传物质、DNA 双螺旋结构的阐明以及生物大分子晶体学的诞生都让人们体会到生物学的这场深刻的变革。正是由于生物学研究本身的需要，再加上化学、物理、数学等学科向生命科学的渗透，生物学研究明显地分成了几个层次。其中在分子和原子水平上的研究促进了分子生物学、生物物理学的诞生。值得一提的是，19 世纪以来，化学学科发展很快，不仅创立了原子、分子学说而且开辟了生物化学的研究方向，逐渐形成了生物化学学科，明确了很多生命过程中的重大化学问题，带动了生命科学走向分子水平。而进行分子水平的研究也确实赋予生命科学不可限量的活力和前景。但是从本质上说，仍迫切需要找到一种合理的遗传物质的结构来解释其功能（这就是核酸）。另一方面，为了认识酶的生化本质及解释大量的生化实验结果，也迫切需要蛋白质的结构数据。也就是说迫切需要研究核酸和蛋白质这两类生物大分子的结构和功能，主要是对核酸和蛋白质的高级结构进行测定、分析，并发掘、研究其功能，对结构与功能的关系进行探讨。这就是结构生物学诞生的科学背景。

1.2 结构生物学时代的到来

首先，让我们按时间顺序回顾一下结构生物学发展的历史。谈到结构生物学发展史，我们不能不提到本与生物学无关的伦琴射线的发现在结构研究上所立下的伟大功勋。1895 年德国伦琴（W.K. Rontgen）发现了 X 射线，也就是伦琴射线。一百多年来 X 射线在物质结构特别是晶态物质结构的阐明上做出了巨大贡献。1912 年劳埃（M.F. Laue）发现了晶体的 X 射线衍射现象。把 X 射线的波长与晶体内部原子间的距离相比拟。当它透过晶体而产生衍射时，将能显现出晶体内部的有序结构。22 年后，1934 年物理学家伯纳尔（Bernat）和克朗福

特 (Crowfoot) 拍摄到胃蛋白酶单晶的 X 射线衍射的第一张照片, 但无法解出其结构。事隔 20 年之后佩鲁茨 (M. Perutz) 的一个重大发现解决了这一难题。1953 年, 他用同晶置换技术解决了胃蛋白酶的结构问题。随后的五六年里, 茨德润 (Kendren) 和 Perutz 分别获得了 $6\text{\AA}$ 分辨率的肌红蛋白和 $5\text{\AA}$ 分辨率的血红蛋白晶体结构。与此同时, Watson 和 Crick 在有限的 X 射线衍射数据的基础上, 通过推理和想象共同建立了 DNA 双螺旋结构模型, 由此奠定了分子生物学的基础。这是一个伟大的功绩, 人们从此开始了对生物大分子三维结构的描述。在 1957~1967 年的 10 年里结构生物学又发展到了一个新的阶段。随着溶菌酶结构阐明之后, 胰凝乳蛋白酶 A 及核糖核酸酶 S 也分别得到了高分辨率的衍射结果, 并根据结构分析了溶菌酶对多糖的水解机制、对糖苷键的断裂机理等。这说明作为结构生物学的主要部分——蛋白质晶体学已成为一门成熟的科学。然而, 虽然它起到了结构生物学奠基者和开路先锋的作用, 但还不能被称之为结构生物学。直至 70 年代, 从生物大分子三维结构的测定进入到生物大分子三维结构与其生物学功能关系的研究阶段, 即不仅是测定单个分子而且是开始测定复合物的结构, 不仅是测定静态的、一个复合物的结构而且是测定动态的、一系列复合物不同运动状态的结构, 并开始特别注意于结构与功能关系的研究的时候, 才正式提出了结构生物学的名称。算下来, 结构生物学名称的提出距今已有 29 年 (1972~2001), 但它的快速发展并逐步形成一个新的学科领域, 则是近几年的事。结构生物学针对生物系统的多层次结构研究也应该包括细胞生物学, 但分子层次的结构生物学一直是其发展的主流。也有另一种看法, 即结构生物学就是指分子层次, 所以不必说结构分子生物学或分子结构生物学而特指分子层次。

在英国出版的权威杂志 *Nature* 1993 年 11 月召开的以结构生物学为主题的讨论会上, 曾任哈佛大学、麻省理工学院教授, 现任美国 Brandeis 大学教授的 Petsko 宣称结构生物学的时代已经开始, 实际上当时多种国际性的结构生物学专业刊物的创立就是很好的证明。仅从 1990 年以来至少有 5 种新的结构生物学专业刊物问世, 它们是:

J. Structural Biology (1990)

Current Opinions Structural Biology (1991)

Macromolecular Structure (1991)

Structure (1993)

Nature 新出版的 *Structure Biology* 专刊 (1994)

同时, 原来的一些著名期刊如《分子生物学杂志 (*J. Mol. Biol.*)》和 *Nature* 等刊物所刊登的与结构生物学有关的学术论文也与日俱增, 另外还有一些著名刊物也是主要刊登结构生物学方面的研究内容, 如: *Biological Macromolecules*、*Structure and Dynamics*、*Protein Science*, *Protein, Structure Function and Genetics*, *Biopolymer*, *Structural Dynamics* 以及 *J Biomolecular NMR* 等。这些刊物的诞生和发展的确宣告了结构生物学时代的到来。

1.3 结构生物学是生命科学的前沿和主流

生物大分子 (蛋白质、核酸、多糖) 发挥其生物功能必须具备的条件是: (1) 相对稳定的、特征的三维结构; (2) 三维结构在各个水平上的运动。我们曾提到生物系统的多层次结构, 而就分子水平看, 生物大分子的结构仍具有层次性, 我们称之为不同级的结构。一级结构是指组成蛋白质的 20 种不同的氨基酸及组成核酸的四种核苷酸的排列顺序, 如 DNA 的一

