

Frontiers of molecular dynamics simulations of protein systems—reexamine from the mesoscience perspective

Ying REN*, Ji XU

State Key Laboratory of Multi-phase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Proteins are the essential parts of living organisms and they participate in virtually every process within cells. An in-depth understanding of the spatiotemporal multi-scale structure and the dominating mechanisms of protein structures would be the basis for scientific exploration of the origin of life, the mechanisms of diseases and the development of new drugs. Due to the limitations of the spatial and temporal resolutions of current experimental methods, computer simulations, especially molecular dynamics simulations, have become one of the most important methods to study the structure and function of protein systems. This article reviewed the progress of molecular simulations and their application in the research of protein systems during the past half century, especially for molecular dynamics simulations and enhanced sampling methods. The time and space involved in protein simulations covers a wide range of scales, which makes it a great challenge to simulate the spatial-temporal multi-scale structures quickly and accurately, or to investigate the physiological process and the underlying dominating mechanisms. Therefore, this article summarized the recent development of the theoretical models and computing algorithms, and their applications in the investigations of the molecular mechanisms of the native structures and structural changes of the structured proteins, the dynamic structure ensemble of intrinsic disordered proteins and the coupled folding and binding with target protein or other biological molecules, protein complex such as molecular chaperonin, virus particle, etc. Furthermore, the evolution of the popular softwares for molecular dynamics simulations driven by the rapid development of high-performance super computers, and their acceleration of the spatial-temporal scales in molecular dynamics simulations of protein systems, were further discussed. At the end of the article, based on the rapid development of mesoscience theory and its successful applications in a variety of complex systems, the future simulation methods and theoretical research of protein systems were prospected.

Learning key points:

- (1) In the last half century, protein molecular models and molecular simulation methods, especially the coarse grained model, non-equilibrium molecular dynamics simulation methods and enhanced sampling methods, have been developing rapidly.
- (2) With the development of computer hardware and simulation software, protein simulation had been extended from the study of protein structure at molecular scale to the structure and function of protein complex.
- (3) From the perspective of mesoscience, the multi-level and multi-scale spatio-temporal dynamic structures of biological macromolecules such as proteins are originated from compromise in competition between different dominating mechanisms.

Key words: protein; molecular dynamics simulation; spatial-temporal multi-scale structure; mesoscience

收稿: 2018-07-09, 修回: 2018-09-04, 网络发表: 2018-09-19, Received: 2018-07-09, Revised: 2018-09-04, Published online: 2018-09-19

基金项目: 国家自然科学基金资助项目(编号: 91434104); 中国科学院过程工程研究所科学研究中心项目(编号: COM2015A005); 多相复杂系统国家重点实验室开放课题(编号: MPCS-2017-D-01)

作者简介: 任瑛(1981-), 女, 河北省保定市人, 博士, 副研究员, 化学工程专业, Tel: 010-82544839, E-mail: yren@ipe.ac.cn.

引用格式: 任瑛, 徐骥. 蛋白质体系分子动力学模拟的前沿进展—从介科学角度重新审视. 过程工程学报, 2018, 18(6): 1126–1138.

Ren Y, Xu J. Frontiers of molecular dynamics simulations of protein systems — reexamine from the mesoscience perspective (in Chinese). Chin. J. Process Eng., 2018, 18(6): 1126–1138, DOI: 10.12034/j.issn.1009-606X.218238.

蛋白质体系分子动力学模拟的前沿进展——从介科学角度重新审视

任瑛*, 徐骥

中国科学院过程工程研究所多相复杂系统国家重点实验室, 北京 100190

摘要: 蛋白质是生命的物质基础, 是生命活动的主要承担者, 对蛋白质时空多尺度结构及其控制机制的深入理解是探索生命起源、病理认知及新药开发的基础. 受实验表征手段及时空分辨率的限制, 计算机模拟已成为研究蛋白质体系结构及功能的重要手段之一. 由于蛋白质体系模拟所涉及的时间和空间跨度均相当大, 因此, 准确且快速地描述其时空多尺度结构, 从而分析体系的控制机制及相关生理过程, 成为分子动力学模拟面临的巨大挑战. 本工作对近半个世纪以来的分子模拟方法, 特别是分子动力学方法和相关的增强采样技术在蛋白质体系研究中的应用进行了总结, 综述了近年来分子动力学的理论模型和算法的发展, 并介绍了这些方法在结构化蛋白质的天然结构与构象变化、固有无序蛋白质的动态结构及其结合底物的动力学过程及分子机理、分子伴侣及病毒等蛋白质复合物体系中的研究成果; 汇总了高性能计算的飞速发展所带动的分子动力学模拟软件的变革, 拓展了蛋白质模拟的时空尺度, 重点阐述了大规模高性能分子动力学模拟在蛋白质研究中的应用; 最后, 基于介科学理论的飞速发展及其在多种复杂体系的成功运用, 对未来蛋白质体系的模拟方法和理论研究的发展趋势进行了思考和展望.

要点:

- (1) 近半个世纪以来, 蛋白质分子模型和分子模拟方法, 特别是蛋白质分子的粗粒化模型、非平衡分子动力学方法及增强采样技术发展迅速.
- (2) 随着计算机硬件及模拟软件的发展, 蛋白质体系的模拟研究已从单分子层次的结构研究拓展到复杂蛋白质复合物的结构及功能研究.
- (3) 从介科学角度审视蛋白质等生物大分子体系多层次、多尺度的时空动态结构, 其物理根源在于不同控制机制在竞争中的协调.

关键词: 蛋白质; 分子动力学模拟; 时空多尺度结构; 介科学

中图分类号: Q6

文献标识码: A

文章编号: 1009-606X(2018)06-1126-13

1 前言

蛋白质是生命体中含量最丰富的功能性高分子化合物之一, 广泛分布于体内所有的器官组织中, 在物质代谢、机体防御、信息传递、组织修复等方面发挥着不可替代的重要作用. 化学、物理学、数学、信息学等学科的最新研究成果已广泛应用于生物学中, 深入到生物学、医学和农业等多个领域. 各学科在生物分子水平上密切联系并相互渗透, 使生物型功能材料、仿生分子器件、多肽蛋白药物、医学造影与生物检测等领域都取得了飞速发展, 生物学及其相关领域也已成为世界上最活跃的研究领域之一.

生命体系呈现典型的多层次、多尺度和介尺度结构. 生物大分子层次包括氨基酸和核苷酸、二级结构、蛋白质/核酸等生物大分子; 细胞层次包括蛋白质等生物大分子、多分子(包括生物大分子和其它分子)形成的超大分子机器、亚细胞器/细胞; 器官层次包括细胞、组织、功能器官; 生命体层次则包括器官、功能系统(如消化系统、血液系统和神经系统)和完整生命体. 对4个层次涉及的边界尺度, 即基本小分子、生物大分子、细胞、器官和生命体, 传统理论研究已较深入, 并逐步形成了不同的学科: 分子生物学、细胞生物学、组织学和系统生

物学. 但对4个层次中介于各自的边界尺度之间的介尺度问题认识十分有限, 分别对应非编码RNA、蛋白质动态结构、细胞器调控、组织和功能系统中的瓶颈问题, 成为现代生物学和医学研究的焦点问题, 久攻不破^[1]. 唯有从分子层次上着手研究生命现象的物质基础, 才能从本质上探讨生命活动的规律, 理解相关疾病的病理及研发新药等.

由于生物体往往处于动态非平衡状态, 蛋白质分子并非呈单一的静态结构, 诸多环境因素, 如温度^[2]、pH值^[3]、离子浓度^[4]及细胞内大分子拥挤环境^[5]等, 都会影响蛋白质的构象及生理功能. 蛋白质的构象可塑性使其能对外界环境因素的变化作出响应, 或与其它配体蛋白或小分子物质相结合以完成特定的生理功能. 作为理解各种生理过程机理的首要条件, 需要精确描述蛋白质在不同环境因素下的动态构象变化. 传统分子生物学实验手段, 如X光晶体衍射、核磁共振、荧光共振能量转移及单分子力谱等, 受限于其较低的时间或空间分辨率, 难以精确测量蛋白质构象的动态变化, 亟需发展能研究蛋白质时空多尺度微观结构的研究方法; 另一方面, 理论计算特别是分子模拟由于其很高的时空分辨率, 是从原子或分子尺度研究蛋白质等生物分子的微观结构和动态变化的重要工具, 被称为计算式“显微镜”^[6], 成为

联系宏观生物学实验现象和其微观本质的重要桥梁,在生物学领域备受关注.近年来,蛋白质体系模拟和实验研究逐步结合,为建立生命体系介科学理论奠定了初步基础.