级结构即为脱氧核糖核酸的种类、数量、排列位置和键连接关系。二级结构是指由相邻氨基酸或核苷酸折叠形成的一种规则结构，如蛋白质的 α 螺旋、 β 折叠的构象，DNA 的双螺旋结构都属于二级结构。三级结构主要是指在一级、二级结构的基础上蛋白或核酸链的卷曲，如基因转录调控蛋白的三种构象：螺旋转折螺旋（Helix Turn Helix）、锌指蛋白（Zinc Finger Protein）和亮氨酸拉链式（Leucine Zipper）结构以及质粒 DNA 的超螺旋结构、核酸酶的锤头型结构都属于三级结构。从结构生物学的内涵来看，每个生物大分子都有其特定、固有的三维空间结构，也就是说构成生物大分子的每个原子在空间都占有一个固定的位置，可以测出它的坐标。测定生物大分子的三维结构是研究生物大分子的核心，也是研究其功能的基础。结构生物学就是以生物大分子三维结构的测定为手段，以其三维结构与功能的关系研究为内容，以探讨和阐明生物学各前沿领域中分子层次作用机制和原理为目的的学科。它以生命物质的精确空间结构及其在各个水平上的运动为基础阐明生命现象的本质，在分子层次上从三维结构角度研究当前生物学中各个前沿领域中的重要科学问题。几乎每一个重要的生物大分子及其复合物的高分辨三维结构的测定都是从分子水平上阐明了一个基本的生命现象。反过来说，当前分子生物学中的每个重要问题都是在其三维结构解析的基础上给予阐明的。从近几十年结构生物学及其相关领域获诺贝尔奖的情况可以看出：过去、现在和将来，分子生物学中的每一个前沿突破必将是与其三维结构的突破密切相关。1962 年诺贝尔生理与医学奖是关于核酸双螺旋结构的阐明；1962 年诺贝尔化学奖是关于血红蛋白和肌红蛋白结构的阐明；1964 年的诺贝尔化学奖属于维生素 B₁₂ 的结构和胰岛素结构的阐明；1982 年的诺贝尔化学奖则是关于病毒结构的阐明；1988 年的诺贝尔化学奖是由于解出了细菌光合反应中心——含有四个蛋白质分子的膜蛋白复合物的结构，并首次在分子水平上解释了光能是如何转化为化学能的；1997 年的诺贝尔化学奖的一半也是由于基于结构的三磷酸腺苷（ATP）合成的基本酶学机制的阐明。

2000 年 6 月 26 日美英正式在白宫向世人宣告：人类基因组的工作草图已经完成。这是生命科学历史发展中的一次新飞跃，它标志着生命科学正在迎来一个崭新的时代——后基因组时代（Post-genome era）。在后基因组时代里，生物学的中心任务是揭示全部基因组的功能，而揭示基因组功能的基本途径就是进一步集中研究蛋白质的结构与功能，在获得基因组序列基础上全面了解相关蛋白质及其复合物、组装体的精细三维结构及其在各个水平上的运动和相互作用，特别是结构与功能的关系。事实上生命活动的规律及疾病发生、发展的分子机理是很难仅用基因的知识去阐明的。例如 20 世纪 80 年代就首次发现了癌基因，但作用机制始终不清楚。直到 1997 年第一个癌基因蛋白 Src 及其同源蛋白精细三维结构的测定成功才揭示了这种具有正常功能、而在一定条件下可以发生癌变的分子机制，进而开辟了运用分子开关机理，使 Src 分子保持关闭状态而防治癌症的新途径。这一实例充分说明蛋白质三维结构及其功能的研究在阐明基因功能中起着多么大的作用，也说明在后基因组时代，结构生物学研究将具有战略性关键地位，是生命科学的前沿和主流。

1.4 结构生物学主要研究手段简介

结构生物学实验研究手段中除了经典的研究生物大分子结构的主要三大手段 X 射线晶体学、电子晶体学及多维核磁共振外，还包括扫描隧道显微镜及各种谱学技术，理论研究手段则以计算机分子图形学及量子力学等为主。

1.4.1 X 射线晶体学

X 射线晶体学这样一个最初专门用来测定无机或有机小分子化合物结构的物理学工具的确为结构生物学研究帮了大忙。自 1912 年 Laue 发现晶体对 X 射线衍射,就预示了 X 射线晶体学时代的到来。最初在 1913 年布拉格 (W.L.Bragg) 测定了氯化钠和金刚石的晶体结构,从此, X 射线晶体结构分析就成为测定晶体和分子结构的主要方法。在早期,著名的霍奇金 (D.Hodgkin) 教授应用这个方法测定了胆固醇、维生素 D、青霉素和维生素 B₁₂ 的结构,以后随着结构生物学的发展,蛋白质晶体学已逐渐成为一个非常重要的独立分支学科, X 射线晶体学成为一个非常重要的研究手段。生命科学中的每个重要新发现几乎都伴随着晶体结构的阐明。虽然其他方法也在飞速发展,但直到 1998 年 4 月,从 PDB (Protein Data Bank) 发布的数据看 X 射线晶体结构分析方法所测定的结构仍占 81.9%。可以说 100 年间 X 射线晶体学在物质结构,特别是晶态物质结构的发现上一直占主导地位,目前它正努力在四维结构变化上,在第三代同步辐射新技术的应用方面使得生物大分子精细结构的研究达到更高的水平。

1.4.2 电子晶体学及电子显微学三维重构

X 射线晶体学是借助于 X 射线与晶体的相互作用来研究晶体结构,而电子晶体学则是借助于电子与晶体的相互作用来研究晶体结构的科学。因为电子与物质的相互作用远强于 X 射线,所以电子晶体学适合研究微小晶体与薄膜的结构。像膜蛋白,由于它一般很难分离纯化,也很难得到单晶,又不溶于水,所以电镜就成为很有希望的研究方法。另外, X 射线晶体学是从 X 射线衍射数据来获得晶体的结构信息,电子晶体学则还可以通过磁透镜聚焦,用电子显微成像来获得晶体的结构信息。De Rosier 和 Klug 两位科学家于 1968 年提出的电镜三维重构的思想就是用电子显微成像后,再把试样沿不同方向投影,将得到的一系列投影像进行电子显微学三维重构。这在膜蛋白结构的测定上起了很大的作用。

1.4.3 多维核磁共振方法

核磁共振 (Nuclei Magnet Resonance, NMR) 是测定溶液中生物大分子结构、构象及其相互作用并能给出原子分辨水平详细结构信息的最有力的工具之一。

1957 年 Jardetzky 首先用 NMR 技术研究氨基酸。同年, Saunders 等也发表了关于核糖核酸酯的第一篇 NMR 论文。此后, NMR 方法不断改进,直至二维和多维核磁共振的诞生使生物大分子的结构测定成为可能。X 射线衍射确实是测定生物大分子结构的主要方法,但它对于液体和非晶态固体生物样品中生物大分子的结构分析就受到一定限制。当一个样品的三维结构已确定,要研究它的运动和变化则要靠 NMR 技术。NMR 最主要的优势在于它测定的是液体,它所测定的结构可以是处于生理环境下的动态构象群,所以更能反映生物大分子有机体内的实际情况。NMR 技术目前发展很快,用 NMR 已经可能测定出几万分子量的未知蛋白的结构,像 Wuthrich 等人用二维核磁共振测定蛋白完整的空间结构也已经获得成功。随着 NMR 一整套新技术的运用,如超导核磁、快速傅里叶变换、三共振实验等等, NMR 已经能在溶液或非晶态中测定生物大分子三维结构及其动力学变化。另一方面,最近国际上少数先进的 NMR 实验室也在开展类似于液态分子构象的固体 NMR 的研究工作,比如在固态下,生物膜的高分辨研究也取得了一些重要进展。而且 NMR 方法测定的结构数量也有了迅速的增长,其完整结构测定的分子量上限已突破了 35 000,预计未来几十年内, NMR 方法的测定结

果在结构生物学上将会与 x 射线衍射晶体学的结果并驾齐驱，互相补充。从技术需要来看，优秀的国家实验室应在以上三个主要方向并向发展。

1.4.4 扫描隧道显微术在结构生物学中的应用

第一台扫描隧道显微镜 (Scanning Tunneling Microscope, STM) 是瑞士的葛·宾尼 (Gerel Bining) 和海·罗需尔 (Heinrich Rohrer) 于 1982 年发明、研制出来的。它的出现使人类能够实时地观察单个原子在物质表面的排列状态以及与表面电子行为有关的物理化学性质。这两位科学家因此获得诺贝尔物理学奖。它的原理主要是利用隧道效应，使极细小的一个探针与被研究的物体表面构成两个电极，当它们之间的距离小于 1nm 时，在外加电场的作用下，电子就会穿过两个电极之间的壁垒流向另一电极。隧道电流强度对针尖与样品表面之间的距离非常敏感，所以能通过电流强度测出样品表面的起伏。

迄今为止，用 STM 不仅研究了核酸在天然状态下的结构，包括 B-DNA、A-DNA、Z-DNA、单链 DNA 及 RNA、polyA·polyU 等，还研究了结构蛋白包括胶原蛋白、骨架蛋白及功能蛋白包括溶菌酶、猪胃蛋白酶和蛋白与核酸的复合物的表面结构，均取得了很好的结果。

1.4.5 各种谱学方法

紫外、红外，激光拉曼、荧光、圆二色性光谱等谱学手段能给出生物大分子二级结构及一些溶液构象方面的粗略信息。虽然我们不能仅靠谱学方法获得完整确切的结构信息，但通过谱学方法可以方便、快捷地获取三维结构的运动性方面的信息。特别是激光拉曼技术在结构生物学中的应用已引起了科学家的极大兴趣。

1.5 结构生物学研究的新进展

本节主要介绍与生物大分子相关的结构与功能研究方面的新进展。

1.5.1 生物大分子三维结构的测试在高速发展

因为结构生物学不是描述性科学，不是推论性科学，它所研究的内容必须是定量的，是原子水平上的三维结构，所以生物大分子三维结构测定发展速度就体现了结构生物学的发展，也是其发展的重要标志。

首先，精确测定的生物大分子结构近年来呈指数增长势态。1988 年结构测定速度是每年 129 个，而到 1995 年激增到每天 3.3 个，仅两年后，1997 年就增为每天 5.1 个。这足以说明结构测定发展之快，测定方法之成熟和普遍。其次，近几年内还突破解决了很多难度高、意义重大的结构，使得以精确三维结构为基础揭示重要生命过程的研究达到前所未有的深度和广度。举例如下：

细菌光合作用中心复合物及细菌集光蛋白复合物三维结构的阐明，比较完整地揭示了细菌光合作用的运行机理及光能高效传递的时空关系和分子机制。

细菌视紫红质高分辨 (0.25nm) 精细三维结构的测定是几个高分辨膜蛋白三维结构测定成功的实例之一。

与重要疾病相关的重要结构——全长重组朊病毒 (prion) 蛋白的 NMR 溶液结构的突破，为研究朊病毒的特殊致病机理——构象转换致病机理奠定了结构基础。

在首次测定抗体—抗原复合物结构之后，近年来对人组织相容性抗原及多种 T 细胞及其复合物的晶体结构的解析，使对免疫反应机制与规律有了越来越深入的了解。

1997 年底准确确定的由 8 个组蛋白亚基和 146 个 DNA 碱基对组成的核小体三维结构，是目前在原子水平上测定的最大也是最复杂的蛋白质—核酸复合物三维结构。这一成就为基因转录、DNA 复制提供了精确的结构基础。

另外，人生长激素与受体复合物、类人猿病毒 40、人鼻病毒 14、人免疫缺损病毒 (HIV) 蛋白酶、与 HIV 有强结合能力的糖蛋白 CD4、一氧化氮合成酶、微管蛋白以及 RNA 核酶三维结构的报道都是里程碑式的发展，都是世人瞩目的结构生物学的重大成就。

1.5.2 研究技术方面的新进展

复杂生物大分子精细结构的研究极大地依赖于复杂的技术条件。近年来对结构生物学影响最大的是高能量第三代同步辐射仪的出现，*Science* 杂志将它列为 1997 年 10 大科学成就之一。第三代同步辐射仪可以提供多种波长、比第二代同步辐射仪强 100 倍的高亮度光源和精细 100 倍的细微聚焦能力。这些先进技术的应用对于生物大分子三维结构的研究产生了重大的影响。首先它极大地降低了对晶体大小的要求，X 射线晶体衍射成功的前提是必须有足够大的优质单晶，而使用第三代同步辐射光源可使对晶体要求的尺度从 0.1mm 降至 20~40 μ m，从而使原先过小的膜蛋白的晶体都得到了成功的分析。另一方面它还克服了在一些过大的晶胞，如病毒的分子组装体、复杂的蛋白质、DNA、RNA 及其复合物中由于含原子数极多、衍射点过多或衍射点强度过弱而造成晶体解析的困难。这具有重要的生物学意义，因为几乎所有的生物功能都是通过生物大分子相互作用实现的。目前结构生物学研究已从单分子进入到研究分子间相互作用的复合物以及许多分子构成的复合体，如酶和底物、生物激素与受体、抗原与抗体、DNA 与其调控蛋白等都是令人瞩目的研究对象，特别是一些复杂、庞大的分子组装体结构如高等植物的光合作用系统是由 60 个不同的蛋白质和三条 RNA 链组成的分子量为 230 万的核糖体结构，这些结构的解析对光合作用、激素作用、DNA 复制、基因调控、遗传信息的转录翻译乃至肽链的折叠、卷曲等重要生命过程的分子机理的阐明有重要意义。所以测定生物大分子复合物以及亚细胞器、细胞器的精细三维结构是结构生物学的重要目标。这无疑需要先进的技术条件。

另外，第三代同步辐射仪的先进性还在于大大减少了在同晶置换法中重原子衍生物制备的困难，使得结构测定的速度加快；同时也可以极快速度获取衍射数据，从而使研究快速运动和动力学过程成为可能，使结构生物学研究从生物大分子静态的结构研究进入了动态的结构研究和动力学分析。