本工作主要介绍蛋白质体系的分子动力学模拟方法和该领域超级计算的发展及研究现状,并基于介科学理论的发展展望未来蛋白质理论研究的发展趋势.

2 分子动力学模拟方法及蛋白质模型

2.1 分子动力学模拟方法

分子模拟利用计算机构建分子模型模拟分子的微观结构和动态行为,从而获得分子系统的化学和物理性质.分子模拟既可采用量子力学方法,从原子和电子的尺度研究化学反应路径、过渡态及反应机理等实验手段无法研究的过程,也可采用经典分子动力学(Molecular Dynamics, MD)、分子力学和蒙特卡洛模拟方法,从分子尺度考察构象变化等物理过程.其中,分子动力学方法采用势能函数,通过数值求解牛顿运动方程,计算出各原子的受力,再以一定的时间步长对体系进行积分,获得体系中各原子在每个时间步长的运动情况,并通过周期性边界条件对研究对象进行周期性复制,以避免物理上非真实的边缘效应.重复进行受力计算和原子的位置更新,便可获得模拟体系的运动轨迹及模拟系统,并可依据统计物理学理论获得体系的宏观性质和动力学特性.分子动力学模拟中,原子之间的相互作用是通过若干相互作用势函数描述的,不同种类的作用势函数及其参数统称为力场.力场是分子动力学模拟的基础,模拟的准确性主要取决于力场的精度.自 1957 年 Alder 等^[7]首次进行分子动力学模拟以来,分子动力学模拟逐渐成为研究蛋白质微观结构及变构机理的主要工具之一.

2.2 蛋白质模型

蛋白质模拟中应用最广泛的模型是全原子模型,该模型可精确描述蛋白质分子中的每一个原子,从而真实反映模拟体系中各原子间的相互作用.通常通过拟合实验观测数据和量子力学从头计算的数据获得全原子模型的力场参数.目前生物大分子体系模拟中广泛应用的力场有 CHARMM 力场^[8]和 AMBER 力场^[9]等,且涵盖的原子类型仍在不断扩大.虽然全原子模型能充分描述体系的微观信息,但因其从原子和化学键的振动等微观细节开始计算大尺度的构象变化,时间步长一般为 $(1\sim 2)\times 10^{-15}$ s,所以采用全原子模型模拟一个普通的蛋白质分子在毫秒甚至微秒时长的构象变化很难实现,需适当减少模拟体系的自由度以拓展模拟的时空尺度.

粗粒化模型在蛋白质计算中越来越受关注.基于粗

粒化粒子在模型中的定义,每个粗粒化粒子对应全原子模型中的几个原子或原子团,减少了体系中的粒子数,从而显著提升了模拟对象的时间和空间尺度.这种模型适合研究缓慢的蛋白质变构问题或生物分子自组装过程等.

早期研究中常用 Gō 模型,这是一种基于蛋白质分子的晶体结构但更为粗粒化的蛋白质模型^[10],在蛋白质结构模拟中也有不少应用.通常基于实验检测的蛋白质的晶体结构定义,采用位于 C_{α} 原子的作用位点代表各氨基酸残基,天然结构中相互接触的残基间采用近程吸引作用描述,所有不相互接触的残基均采用排斥作用描述,因此 Gō 模型可看作一个拥有光滑能量曲面的理想化模型,以蛋白质天然结构中的相互作用为折叠的主要驱动力^[11].由于粒子数较少且可隐式计算溶剂, Gō 模型的模拟速度远快于全原子模型. Gō 模型常用于蛋白质折叠问题的理论研究,尤其是过渡态和一些折叠中间态等构象可通过计算得到^[12],与大量实验数据定性吻合.但该模型尚有两点缺陷难以克服:首先,其只能用于结构已经测定的蛋白质,且实验测定的构象必须能精确代表溶液中的构象,对实验尚未解析的蛋白质构象无法研究;其次,该模型重点考虑蛋白质或其复合物中的天然接触,对非天然接触作用占主导的体系难以准确计算.

蛋白质模拟中较常用的另一种粗粒化模型是 Marrink 等^[13]开发的蛋白质结构粗粒化方法及相应的 MARTINI 力场,其基本原理是将 4 个重原子(环状结构除外)粗粒化为 1 个粒子,并将粗粒化粒子分为 4 类:极性(P)、介于极性非极性之间(N)、非极性(C)和带电荷(Q),根据接受质子能力不同又将每种粒子类型细分为多种子类型:氢供体(d)、氢受体(a)、氢供体兼受体(da)和既非氢供体也非受体(o).该模型已广泛用于跨膜蛋白、病毒颗粒等复杂蛋白质体系的动力学研究,但因其会丢失原子细节信息,尚无法准确计算蛋白质高级结构的折叠过程^[14].

2.3 增强取样及自由能计算

由于自由能波动较小且整个体系一直沿自由能减少的路径演化,直接分子动力学模拟通常只能在构象空间的一个局部极小区域内取样,很难遍历整个构象空间.目前已发展了多种增强取样方法解决此问题.由于相空间的概率密度与体系自由能密切相关,因而有些方法也被称为自由能计算方法^[15-17].总体而言,增强取样或自由能计算方法大致可分为两类^[18],一类是基于反应坐标(Reaction coordinate)或联合变量(Collective variable)的方法,沿某个或某些选定的反应坐标自由度增强取样能力;另一类是基于温度的方法,通过不同温度间的相互

交换提高自由能的波动范围,从而增强体系所有自由度的取样能力。

在基于反应坐标或联合变量的增强取样方法中,反应坐标一般选取能描述体系特征构象变化较慢的自由度,这样沿反应坐标维度的自由能曲面能描述体系的特征构象变化过程,可直接作用于反应坐标空间局部点的自由能梯度,采用热力学积分(Thermodynamic integration, TI)^[19]等方法计算自由能曲线或曲面;也可在反应坐标空间的自由能曲面上外加偏置作用势以探索更多的构象空间,如伞状取样(Umbrella sampling)法^[20]、Metadynamics 方法^[21,22]等。基于温度的增强取样方法通常通过提高体系模拟温度促使体系快速越过自由能较高的能垒,如副本交换分子动力学(Replica Exchange MD, REMD)方法^[23]。基于反应坐标的增强取样方法和基于温度的增强取样方法各有优缺点。采用基于反应坐标的增强取样方法,需先选取一组体系运动变化较慢的自由度,否则会影响计算的收敛性甚至无法得到收敛的自由能曲面;而对于蛋白质大规模构象变化等复杂问题,选择这样一组反应坐标并不容易。采用基于温度的增强取样方法虽然不需选择特定的反应坐标,但能增强所有自由度的取样能力,所以计算量往往比基于反应坐标的方法大很多。近年来有一些两者结合的方法,如Metadynamics 与副本交换动力学方法相结合^[24],或伞状取样与副本交换分子动力学结合^[25]等,可利用两种增强取样方法各自不同的优点,取样能力往往高于单一方法。

3 蛋白质体系的模拟现状

上述建模方法广泛用于蛋白质系统的结构特征和动力学分析,根据研究对象可有条件地分为结构化蛋白质、固有无序蛋白质、环境诱发的蛋白质变构、复杂蛋白质机器等体系。

3.1 结构化蛋白质的天然结构与构象变化

蛋白质折叠问题伴随着第一个蛋白质晶体结构的解析而出现^[26]。多数天然蛋白质都会折叠成特定的三维空间构象,其氨基酸序列决定最终的三维构象。理解蛋白质折叠机理是一个巨大的挑战,经过多年的理论、实验及模拟研究,近年来对蛋白质折叠问题已有较深入的认识^[27-29]。其热力学研究始于1973年Anfinsen^[30]提出的假说,即蛋白质采取平衡态条件下体系自由能最小的状态,该状态只与蛋白质的氨基酸序列及溶剂条件有关,而与具体的折叠路径和折叠时间等动力学因素无关,蛋白质折叠可以看作一个在肽链构象空间内搜索自由能最小构象的问题。热力学假设为蛋白质折叠问题的体外实验研究奠定了基础,蛋白质的天然状态可通过大