多维核磁共振技术也发展很快。价值 700 万美元的 900MHz 核磁共振仪已于 1998 年夏在美国安装，而 1000MHz 的新一代 NMR 谱仪也已在筹划之中。计算机硬件和软件的迅猛发展以及图像显示技术的进步都更进一步推动了结构生物学研究技术的发展。

1.5.3 学科交叉方面的进展

结构生物学主要是一级学科（数、理、化、天、地、生）生物、物理、化学和计算数学深层次交叉的产物。在对生命不同结构层次进行深入探索中会不断有新的边缘学科形成，由于深层次的学科交叉，生物学将面临着理论上大综合和大发展时期。结构生物学是这个时期的产物，它的诞生也促进了学科间深层次的交叉，同样给各交叉学科注入了新的生机。结

构生物学的研究成果越来越受到生命科学各个相关领域的重视和引用，当前任何一本分子生物学相关领域的专著和教科书都在大量引用结构生物学的研究成果，特别是一些三维结构的研究成果。藉此才能定量、深入地阐明有关生物学机理，使其成为一门前沿学科。凡优秀的分子生物学的论文都含有结构生物学内容，这也说明结构生物学已渗透到生物学中各个领域。随着人类基因组测序计划的完成，生命科学进入了“后基因组”时代，解释所有编码蛋白质的结构和功能已成为生命科学新的前沿与挑战。通过结构生物学与基因组学的深入交叉，使得大规模克隆基因、表达蛋白、生长晶体、测定三维结构以快速获取基础蛋白的结构与功能变得不再遥远。同时这种交叉也促成了一个新的学科领域的诞生，即结构基因组学（structural genomics），其最终目标是使基因组的遗传密码与细胞中蛋白质的精细结构与功能直接联系起来。美国 1993 年即启动了测定细菌模式系统中全套蛋白质三维结构的计划。2000 年后美国 NIH 已设立了 6 个结构基因组学的研究机构，这些工作的推进，一定会对新世纪的生物学研究产生不可限量的影响。

（杨铭）

参考文献

1. Drews J. Drug discovery: a historical perspective. *Science*, 2000, 287 (5460): 1960 - 1964
2. Service R. F. Structural genomics: protein data justice for all. *Science*, 2000, 288: 939 - 941
3. Service R. F. Structural genomics offers high-speed look at proteins. *Science*, 2000, 287: 1954 - 1956
4. Service R. F. Wiggling and undulating out of an X-ray shortage and the automated approach to protein structure. *Science*, 1999, 285 (5432): 1342 - 1346
5. Shapiro L, Lima CD. The argonne structural genomics workshop: Lamaze class for the birth of a new science. *Structure*, 1998, 6: 265 - 267
6. Rost B. Marrying structure and genomics. *Structure*, 1998, 6: 259 - 263
7. 王大成. 结构生物学研究的一些新进展. *生物化学与生物物理进展*, 1998, 25 (5): 396 - 403
8. 王大成. 后基因组时代中的结构生物学. *生物化学与生物物理进展*, 2000, 27 (4): 340 - 344
9. 唐有祺. 生命科学与结构分析. 共同走向科学 (百名院士科技系列报告集). 中卷. 北京: 北京科学技术出版社, 1997
10. 王亚辉, 吴志纯, 李家瑶. 走向 21 世纪的生物学——未来生物学预测. 北京: 华夏出版社, 1992
11. 梁栋材. 我国生命科学的前沿问题. 上海: 上海科学技术出版社, 1994
12. 邹承鲁. 世界科技研究与发展. 1995, 17: 9
13. 刘次全, 白春礼等. 结构分子生物学. 北京: 高等教育出版社, 1997
14. Service R. F. NMR researches look to the next generation of machines. *Science*, 1998, 279 (5354): 1127 - 1128

2 生物大分子的计算机模拟

随着人类基因组计划的完成，解读这部由 30 亿碱基对、约 3 万个基因所构成的“天书”的后基因组时代已经到来。基因（组）学、蛋白质（组）学及生物信息学的研究正在飞速发展，不断有新的功能基因被发现、新的蛋白质结构被阐明。

计算机技术的飞速发展，如多处理器的并行计算工作站以及高清晰显示系统的出现，使计算机模拟已经成为药物研究中的一种重要工具。如今，计算机模拟已不再是少数计算化学专家的专用手段，越来越多的药物化学家也对它发生了浓厚的兴趣。计算机模拟的应用领域在持续扩大，已从最初的对实验结果的处理上升到新药物先导结构的设计、生物大分子结构的模拟；与此同时，在从高度复杂的从头计算到半经验的计算中，阐明分子结构与性质的计算方法也在增加，计算的精度在不断提高。

2.1 生物大分子的计算机模拟方法

2.1.1 量子力学

量子力学计算在近 20 年来取得了很大的进展，随着计算机科学的进步，量子力学计算方法所能计算的体系越来越大，计算的精度越来越高。采用从头计算方法可对 100 个原子的小分子进行中等大小的基组计算，密度泛函理论方法的计算可以使计算的精度更高，采用半经验的量子力学计算方法则可对 1 000 ~ 10 000 个原子大小的蛋白质进行计算。新发展起来的量子力学和分子力学或分子动力学相结合的计算方法，为研究大分子的结构以及小分子与大分子的相互作用提供了新的计算途径。

Schrödinger 方程是所有讨论量子力学时的基本出发点。其依赖于时间的全方程式是：

$$\left\{ -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V \right\} \Psi(r, t) = i\hbar \frac{\partial \Psi(r, t)}{\partial t} \quad (2-1)$$

上式表示质量 m 的一个粒子（如一个电子）在外场 V （也许是分子的原子核产生的静电场）下在空间（由位置矢量 $r = x_i + y_i + z_i$ 表示）和时间（ t ）的运动。 $\hbar$ 是 Planck 常数除以 2π ， i 是 -1 的平方根。 Ψ 是波函数，它描述了粒子的运动，通过波函数，可以得到运动粒子的许多性质。当外场 V 不依赖于时间时，波函数可以描述成空间和时间的积，即 $\Psi(r, t) = \Psi(r) T(t)$ ，此时，Schrödinger 方程可以描述成一个更熟悉的与时间无关的形式：

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V \right\} \Psi(r) = E \Psi(r) \quad (2-2)$$

式中： E 是粒子的能量，

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (2-3)$$

将简化的 Schrödinger 方程式的左边缩写为 $\hat{H}$ ，这样 Schrödinger 方程可以描述为 $\hat{H}\Psi =$

子需要考虑，因而需要耗费大量的计算机时。分子力学方法忽略电子的运动，而只将体系的能量看作是原子核位置的函数，因此，它可对含有大量原子的体系进行计算。在某些情况下，分子力学计算能得到与最高级量子力学计算同样精确的结果，而耗时却只有量子力学的几分之一。基于分子力学的简单与有效，它是应用得最为广泛的一种计算方法，现在在个人计算机上也能采用这一方法对某些较简单的体系进行计算。

分子力学计算是基于几个合理的假设，最首要的就是 Born-Oppenheimer 近似。今天所采用的许多分子力学力场都能归纳为一个简单的由体系中分子内、分子间的四个组分所构成的一个通式。

$$\begin{aligned}
 E(r^N) = & \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,o})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,o})^2 \\
 & + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (2-4)
 \end{aligned}$$

$E(r^N)$ 说明分子的势能是 N 个粒子（通常是原子）的位置的函数。这四种作用方式可用示意图表示（图 2-1）。上式中第一项是键长相关项，势能随键长 l_i 偏离参考值 $l_{i,o}$ 的大小而变化；第二项是键角相关项，与键长项类似，分子的势能随键角 θ_i 偏离参考值 $\theta_{i,o}$ 的大小而变化；第三项是分子的扭转势能项，分子的能量随着键的旋转而变化；第四项是非键相互作用项，它不仅包括不同分子间的所有原子（ i 和 j ）间的非键相互作用，同时也包括同一分子中相距三个键以上的原子间的非键相互作用。在简单的力场中，非键相互作用通常采用 Coulomb 势能项表示静电相互作用，采用 Lennard-Jones 势能项表示范德华相互作用。有些更复杂的力场还含有其他项，如氢键相互作用项等，但所有的力场总包含有上述四项。该通式的最大特点是各项能通过内部坐标的变化来描述，如键长、键角及键的旋转或原子的相对运动，这样就理解力场参数的变化如何影响计算能力。

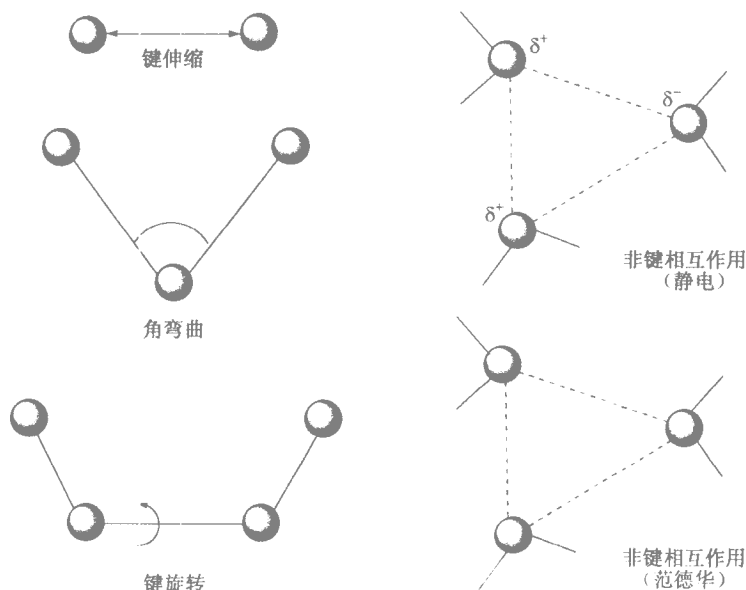


图 2-1 分子力学所描述的原子间相互作用的几种方式

分子力学的力场具有如下几个特征：

(1) 在设定力场时，不仅需要设定其函数形式，而且也需要设定参数，例如式中的 k_i 、 V_n 和 σ_{ij} 。两种力场也许具有相同的函数形式，但可能它们具有非常不同的参数。另外，具有相同函数形式但不同参数的力场和具有不同函数形式的力场可能具有类似的精度。一个力场应是单一的体系，那种将能量分成几个单元，各单元各自从其他力场选取参数，然后将来自不同的力场的参数组合成新的力场的方式，在严格意义上讲是不正确的。尽管如此，力场的某些部分是相对独立于其他部分（特别是键与键角项）的，因此，在某些情况下可以做适当的近似。

(2) 分子模拟用的分子力场原是为了再现分子的结构性质，但它也可以预测分子的其他性质，如分子光谱。然而分子力学在预测光谱时，往往不具备很高的准确性。通常一个力场是设计用于预测某些性质，并对其进行了相应的参数化，如果用这个力场去进行未包含参数的性质预测时，可能会出现力场不能进行设定的情况。

(3) 函数形式和参数的普适性是力场的一个重要特征。这里的普适性是指一套参数可以用于模拟一系列相关的分子，而不是对每个单一的分子需要设定一套新的参数。普适性在用于预测分子的性质时显得尤为重要。

(4) 在分析分子力学时必须牢记的一点是力场是经验的，就力场而言，没有“正确”的形式。当然，如果一个函数形式表现出要好于另一种形式，那么这个力场要优于另一个。大多数通常使用的力场都具有非常相似的形式，并都在试图说明是最佳的函数形式。分子力学力场所使用的函数形式往往是精确度与计算效率的折中结果，最精确的函数形式也许常常不具备满意的计算效率。但随着计算机计算能力的增加，能计算越来越复杂的模型。

(5) 原子形式是大多数力场所有的一个概念。原子形式不仅只是指原子的原子序数，它通常还包括原子的杂化状态和某些情况下原子所处环境。例如，在大多数力场中必须区分 sp^3 杂化碳原子、 sp^2 杂化碳原子和 sp 杂化碳原子等。有些力场除区分了原子的杂化状态外，还针对原子相邻的环境进行了区分。如广泛用于计算小分子的 MM2/MM3 力场对 sp^3 、 sp^2 、 sp 、羰基、环丙烷、自由基、环丙烯和碳翁离子的碳进行了区别。AMBER 力场则将五元环并六元环（如色氨酸）的连接碳与独立的五元环（如组氨酸）的碳区别对待。

2.1.3 分子动力学

分子动力学是建立在牛顿力学基础上的一种分子模拟方法，并已成为生物物理学研究中不可缺少的工具之一。由于分子力学优化得到的仅仅是其初始结构附近的局域能量极小构象，难以实现对生物大分子的构象搜索，运用分子动力学模拟退火方法就可以较好地解决这个问题。