量试管实验而不是复杂的体内实验进行研究。采用最小自由能方法计算蛋白质结构的代表是Rosetta 算法及相应计算软件,关键在于自由能的精确计算及高效的构象搜索算法^[31]。在多届蛋白质结构预测技术的关键测试(Critical Assessment of Techniques for Protein Structure Prediction, CASP)大赛中,Rosetta 作为唯一采用从头计算的方法,其对蛋白质天然结构的预测可达到同源模板法等其它方法的准确度^[32],在第七届CASP 大赛中,Rosetta 在每个类别的蛋白质结构计算中都成为最好的预测器之一^[33]。

但肽链整个构象空间随链长呈指数增长,其构象数量随肽链增长很快变得非常大,详尽遍历整个构象空间并找到自由能最小的构象需要几乎无限长的时间,这与大多数蛋白质微秒级的快速折叠速度不相符,通常被称为Levinthal 悖论^[34]。蛋白质怎样完成快速折叠成为一个重要的动力学问题^[35];另一方面,科学家已发现部分蛋白质如 α 水解蛋白酶(α -Lytic protease)^[36]和丝氨酸蛋白酶抑制剂(Serpins)^[37],它们在生物体内行使生理功能的构象并不是热力学上自由能最小的构象,而是一个自由能局部极小的亚稳状态构象。由于分隔亚稳状态的自由能垒很高,这些构象是动力学稳定的^[38,39],能在生物体内存在较长时间并行使其相应的生理功能。

上述能量曲面理论是天然折叠蛋白质在平衡态溶液条件下得出的,但生物体内环境非常复杂,往往伴随剪切流场、光照、pH 值变化、浓度梯度、大分子拥挤环境、机械力等,导致蛋白质构象变化,如光脉冲能促使光敏蛋白由 β -折叠状态转变成无规则卷曲状态^[40]、机械力能引发肌联蛋白激酶去折叠^[41],从而激活其ATP 结合过程等。对于此类在非平衡环境因素下行使生理功能的蛋白质,外界环境因素主导下,可能有与体系自由能相互竞争的其它控制机制共同主导蛋白质动态构象变化和生理功能的行使。以凝血过程为例,当血管壁破裂时,血清中的血小板会粘附在受损血管壁周围,起到凝血止血的作用。该生理过程开始是血小板粘附受体糖蛋白(Platelet adhesion receptor glycoprotein Iba, GPIba)与血管假性血友病因子蛋白(Von Willebrand Factor, VWF)在血液流场驱动下结合形成复合物,进而引发一系列凝血过程,促使血栓形成达到止血目的^[42,43]。GPIba 与VWF的结合位点主要涉及一个包含12个氨基酸残基的 β -switch 区域,如图1^[44]所示。韩孟之等^[44]对此过程开展的分子动力学模拟表明,无流场时,无规则卷曲状态 β -switch 区域不会由loop 状态自发折叠成 β -折叠状态,而 β -折叠状态则会逐渐去折叠,表明loop 结构对应的体系自由能最小;进一步对流场中的 β -switch 区域

变构过程进行模拟, 结合自由能曲面证明了流体力学作用会导致蛋白质变构, 从自由能最小的 loop 构象越过能垒, 转变为自由能相对较高的 β -折叠状态, 即流体力学

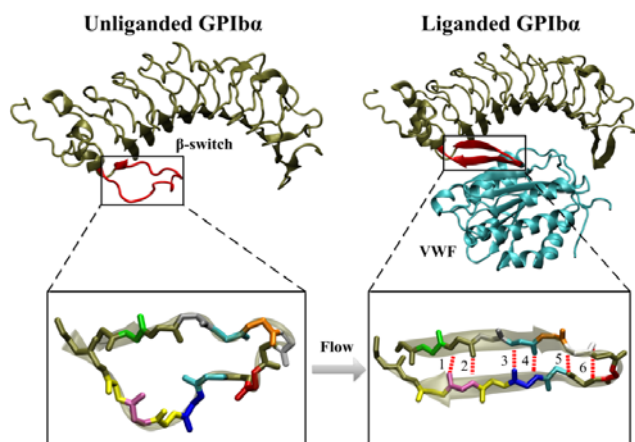


图 1 血小板糖蛋白 GPIIb/IIIa 未结合态与结合态结构示意图^[44]
Fig.1 Crystal structure of unliganded and liganded GPIIb/IIIa subunit^[44]

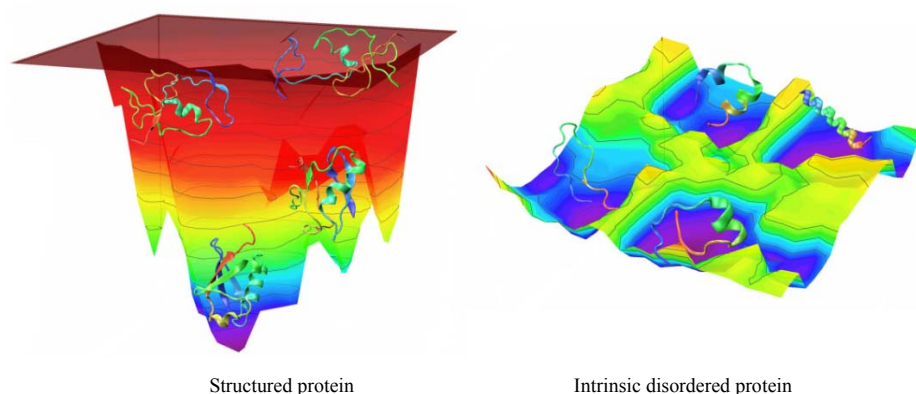


图 2 结构化蛋白与固有无序蛋白的能量曲面示意图

Fig.2 Schematic diagram of the free energy landscape of structured protein and intrinsic disordered protein

固有无序蛋白质的性质使其存在于溶液中时呈随机状态, 但当其与结合配偶体结合时倾向于折叠成规则的三维构象。结合伴侣可以是其它固有无序蛋白^[49,50]、天然折叠蛋白^[51]或核酸^[52,53]等, 固有无序蛋白与其结合后呈规则的二级或三级结构并行使特定的生理功能, 该过程称为结合与折叠的耦合过程(Coupled folding and binding process)^[51,54], 目前已有两种假说描述该过程: 诱导契合(Induced fit)模型^[55]和结构选择(Conformational selection)模型^[56]。诱导契合模型认为固有无序蛋白先以无规则卷曲状态结合到配体伴侣上, 该过程以形成非天然接触作用为主, 在配体伴侣的诱导作用下, 固有无序蛋白逐渐形成天然接触作用并折叠成最终的结合态构象, 即结合过程先于折叠过程进行。与之相反, 结构选择模型则认为固有无序蛋白在溶液中单独存在时会在

作用可视为驱动体系结构变化的另外一种控制机制。

3.2 固有无序蛋白质的动态结构

近年来的研究表明, 生物体内许多功能蛋白质并不会折叠成特定的三维空间构象, 而是以动态构象系综形式存在, 这类蛋白质被称为固有无序蛋白质(Intrinsically disordered protein, IDP)。固有无序蛋白在溶液中单独存在时不能折叠成特定的三维构象, 因此早年在结构生物学研究领域往往被忽略。1999年Wright等^[45]改变了人们对传统蛋白质结构功能理论的认识, 此后固有无序蛋白被广泛研究。后续研究表明固有无序蛋白在生物体内广泛存在, 预测近 1/3 的真核细胞蛋白包含无序序列^[46], 25%的人类基因组编码的蛋白完全无规则, 40%的蛋白质包含至少 30 个氨基酸残基长度的无规则结构域^[47,48]。固有无序蛋白在生物体能行使许多重要的生理功能。如图 2 所示, 由于不存在单一的天然态构象, 固有无序蛋白的能量曲面非漏斗形, 体系自由能趋于最小并不是固有无序蛋白构象变化和行使生理功能的唯一控制机制。