在动力学轨迹中，周期性优化有助于通过确定局域极小构象而分析动力学波动。图 2-2 中与椭圆形能量面叠合的不规则曲线代表粒子从 a 点开始的动力学轨迹，最初的梯度使粒子朝下运动，但由于能量守恒，当它接近最小值时，它的运动速度增加，动力使分子来回运动，在轨迹中到达许多位点，然而，如果使粒子在任意一点停止运动并进行最小化，只能找到一个点。即在椭圆形的能量面，从 a 、 b 、 c 或 d 的任意一点出发，都能收敛于同一点，因此，局域极小可用来表征动力学采集的许多相邻的点。对于有更多自由度的分子，动力学计算时的周期优化产生相对较少的结构（与不连续的动力学结构相比）供详细结构与能量研究用。

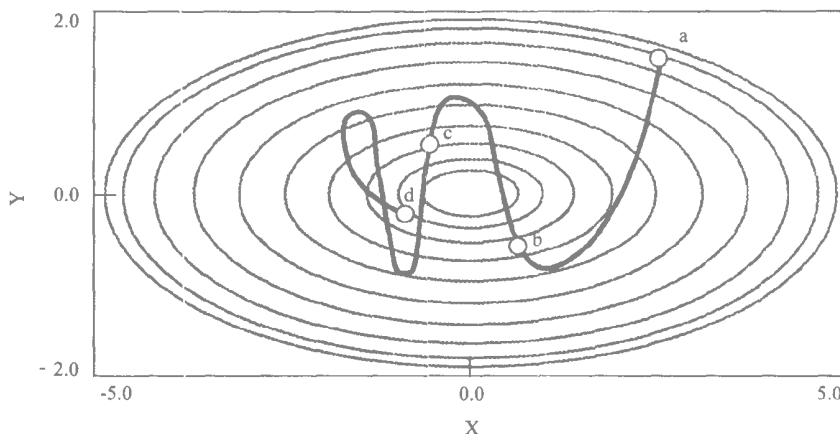


图 2-2 简单能量面的动力学轨迹

运用分子动力学方法已成功地进行了大分子体系低能构象的模建和 x 射线晶体衍射以及 NMR 实验结果的处理。

分子动力学的一个不足之处是在一个轨迹中采集了成百上千个构象，因而使分析的量相应增加。另外，在 300K 以下，某些体系的转变可能需要毫秒至分钟的时间，但现在动力学计算只需进行数十至百纳秒。

2.2 蛋白质三维结构的模建

蛋白质是由氨基酸序列组成，常见的氨基酸一共有 20 种，这些氨基酸通过酰胺键连接形成多肽链。所有天然氨基酸的 α -碳都具有相同的立体化学，而侧链则具有不同的大小、形状、氢键能力和电荷分布，这样使蛋白质具有广泛的功能。蛋白质的生物功能常常依赖于其采取的构象，x 射线衍射和 NMR 是提供蛋白质详细信息的两种方法，但是蛋白质结构的实验确定速度远远落后于蛋白质序列测定速度，因此人们对从氨基酸序列出发理论预测的蛋白质三维结构模建方法产生了浓厚兴趣。

许多力场可用来模拟蛋白质，如 AMBER、CHARMM、CVFF、ENCAD、GROMOS 96 和 OPLS-AA 力场等，另外还有专用于模拟蛋白质 - 配体相互作用的 MMFF 力场。这些力场都含有式 2-4 的基本形式，它们各自就非平面扭曲、Urey-Bradley 相互作用或相距 3 个键以上的原子的相互作用等进行了表述。除 AMBER 力场外，其他力场的参数采用经验方式获取。AMBER 力场采用 6-31G* 基组的 *ab initio* Hartree Fock 计算，获取残基或分子的静电势，使部分电荷能最佳再现量子力学电势。

2.2.1 蛋白质结构的几种基本概念

蛋白质的第一个 x 射线衍射结构显示蛋白质不采取规则或对称的结构，而且是高度不规则的结构。但是一些结构模式在蛋白质结构中常常出现，最常见的模式是 α -螺旋 (helix) 和 β -折叠股 (strand)，它们组成了蛋白质的二级结构。Pauling 早在第一个蛋白质结构被发现之前曾预测 α -螺旋是多肽结构的一种稳定元素，他的预测是基于对小分子晶体结构和多肽单元的几何学分析，可以作为分子模拟预测能力的一个经典例子。 β -折叠股常常形成伸展结

构，称之为 β -折叠片 (sheet)，在 β -折叠片中，链通过平行或反平行形成氢键。二级结构间通过“环” (loop) 区相连，某些环区结构具有共同的构象，如 β -折叠形成的“ β -转角” (turn)。

如果我们不考虑键角与键长的小的变化，那么蛋白质的结构可以通过其旋转键的扭角来确定 (图 2-3)。蛋白质的主链骨架有三个扭角 ϕ 、 ψ 和 ω ，其侧链则可由 χ_1 、 χ_2 等表征。酰胺键偏离平面的旋转具有很高的能量势垒，因此， ω 很少明显偏离 0° 或 180° ，而且多采用 180° 的反式构象。蛋白质的 X 射线晶体衍射结果显示，大多数氨基酸占据 Ramachandran 等势图中的一个低能量区域，但侧链也有采取非常见的高能量构象的情况，侧链的构象常常与骨架的结构相关，某些骨架的构象只在一些特殊侧链结构中才出现。

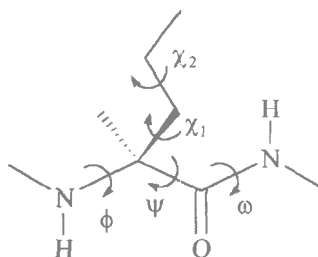


图 2-3 描述蛋白质结构的扭角

2.2.2 蛋白质结构预测的从头计算

从头计算法 (*ab initio*) 是希望解决蛋白质折叠的最渴求的方法，解决问题的办法是搜索分子的构象空间，找到最合适的构象。由于可能的构象数目庞大，因此通常试图能找到能量最低的构象，因为自然界出现的结构应该是能量最低的。几种经验形式的力场常被采用，但它们常因溶剂化作用而引发争议。

Scheraga 及其同事提出了许多搜索多肽和蛋白质构象空间的方法，每种新的方法都经标准的实验分子 (H-Tyr-Gly-Phe-Met-OH) 测试。其中之一的方法是“拼合” (built-up) 法，该方法由三维氨基酸模板构建多肽，每种模板对应于 Ramachandran 图中的低能区域。为搜索多肽构象空间，首先将所有可以符合最先两个氨基酸的可能的模板对接构建二肽片段，每个二肽片段经能量最小化而得到的能量最低构象接着连接第三个氨基酸。多肽采取这种方式逐个连接而成，并经能量优化，而且在每一步都选取能量最低结构。