多种构象之间快速转变, 其中包含结合态构象, 配体伴侣会选择性地结合呈结合态构象的固有无序蛋白, 即折叠过程先于结合过程。近年来也提出了一些混合模型(Mixed model), 即在结合与折叠的耦合过程中诱导契合与结构选择共同发挥作用^[57]。虽然已进行了大量实验和理论研究, 但对该过程的控制机制尚不完全清楚, 还需开展更深入的理论研究工作。

固有无序蛋白结合与折叠耦合过程的时间尺度往往在微秒~毫秒级别, 同时由于模拟需同时包含固有无序蛋白和配体, 采用直接分子动力学模拟对这样的体系进行微秒~毫秒级模拟尚难以实现。为减少计算量, 常采用基于结构的 Gō 模型或通过隐式溶剂模型进行研究。通过模拟固有无序蛋白 Cdc42 与 CBD 结合的过程发现, Cdc42 蛋白的结合与折叠过程并不是分离的, 而是密切

耦合的^[58]。固有无序蛋白 NCBD 与 ACTR 的结合过程也采用 Gō 模型进行了模拟^[59], 发现结合与折叠的耦合过程介于诱导契合与结构选择之间。对固有无序蛋白 pKID 及其非磷酸化类似物 KID 蛋白体系的 Gō 模拟^[60]发现, pKID 与 KIX 的结合先于 pKID 折叠进行, 与实验结果^[61]相符, 增加 pKID 折叠比例会降低其与 KIX 结合的速率, 符合诱导契合假说。为减少计算次数, 也可通过隐式溶剂模型的原子分子动力学研究这些系统。对 pKID 及 KID 蛋白通过副本交换分子动力学模拟^[23]发现, 两体系的螺旋含量相似, 但构象空间不同, 磷酸化的 pKID 能限制其部分片段构象空间的取样, 因此 pKID 能与 KIX 结合而 KID 不能^[62]。固有无序蛋白 NCBD 与 ACTR 的结合与折叠过程也采用上述方法进行了 100 ns 的副本交换分子动力学模拟^[63], 但没有得到收敛的构象系综, 表明模拟固有无序蛋白结合和折叠过程比较困难。对肿瘤抑制蛋白 p53 与 S100 钙结合蛋白 B [S100B(β)] 的结合与折叠过程也采用隐式溶剂增强取样方法进行了 MD 模拟^[25], 得到了收敛的二维结合与折叠自由能曲面, 发现二者的结合过程符合诱导耦合模型。虽然采用隐式溶剂能模拟的时间尺度大大超过显式溶剂, 但隐式溶剂模型在描述蛋白质与水分子的相互作用时仍存在诸多限制^[64], 阻碍了其广泛应用。

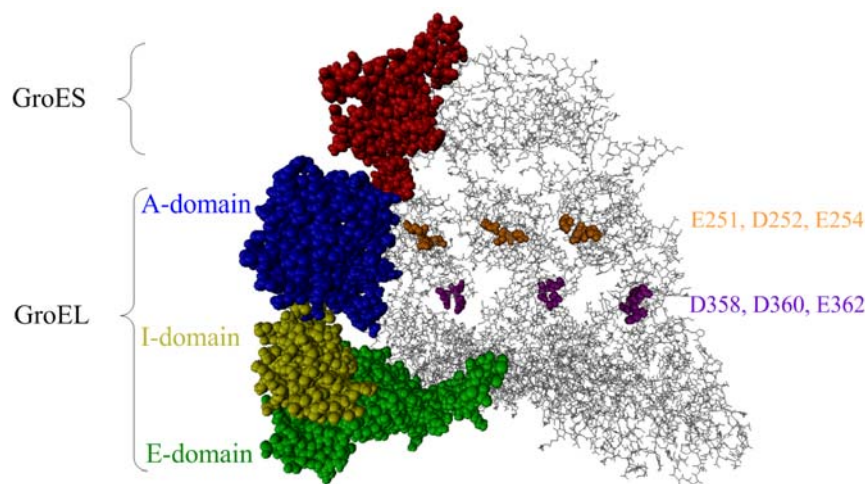
随着计算机硬件水平的提升和增强取样方法的发展, 对较小的固体无序蛋白体系已可进行显式溶剂模型的全原子分子动力学模拟。韩孟之等^[57]采用 Parallel tempering 方法研究了麻疹病毒核蛋白的碳端结构域上的 α -螺旋分子识别因子 α -MoRE(残基编号 484-504)的构象系综, 指出其有部分片段极易形成螺旋结构, 并采用 Metadynamics 与 Parallel tempering 相结合的增强取样方法研究了 α -MoRE 与 XD 蛋白结合与折叠的耦合过程, 获得了其自由能曲面, 提出 α -MoRE 与 XD 结合过程中, XD 蛋白先选择性结合呈螺旋构象的 α -MoRE 残基, 再诱导其它残基折叠成完整 α -螺旋构象, 即诱导契合模型和结构选择模型共同发挥作用。采用类似方法研究了肿瘤抑制蛋白 p53 蛋白碳端结构域(p53 CTD)与 S100B(β)结合并折叠为 α -螺旋构象^[65]、与去乙酰化酶(Sirtuin)结合并折叠为部分转折构象^[66]、与细胞周期素 A2(Cyclin A2)结合并呈无规则卷曲状态^[67]的过程, 由计算得到的自由能曲面可知, 上述 3 种复合物中 p53 CTD 的构象均对应较大的自由能, 只有通过底物蛋白的结合作用才能实现其变构^[68]。这些工作也表明, 固有无序蛋白的动态结构受其自身序列结构和环境中存在的分子对其施加的力作用, 两种机制协调控制。

3.3 蛋白质复合物

由大量蛋白质与蛋白质或其它分子形成的复合物如病毒、蛋白质机器等, 多具有特定复杂结构及生物学功能, 近几年蛋白质复合物是蛋白质领域的研究热点之一。鉴于此类体系所含的生物分子数量众多, 且其发挥生命功能一般涉及结构转变、能量传递等, 采用分子模拟研究如此大的时空尺度需要庞大的计算量, 因此研究工作多为对较小的蛋白质复合物开展全原子模型 MD 计算, 或通过粗粒模型进行研究。

分子伴侣 GroEL/ES 系统是生物体内高度保守的一类蛋白质复合物, 由 GroEL 及辅助分子伴侣 GroES 组成, 可辅助分子量高达约 60 kDa 的底物蛋白在其内部空腔内折叠至天然状态^[69-74]。GroEL 蛋白是同型寡聚复合物, 由 14 个分子量约 57 kDa 的亚基形成 2 个背靠背的七聚物双环, 每个环由 7 个亚基组成, 未折叠的底物蛋白在 GroEL 的空腔内折叠, 折叠好后释放至外部。尽管对 GroEL/ES 的结构及其辅助底物蛋白折叠的基本机理已有了大体了解, 但对部分折叠的底物蛋白释放到分子伴侣的顺式空腔后同其内壁作用并继续折叠至其天然状态的动力学过程仍不甚清楚。Tang 等^[75]对硫氰酸酶在一系列空腔大小及内壁静电作用不同的分子伴侣突变体中的折叠情况进行了实验比较, 证实空腔体积、空腔内壁电荷分布均会显著影响硫氰酸酶在空腔内的折叠速度。因实验很难观察到折叠过程中底物蛋白构象的变化过程, 故就此问题开展了一系列分子模拟研究。Van der Vaart 等^[76]模拟发现硫氰酸酶在同 GroEL 的 A-domain 相互作用的过程中自身可去折叠。Takagi 等^[77]采用 Gō 模型对 8 个残基数少于 100 的小蛋白质受限圆柱形空间的现象进行了隐式溶剂的分子动力学模拟, 已确定蛋白质的折叠速率取决于蛋白质与空腔的相对大小。如图 3 所示, 本课题组^[78]根据相关实验, 构建了野生型分子伴侣、2 个改变空腔体积的突变体及 2 个改变空腔内壁静电作用的突变体, 对硫氰酸酶在 5 个不同分子伴侣的空腔内的折叠动力学进行了 200 ns 全原子分子动力学模拟, 对该过程的微观机理进行了分析和讨论, 认为 GroEL/ES 复合物通过对底物蛋白施加作用力而主动促进其折叠。

驱动蛋白是一种由多个蛋白质分子组成的分子马达, 利用三磷酸腺苷(ATP)水解释放的化学能驱动自身及其携带的货物分子沿微管定向运动, 在细胞有丝分裂、细胞器转移、精子运输等过程中发挥重要作用^[79,80]。目前, 对 ATP 释放的化学能如何转化为驱动蛋白的机械能、驱动蛋白在微管上的运动方式了解甚少。



(Only the cis-cavity is shown for clarity)

Clusters of negatively charged residues exposed toward the cis-cavity are highlighted in orange (E251, D252, E254) and purple (D358, D360, E362) respectively. GroES, the A-domain, I-domain and E-domain of GroEL are shown in red, blue, yellow and green, respectively

图 3 GroEL/ES 复合物结构图^[78]

Fig.3 The structure of GroEL/ES complex^[78]

曹添亮等^[81]结合实验检测了驱动蛋白单体的 3 种特征结构(Kinesin-ATP, Kinesin-ADP, Kinesin-APO), 如图 4 所示, 驱动蛋白通常呈长约 80 nm 的杆状结构, 由马达、颈链、茎部及尾部 4 个结构域组成. 通过分子动力学模拟归纳了驱动蛋白在 ATP 循环中的构象变化, 并通过拉伸分子动力学模拟了驱动蛋白在微管上的迈步

过程, 分析了典型介尺度结构在化学能和机械能转化过程中所发挥的作用^[82], 为驱动蛋白体系的机理分析提供参考.

病毒是一类结构复杂的蛋白质及其它生物分子的复合物, 一般由蛋白质外壳和内部的遗传物质组成, 对其结构及功能的探索一直是生命领域的研究热点. Freddolino 等^[83]采用全原子模拟研究了由 100 万个原子组成的烟草花叶病毒, 揭示了 RNA 在维持病毒衣壳的球状结构中所发挥的重要作用. Zhao 等^[84]基于冷冻电子显微镜观察到人类免疫缺陷病毒衣壳中的 HIV-1 衣壳蛋白的外观结构, 对 216 个 HIV-1 衣壳蛋白六聚体和 12 个五聚体形成的含 6400 万个原子的病毒衣壳进行了全原子计算机模拟, 获得了病毒衣壳的微观结构. Xu 等^[85]基于实验观察到的流感病毒的外部结构及已解析的跨膜蛋白、核蛋白等单体结构, 构建了由 8 段包裹核蛋白的 RNA 链、衣壳蛋白层、外层膜及跨膜蛋白组成的流感病毒颗粒(图 5), 并对此含 3 亿个原子的流感病毒在溶液环境中的动态结构进行了全原子模拟, 展示了病毒颗粒的结构柔性, 为抗流感病毒的药物研发提供了珍贵信息. Rapaport^[86,87]建立了基于形状的粗粒模型, 将蛋白质亚基处理为含少量接触位点的多面体, 用于研究病毒外壳(Capsid)的自组装过程. Arkhipov 等^[88]将乙肝病毒的衣壳处理为 3600 个粗粒化粒子, 采用 MD 研究了其在外力作用下的非可逆形变.

3.4 超级计算及其在蛋白质体系的应用

早期计算机硬件的计算能力较弱, 相应的 MD 模拟软件也以串行代码为主. 最初能模拟的体系规模只有几

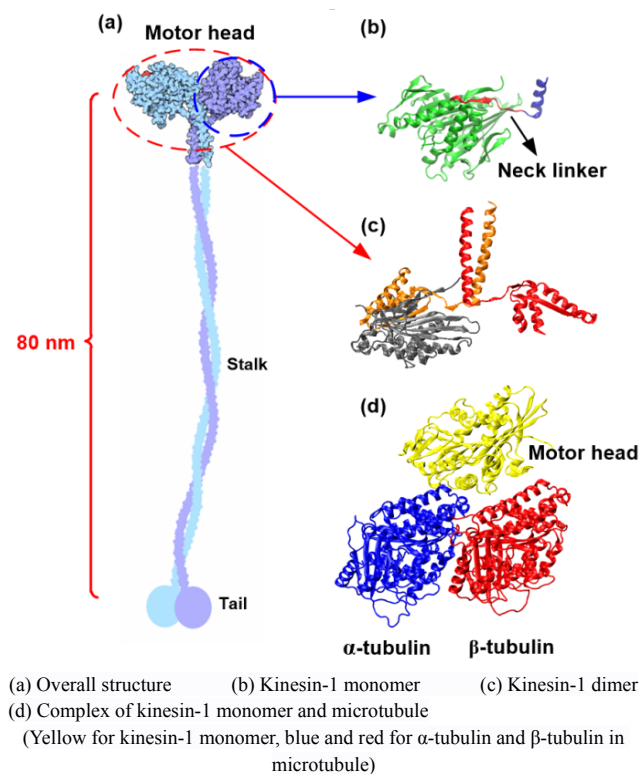


图 4 驱动蛋白体系结构^[81]

Fig.4 The structure of kinesin complex^[81]

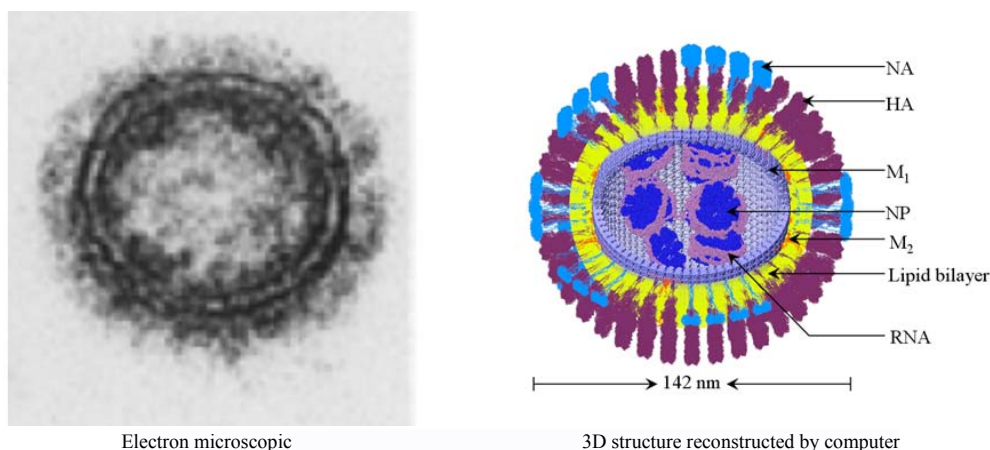


图 5 流感病毒颗粒的电镜检测图片与计算机复建的流感病毒三维结构^[85]

Fig.5 An influenza virion examined by electron microscopic and the 3D structure reconstructed by computer^[85]

百到上千个原子的 10^{-12} s 尺度. 20 世纪 80 年代末, 随着计算机硬件及计算能力的发展, 并行计算架构逐渐得到广泛推广和应用, 模拟的时空尺度也大大增加. 为适合新的计算机硬件架构, 对串行分子动力学代码进行并行化, 诞生了第 1 个可扩展的分子动力学模拟程序 EGO^[89], 这是现代并行 MD 模拟程序的先驱. 随着计算能力的进一步发展, 分布式共享内存的并行计算架构及通过高速网络互联的超大规模并行计算集群逐渐成为高性能计算的主要工具. 早期的分子动力学模拟采用的并行分解算法一般是粒子分解或受力分解等相对比较简单算法. 当模拟计算的体系规模进一步增大、计算所用的中央处理器(CPU)核心数进一步增多时, 其并行效率急剧下降. 为获得最高的计算性能, 进一步开发了空间区域分解算法^[90], 将整个模拟体系的各区域分解到不同处理器进行计算, 特别适合超大规模体系的模拟计算. 同时, 分子动力学模拟的并行化策略也逐步完善, 一些算法如动态负载均衡算法^[91]的引入, 可使计算能力得到更充分的发挥. 基于上述先进的并行分子动力学模拟算法, 提出了许多用于建模可扩展分子动力学的新软件包, 如 GROMACS^[92], NAMD^[93], LAMMPS^[94] 及 Desmond^[95]等, 在蛋白质体系中尤以前两者最为常用.