最简单的 Scheraga 随机搜索方法是随机二面角搜索，在每一步选取一个二面角并随机旋转，得到的结构经能量优化后，按 Metropolis 临界条件决定取舍。静电驱使的 Monte Carlo 方法是一种较为复杂的随机搜索方法，它考虑了多肽和蛋白质中远距离静电相互作用的重要性。它是基于酰胺单元的偶极对蛋白质静电场的影响。

蛋白质比多肽具有更多的可旋转的键，因此常常采取某些蛋白质的简化模型以便使问题容易处理。简化模型的出发点是假定简化模型的势能面能再现详细模型的一般特征，区别只是没有精细结构，由于具有较少自由度的模型的势能面比详细模型的势能面具有较少的能量极小构象，因而出现了许多简化模型用于研究蛋白质的构象空间，许多模型与聚合物的 Monte Carlo 模拟模型类似，如 lattice 和 bead 模型。

2.2.3 组合量子力学 / 分子力学模拟 (QM/MM)

QM/MM 模拟为研究包含蛋白质和 / 或溶剂环境的生物系统的量子力学过程提供了一种有力的手段, 该方法利用需要量子力学处理的过程, 如键的断裂和形成, 通常限于系统的一个小区域, 如活性部位, 该区域只由几十个原子组成, 可由半经验方法、DFT 或 Hamilton 从头算法处理, 而体系的其余部分由经典的 Hamilton 力场描述。量子力学和经典的分子力学的相互作用区域通常由静电和范德华相互作用描述。

QM/MM 方法的一个困难是如何获得合适的量子力学与分子力学的界面。当量子力学仅为分子的一个部分而不是整个分子时, 量子力学与分子力学区域间的共价键必须切断, 这种情况在生物分子的模拟时很常见, 因为量子力学区域通常由活性部位和底物组成。共价键的断开导致在量子力学区域不饱和价态的形成, 因而产生了在量子力学计算时如何处理描述所谓前线键的问题。其中的一个解决办法是采用连接原子的方式, 即将不饱和价态用氢、卤素或特殊的参数化的量子力学原子饱和; 另一个办法是区域自洽场 (LSCF) 方式, 将连接量子力学与分子力学原子之间的键按严格的区域化键轨道处理, 这些区域化键轨道由系统的小类似物的量子力学计算得来, 并假设可移植到所研究的体系, 在整个自洽场 (SCF) 计算过程中, 它被冻结。LSCF 方法对每个新体系都需要一个新的参数, 而通用化的杂化轨道 (CHO 则具有普遍可移植性, 在 CHO 方法中, 冻结的轨道位于特殊参数化的原子, 因此再现整个量子力学系统的键性质。诸如此类的进展预示着 QM/MM 界面问题在不久的将来会被解决。

QM/MM 的应用包括势能面和过渡态结构的确定, 这些通常通过酶 - 底物复合物按反应坐标进行几何优化获得, 计算量能通过优化预先设定的活性区域半径内的原子并固定其他原子而减少。这种计算方式最近被用来研究酚羟基酶、二氢叶酸还原酶等的反应机理。QM/MM 的另一个应用是自由能计算。这种最新的模拟方法采取 Monte Carlo 或分子力学技术按反应坐标相对空间进行采样, 反应坐标通常来源于更小模型体系的高级计算。通过这种模拟, 能计算反应自由能和活化能, 计算出反应速率。采用 QM/MM 自由能计算方法, Cummins 等研究了抑制剂结合二氢叶酸还原酶的离子态。在研究乳清酸核苷单磷酸脱羧酶 (ODCase) 的反应机理时, QM/MM 方法研究结果显示, ODCase 的催化能力来源于基态的去稳定, 而不是过渡态的稳定。基态的去稳定由底物的反应部分的去稳定产生, 这种不稳定作用可由酶与底物的其他部分的强相互作用得到补偿。

2.2.4 同源和比较模拟方法

某些蛋白质在三维结构上有惊人的相似之处, 如胰岛素、胰凝乳蛋白酶和凝血酶。这三种蛋白质都是胰岛素样丝氨酸蛋白酶家族的成员。但也有在生物学上无关的蛋白质具有结构相似性的情况。例如许多蛋白质由 8 个扭曲的平行 β -折叠通过 α -螺旋连接排列桶状结构。同源模拟方法根据蛋白质结构的相似性以已知蛋白质的结构为模板构建蛋白质的三维结构。进行这种模拟首先必须确定采用何种蛋白质作为模板, 然后确定怎样将未知结构的氨基酸与已知结构的氨基酸相配。如果知道蛋白质的生物功能, 常可以相对较直接地确定选取何种蛋白质作为构建的模板, 如果蛋白质的功能未知, 可通过搜索蛋白质结构数据库推出该蛋白属于何种蛋白家族, 这样常可得到蛋白质的特殊功能和某些结构特征。另外, 在某些情况下还可以选取与未知蛋白质的氨基酸序列最接近的蛋白质作为模板, 此时称之为比较模拟较称同源

模拟更合适，因为后者具有未知蛋白与模板蛋白有相似功能的含义，而比较模拟则不需要这层意义。

同源模建方法是目前从已知蛋白质序列预测其三维结构时最常被采用的方法，同源模型的功效取决于其用于什么目的。在某些情况下，只对蛋白质的大致折叠形式感兴趣时，一个相对较低分辨率的模型是可接受的；而在另外一些情况下，如药物设计时，一个拙劣的模型将产生严重的误导，此时有一个劣质的模型还不如没有模型。整体说来，一个模型的好坏在很大程度上取决于两方面因素：未知蛋白质与模板蛋白质之间序列相似率和插入或删除的结构。在模建时，可以得到准确的氨基酸序列排列，错误的排列常会导致错误的结构，这种错误往往在有大环区插入时容易产生，因此也提出了对解决这类问题的新策略的要求。

2.2.5 应用实例

分子系统的计算机模拟能在各种不同分辨率水平上提供丰富的信息，这种方法对于了解蛋白质的折叠机理具有重要的意义。简单的计算方法是将蛋白质的氨基酸残基看作是一个或两个相连的珠子，较高分辨率的计算模型则是考虑绝大多数或所有蛋白质的原子，而更高分辨率的模拟则是包含整个蛋白质和溶剂分子的分子动力学模拟。模拟的力场参数来源于实验和气相的量子力学计算，并已在小体系与实验结果进行过许多比较。但由于体系的复杂性、原子数量大（常常超过 10 000），以及需要时间步长在 $1 \sim 2 \text{ fs}$ ，这种模拟计算常常只在纳秒级。直接的折叠模拟只用在小肽片段。实验结果表明，小蛋白在达到形成部分二级结构的边界稳定状态的时间在 $10 \mu\text{s}$ 级，同时小蛋白能在 $20 \mu\text{s}$ 内折叠，并认为最短时间是 $1 \mu\text{s}$ 。Kollman 等采用 Cray T3E 并行计算机（256 个 CPU）用 Cornell 力场在 300K 下对含有 36 个氨基酸残基的 HP-36 首次进行了长达 $1 \mu\text{s}$ 的从非折叠出发（A），并含有约 3 000 个水分子（TIP3P 水模型）的折叠模拟（图 2-4）。HP-36 是最小的一个能自动折叠的蛋白，它只含有天然氨基酸并且不需要二硫键、寡聚或配体结合来稳定，它在水溶液中的熔点为 70°C ，该蛋白的折叠时间估计在 $10 \sim 100 \mu\text{s}$ ，是折叠最快的蛋白之一。非折叠起始结构来源于 McKnight 等获得的 NMR 结构（图 2-4C）经 1 000K 下 1.0 ns 的动力学模拟，经此处理后，蛋白质处于伸展状态，只有小于 3% 的原始接触，并不含螺旋结构。模拟的时间尺度接近于小蛋白的折叠，模拟结果发现在达到完全折叠状态前，须经过一个折叠中间体，其寿命在 150 ns 。

2.3 核酸结构的模拟

核酸的生物功能——信息储存、复制、转录及翻译等都最终与碱基对序列的化学因素及结构、分子运动和能量的生物物理因素相关。X 射线衍射、NMR 光谱、凝胶电泳、微量热分析等技术常在实验中用于核酸的生物物理性质研究。模拟溶液中大分子结构的最有力的方法是分子动力学计算机模拟方法。分子动力学提供了一种在显微水平描述大分子和溶剂的原子结构和运动的方法。在实践中，分子动力学仍处在发展中，模拟结果依赖于所用力场的近似方式、模拟程序和系统运用。尽管现在的计算时间已可达纳秒级，但有时仍不能满足所计算的核酸体系的要求。

用分子动力学方法模拟核酸的行为是一个极具挑战性的问题，DNA 和 RNA 在中性和生理 pH 条件下是聚电解质，骨架的磷酸酯被充分离子化。可以预见静电电荷 - 电荷相互作用将是决定结构和推动动力学模拟的主要动力。按照 Coulomb 定律，电荷 - 电荷相互作用随距

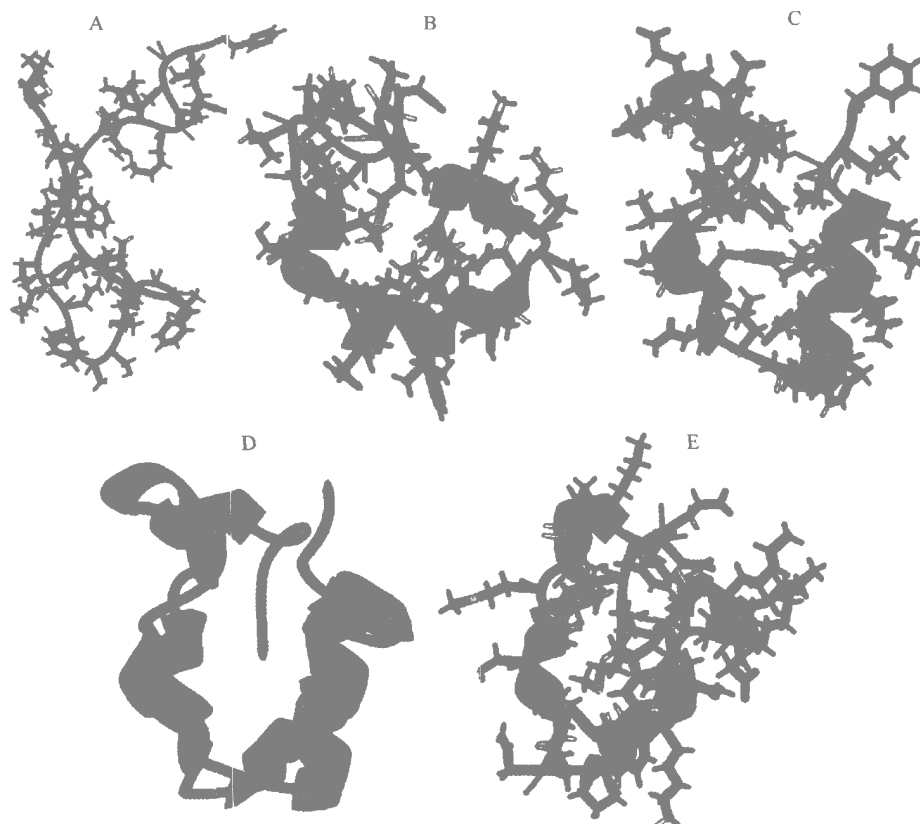


图 2-4 HP-36 的分子动力学模拟

(A) 非折叠初始状态；(B) 部分折叠 (980ns)；(C) NMR 折叠结构；
(D) NMR 结构与最稳定模拟结构的比较；(E) 最稳定模拟折叠结构

离的增大变化缓慢，当采用数字式分割时最易带来误差，另外，与蛋白质相比，核酸具有较大的面积体积比，使得溶剂化效应更加重要，在某些情况下甚至占主导作用。因此，核酸的分子动力学模拟需对远距离静电相互作用和溶剂效应有一个准确的理论处理。

从 1953 年到现在根据实验结果提出的 B-DNA 结构如图 2-5 所示。许多分子动力学模拟方法的研究都是以 d (CGCGAATTCGCG)₂ 序列为对象，Dickerson 等在 1980 年报道了该序列的晶体结构，这是第一个 Watson-Crick 双螺旋的单晶结构（图 2-5c）。该序列含有 *EcoRI* 限制性内切酶的识别区，因此也为 *EcoRI* 十二聚。现在已经获得了许多 DNA 和 RNA 的晶体结构，它们都成为评价分子动力学模拟结果的基础。

早期核酸结构的分子动力学模拟困难重重，粒子网 Ewald (PME) 力场可称为“第一代”力场，它采用了溶剂距离相关静电筛选模型和远距离相互作用分割方法。到了 1995 年，三个完全原子力场的出现，包括一整套核酸参数的出现，尤其是针对溶剂作用模拟的清晰描述，可以称为“第二代”力场的到来。在各种核酸模拟方法中，AMBER、CHARMM 和 GROSMOS 是三种典型代表，其中又以前两种应用最为普遍。值得指出的是，核酸模拟时，带有截止点 (cutoff) 的分子动力学方法不稳定，因此必须采用更准确的处理远距离静电相互作用的方法，如 Ewald 类方法 (PME 或 PPPM) 或较简单一点的力移动方法。

绝大多数生物大分子都处在水环境中，有关液态水的势能函数研究了 30 多年。当今应