近年来, 图形处理单元(Graphics Processing Unit, GPU)的计算能力飞速发展, 在科学计算中发挥着越来越大的作用. 在 MD 模拟领域, 采用闲置计算资源进行分布式 CPU/GPU 计算的著名项目是斯坦福大学化学系 Pande 教授主持的 Folding@home 项目^[96]. 该项目于 2000 年启动, 是目前世界上最大的分布式计算项目, 致力于研究蛋白质折叠、误折叠、蛋白质聚合及由此引起的相关疾病等, 包括阿尔兹海默症、牛海绵状脑病(疯牛病)和癌症等. 对较小的蛋白体系, 目前已可成功模拟

5~10 μ s 的折叠过程^[68]. 同时, MD 模拟中的许多算法均已实现了 GPU 多线程并行计算模式, 主流的 MD 模拟软件也实现了 GPU 加速计算. 本课题组^[97]开发了基于 GPU 的通用分子动力学模拟软件包 GPU_MD-1.0.5, MD 模拟的各步骤, 包括邻居列表建立、近程非成键受力计算、成键作用受力计算、更新步骤、约束算法及长程静电作用的 PME 计算^[98]等全部在 GPU 上进行, 计算过程中除输出步骤外不需 GPU 与 CPU 间的任何数据通讯, 计算效率较高. 后续工作又对其进行了并行化, 使其可在多个 GPU 卡上并行运行, 以进行超大规模体系的模拟研究. 应用此程序及 Mole-8.5 超级计算机的约 1300 GPU 卡^[99], 首次对直径 135 nm、总原子数超过 3 亿个的流感病毒水溶液体系进行了全原子模拟^[85]. 该工作涉及 6 种共 4527 个蛋白质分子和 8 段 RNA 结构, 是目前国际上规模最大、结构最复杂的生物分子体系原子层次的 MD 模拟, 可用于研究流感病毒的结构及遗传复制等生命过程、研究流感药物和流感病毒的作用, 并将其用于开发新的抗病毒药物.

除 GPU 外, 针对特殊问题还可通过开发特殊的计算机硬件加速计算, 如 Shaw 等^[100]针对蛋白质体系的 MD 模拟开发了 Anton 超级计算机, 通过特殊的芯片架构及多路并联的网络整体设计与与此硬件结构相匹配的 MD 程序, 极大提高了计算速度. 最初设计的 Anton 超级计算机拥有 128 个处理器, 能在可接受的时间内对一个具有 24000 个原子的蛋白质水溶液体系进行 17 μ s 的 MD 模拟^[101]. 目前最新推出的 Anton 已有 512 个处理器, 对 23558 个原子的体系可达到 17 μ s/d 的计算速度, 成功地将蛋白质分子 MD 模拟的时间尺度提升到微秒级^[102]. 采用 Anton 的硬件条件, Lindorff 等^[35]对 Villin 等 12 个含几十个残基的小型蛋白质通过 MD 模拟观察

到了 400 多次可逆的折叠和去折叠过程, 模拟结果与实验结果的误差在 0.45 nm 以内; 并对 3 种蛋白质(其中两种为多结构域蛋白)开展了 10 μ s 时间量级的 MD 模拟, 发现蛋白质分子呈非平衡特性, 且在小时时间尺度(10^{-12} s)和大时间尺度(10^{-5} s)呈现相似的动力学特性, 挑战了传统采用平衡态方法研究蛋白质体系结构的思路^[103].

4 从介科学角度展望蛋白质体系研究

生命科学、物质加工科学和自然界这 3 个系统毫无例外地呈现多层次, 每一层次又是多尺度结构. 传统研究方法关注每一层次的单元尺度和系统尺度, 而认识这种多尺度问题的关键在于介尺度结构, 即介于单元和系统之间的尺度上表现出的动态非均匀结构, 或来源于这类动态结构的静态结构, 是各领域面临的共同挑战^[104]. 我国已于 2012 年启动了“多相反应过程中的介尺度机

制及调控”重大研究计划^[105], 旨在推广介科学研究方法在不同复杂体系中的应用及完善相应的介科学理论.

在由蛋白质、核酸等生物分子组成的生命体系中, 对氨基酸、蛋白质等生物分子、细胞、器官和生物体等边界尺度的实验和理论研究相对较深入, 但对各层次上介于各自边界尺度之间的介尺度问题, 如蛋白质折叠及其在环境中的变构现象、蛋白质机器等生物大分子自组装体等体系的动态结构等, 对其介尺度结构及控制机制的理论认识十分有限, 已成为生命科学研究的瓶颈. 如图 6^[104]所示, 同化工、材料等复杂体系中的介科学问题类似, 生命体系中的这些介尺度结构及现象也具有非线性、非平衡的复杂性, 物理上都可归因于不同控制机制在竞争中的协调, 需采用介科学研究方法才能从根本上描述蛋白质等生物大分子体系多层次、多尺度的时空动态结构^[106,107].

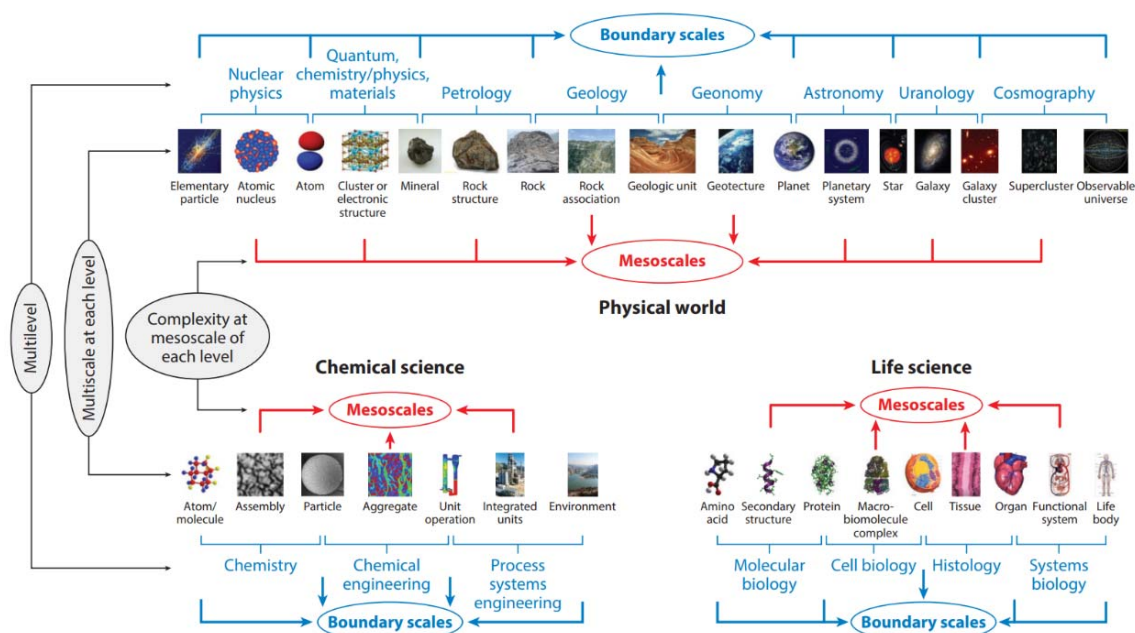


图 6 物质世界多层次、多尺度的特性及介尺度结构的复杂性^[104]

Fig.6 The multilevel, multiscale characteristics and mesoscale complexity of our world^[104]

对复杂蛋白质体系目前多采用关联型多尺度计算方法进行研究, 如通过量子化学计算研究蛋白质分子内部或蛋白质同其它分子间不同原子或化学基团间的作用力或作用势, 通过数学拟合得到力势曲线, 应用于小规模的原子模型计算中, 再针对原子模型计算所得化学基团间的相互作用及溶剂化自由能等, 拟合得到粗粒化粒子间的作用势, 开展多分子体系的粗粒化模拟. 这种关联型多尺度计算方法虽然通过由下至上的半经验性处理获得了不同尺度的模型参数, 但由于其对介尺度结构作了不同粒度的均匀化处理, 因此适用范围受到一定

限制. 例如依此思路开发的 Martini 粗粒化力场将每个残基粗粒化为一个或多个粒子, 可用于蛋白质分子间聚集过程的研究, 但无法描述蛋白质折叠及变构现象. 一个可能的研究方向是开发基于蛋白质分子介尺度结构的粗粒化模型, 模拟参数依赖于介尺度结构, 而非通过数学拟合得到.

借鉴流态化、湍流等复杂体系的介科学理论, 复杂系统中存在两种或多种控制机制, 不同控制机制在竞争中协调, 导致体系呈现复杂的时空介尺度结构. 因此, 对蛋白质体系中控制机制和稳定性条件的理论研究是

建立其介尺度模型的前提。蛋白质分子通过时空多尺度的动态结构行使生理功能,对此非平衡体系,自由能并非决定蛋白质结构的唯一因素,而是受多种控制机制协调作用,这一观点目前已得到初步证实,但控制机制和稳定性条件仍需进一步明确。

从介科学角度重新审视蛋白质体系的研究发展,有如下建议:基于对蛋白质分子介尺度结构的模拟计算、实验检测和理论分析,量化表达其控制机制在竞争中的协调关系,并建立蛋白质分子的稳定性条件,据此建立描述蛋白质分子动态结构的简化模型;在蛋白质单分子模型的基础上,考虑蛋白质分子间的相互作用,形成一套宏尺度模型,通过浓度、温度、时间等变量求解多种蛋白质随时间和空间的分布变化及动态结构,并可拓展到多蛋白质体系,如复杂蛋白质机器、细胞器、细胞等,研究生命体中不同蛋白质的复杂作用网络。该工作对了解生命过程、开展病理药理研究等具有重要意义。

5 结语与展望

深刻认识蛋白质体系的时空多尺度动态结构是探索生命奥秘的关键。近半个世纪以来,分子动力学模拟等计算机模拟方法不断发展完善,超级计算能力也迅速提升,对蛋白质结构的研究日益深入,但尚有诸多分子机理有待进一步揭示。蛋白质体系模拟计算中存在的问题及发展方向可总结如下:

(1) 单个蛋白质分子结构演化的动力学行为并不能准确反映其系综平均特性,因此,应对单分子开展长时间分子动力学模拟,或对多种不同初态结构分别获得模拟轨迹以计算统计学特性;另一方面,增强取样方法与分子动力学计算耦合可获得体系的自由能曲面,便于识别蛋白质稳态结构及关键中间态结构等。

(2) 大规模并行计算的发展为蛋白质体系的分子动力学模拟计算带来了难得的机遇,使超大规模蛋白质复合物体系的微观研究成为可能。

(3) 采用非平衡分子动力学模拟可研究蛋白质体系在流场、浓度梯度等外场作用下的结构转变,推动传统分子动力学模拟从研究蛋白质体系的平衡态结构拓展至研究其生命功能的行使过程。

(4) 从介科学的角度重新审视生命体中蛋白质分子的动态结构及其功能运动,控制机制和稳定性条件的识别和建立是关键。在此基础上可进一步对蛋白质体系建立简化的分子模型,最终建立考虑蛋白质不同特征结构随环境因素演化过程的宏尺度方程,描述不同蛋白质的复杂作用网络,以从分子水平系统研究生命体内部的时空不均匀结构。

参考文献

- [1] Li J H. Exploring the logic and landscape of the knowledge system: multilevel structures, each multiscaled with complexity at the mesoscale [J]. *Engineering*, 2016, 2(3): 276–285.
- [2] Guo M H, Xu Y F, Gruebele M. Temperature dependence of protein folding kinetics in living cells [J]. *Proc. Natl. Acad. Sci. U.S.A.*, 2012, 109(44): 17863–17867.
- [3] Vila-Viçosa D, Campos S R, Baptista An M, et al. Reversibility of prion misfolding: insights from constant-pH molecular dynamics simulations [J]. *J. Phys. Chem. B*, 2012, 116(30): 8812–8821.
- [4] DeMarco M L, Daggett V. Local environmental effects on the structure of the prion protein [J]. *Comptes Rendus Biologies*, 2005, 328(10): 847–862.
- [5] Ellis R J, Minton A P. Cell biology: join the crowd [J]. *Nature*, 2003, 425(6953): 27–28.
- [6] Lee E H, Hsin J, Sotomayor M, et al. Discovery through the computational microscope [J]. *Structure*, 2009, 17(10): 1295–1306.
- [7] Alder B, Wainwright T. Phase transition for a hard sphere system [J]. *J. Chem. Phys.*, 1957, 27(5): 1208–1209.
- [8] MacKerell A D, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins [J]. *J. Phys. Chem. B*, 1998, 102(18): 3586–3616.
- [9] Cornell W D, Cieplak P, Bayly C I, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules [J]. *J. Am. Chem. Soc.*, 1995, 117(19): 5179–5197.
- [10] Ueda Y, Taketomi H, Gō N. Studies on protein folding, unfolding, and fluctuations by computer simulation: II. a three-dimensional lattice model of lysozyme [J]. *Biopolymers*, 1978, 17(6): 1531–1548.
- [11] Prieto L, de Sancho D, Rey A. Thermodynamics of Gō-type models for protein folding [J]. *J. Chem. Phys.*, 2005, 123(15): 154903(1–8).
- [12] Hills R, Brooks C. Insights from coarse-grained Gō models for protein folding and dynamics [J]. *Int. J. Mol. Sci.*, 2009, 10(3): 889–905.
- [13] Marrink S J, Risselada H J, Yefimov S, et al. The MARTINI force field: coarse grained model for biomolecular simulations [J]. *J. Phys. Chem. B*, 2007, 111(27): 7812–7824.
- [14] Marrink S J, Tieleman D P. Perspective on the Martini model [J]. *Chem. Soc. Rev.*, 2013, 42(16): 6801–6822.
- [15] Kollman P. Free energy calculations: applications to chemical and biochemical phenomena [J]. *Chem. Rev.*, 1993, 93(7): 2395–2417.
- [16] Christ C D, Mark A E, Van Gunsteren W F. Basic ingredients of free energy calculations: a review [J]. *J. Comput. Chem.*, 2010, 31(8): 1569–1582.
- [17] Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration [J]. *Entropy*, 2013, 16(1): 163–199.
- [18] Bernardi R C, Melo M C R, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems [J]. *Biochim. Biophys. Acta*, 2015, 1850(5): 872–877.
- [19] Kirkwood J G. Statistical mechanics of fluid mixtures [J]. *J. Chem. Phys.*, 1935, 3(5): 300–313.
- [20] Torrie G M, Valleau J P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling [J]. *J. Comput. Phys.*, 1977, 23(2): 187–199.
- [21] Laio A, Parrinello M. Escaping free-energy minima [J]. *Proc. Natl. Acad. Sci. U.S.A.*, 2002, 99(20): 12562–12566.

- [22] Barducci A, Bonomi M, Parrinello M. Metadynamics [J]. Wiley Interdiscip. Rev. Comp. Mol. Sci., 2011, 1(5): 826–843.
- [23] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding [J]. Chem. Phys. Lett., 1999, 314(1): 141–151.
- [24] Bussi G, Gervasio F L, Laio A, et al. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics [J]. J. Am. Chem. Soc., 2006, 128(41): 13435–13441.
- [25] Chen J. Intrinsically disordered p53 extreme C-terminus binds to S100B ($\beta\beta$) through “fly-casting” [J]. J. Am. Chem. Soc., 2009, 131(6): 2088–2089.
- [26] Dill K A, Ozkan S B, Shell M S, et al. The protein folding problem [J]. Annu. Rev. Biophys., 2008, 37: 289–316.
- [27] Wolynes P G. Evolution, energy landscapes and the paradoxes of protein folding [J]. Biochimie, 2015, 119: 218–230.
- [28] Compiani M, Capriotti E. Computational and theoretical methods for protein folding [J]. Biochemistry, 2013, 52(48): 8601–8624.
- [29] Dill K A, MacCallum J L. The protein-folding problem, 50 years on [J]. Science, 2012, 338(6110): 1042–1046.
- [30] Anfinsen C. Principles that govern the protein folding chains [J]. Science, 1973, 181: 223–230.
- [31] Rohl C A, Strauss C E M, Misura K M S, et al. Protein structure prediction using Rosetta [J]. Methods Enzymol., 2004, 383: 66–93.
- [32] Das R, Baker D. Macromolecular modeling with Rosetta [J]. Annu. Rev. Biochem., 2008, 77: 363–382.
- [33] Das R, Qian B, Raman S, et al. Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home [J]. Proteins, 2007, 69(S8): 118–128.
- [34] Levinthal C. How to fold graciously [J]. Mossbauer Spectroscopy in Biological Systems Proceedings, 1969, 24(41): 22–24.
- [35] Lindorff-Larsen K, Piana S, Dror R O, et al. How fast-folding proteins fold [J]. Science, 2011, 334(6055): 517–520.
- [36] Sohl J L, Jaswal S S, Agard D A. Unfolded conformations of α -lytic protease are more stable than its native state [J]. Nature, 1998, 395(6704): 817–819.
- [37] Wang Z, Mottonen J, Goldsmith E J. Kinetically controlled folding of the serpin plasminogen activator inhibitor 1 [J]. Biochemistry, 1996, 35(51): 16443–16448.
- [38] Sanchez-Ruiz J M. Protein kinetic stability [J]. Biophys. Chem., 2010, 148(1): 1–15.
- [39] Fisher K E, Ruan B, Alexander P A, et al. Mechanism of the kinetically-controlled folding reaction of subtilisin [J]. Biochemistry, 2007, 46(3): 640–651.
- [40] Schrader T E, Schreier W J, Cordes T, et al. Light-triggered β -hairpin folding and unfolding [J]. Proc. Natl. Acad. Sci. U.S.A., 2007, 104(40): 15729–15734.
- [41] Puchner E M, Alexandrovich A, Kho A L, et al. Mechanoenzymatics of titin kinase [J]. Proc. Natl. Acad. Sci. U.S.A., 2008, 105(36): 13385–13390.
- [42] Savage B, Saldivar E, Ruggeri Z M. Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor [J]. Cell, 1996, 84(2): 289–297.
- [43] Matsushita T, Sadler J E. Identification of amino acid residues essential for von Willebrand factor binding to platelet glycoprotein Ib: charged-to-alanine scanning mutagenesis of the A1 domain of human von Willebrand factor [J]. J. Biol. Chem., 1995, 270(22): 13406–13414.
- [44] Han M Z, Xu J, Ren Y, et al. Simulations of flow induced structural transition of the β -switch region of glycoprotein Ib α [J]. Biophys. Chem., 2016, 209: 9–20.
- [45] Wright P E, Dyson H J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm [J]. J. Mol. Biol., 1999, 293(2): 321–331.
- [46] Pancsa R, Tompa P. Structural disorder in eukaryotes [J]. PLoS One, 2012, 7(4): e34687(1–10).
- [47] Oldfield C J, Cheng Y, Cortese M S, et al. Comparing and combining predictors of mostly disordered proteins [J]. Biochemistry, 2005, 44(6): 1989–2000.
- [48] Uversky V N, Dunker A K. Understanding protein non-folding [J]. Biochim. Biophys. Acta, Proteins Proteomics, 2010, 1804(6): 1231–1264.
- [49] Demarest S J, Martinez-Yamout M, Chung J, et al. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators [J]. Nature, 2002, 415(6871): 549–553.
- [50] Lee C W, Martinez-Yamout M A, Dyson H J, et al. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein [J]. Biochemistry, 2010, 49(46): 9964–9971.
- [51] Dyson H J, Wright P E. Coupling of folding and binding for unstructured proteins [J]. Curr. Opin. Struct. Biol., 2002, 12(1): 54–60.
- [52] Kiefhaber T, Bachmann A, Jensen K S. Dynamics and mechanisms of coupled protein folding and binding reactions [J]. Curr. Opin. Struct. Biol., 2012, 22(1): 21–29.
- [53] Dyson H J. Roles of intrinsic disorder in protein-nucleic acid interactions [J]. Mol. Biosyst., 2012, 8(1): 97–104.
- [54] Wright P E, Dyson H J. Linking folding and binding [J]. Curr. Opin. Struct. Biol., 2009, 19(1): 31–38.
- [55] Huang Y, Liu Z R. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the “fly-casting” mechanism [J]. J. Mol. Biol., 2009, 393(5): 1143–1159.
- [56] Espinoza-Fonseca L M. Reconciling binding mechanisms of intrinsically disordered proteins [J]. Biochem. Biophys. Res. Commun., 2009, 382(3): 479–482.
- [57] Han M Z, Xu J, Ren Y, et al. Simulation of coupled folding and binding of an intrinsically disordered protein in explicit solvent with metadynamics [J]. J. Mol. Graphics Modell., 2016, 68: 114–127.
- [58] Wang J, Lu Q, Lu H P. Single-molecule dynamics reveals cooperative binding-folding in protein recognition [J]. PLoS Comput. Biol., 2006, 2(7): 842–852.
- [59] Ganguly D, Zhang W, Chen J. Synergistic folding of two intrinsically disordered proteins: searching for conformational selection [J]. Mol. Biosyst., 2012, 8(1): 198–209.
- [60] Turjanski A G, Gutkind J S, Best R B, et al. Binding-induced folding of a natively unstructured transcription factor [J]. PLoS Comput. Biol., 2008, 4(4): e1000060(1–11).
- [61] Sugase K, Dyson H J, Wright P E. Mechanism of coupled folding and binding of an intrinsically disordered protein [J]. Nature, 2007, 447(7147): 1021–1025.
- [62] Ganguly D, Chen J. Atomistic details of the disordered states of KID and pKID: implications in coupled binding and folding [J]. J.

- Am. Chem. Soc., 2009, 131(14): 5214–5223.
- [63] Zhang W H, Ganguly D, Chen J H. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins [J]. PLoS Comput. Biol., 2012, 8(1): e1002353(1–15).
- [64] Chen J H. Towards the physical basis of how intrinsic disorder mediates protein function [J]. Arch. Biochem. Biophys., 2012, 524(2): 123–131.
- [65] Rust R R, Baldissari D M, Weber D J. Structure of the negative regulatory domain of p53 bound to S100B ($\beta\beta$) [J]. Nat. Struct. Mol. Biol., 2000, 7(7): 570–574.
- [66] Hoff K G, Avalos J L, Sens K, et al. Insights into the sirutin mechanism from ternary complexes containing NAD⁺ and acetylated peptide [J]. Structure, 2006, 14(8): 1231–1240.
- [67] Lowe E D, Tews I, Cheng K Y, et al. Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A [J]. Biochemistry, 2002, 41(52): 15625–15634.
- [68] Wiewiora R P, Chen S, Beauchamp K, et al. Conformational dynamics of histone lysine methyltransferases by millisecond-timescale molecular dynamics on folding@home [J]. Biophys. J., 2017, 112(3): 189a.
- [69] Hartl F U, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein [J]. Science, 2002, 295(5561): 1852–1858.
- [70] Fenton W A, Horwich A L. Chaperonin-mediated protein folding: fate of substrate polypeptide [J]. Q. Rev. Biophys., 2003, 36(2): 229–256.
- [71] Xu Z, Horwich A L, Sigler P B. The crystal structure of the asymmetric GroEL–GroES–(ADP)₇ chaperonin complex [J]. Nature, 1997, 388(6644): 741–750.
- [72] Braig K, Otwinowski Z, Hegde R, et al. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å [J]. Nature, 1994, 371(6498): 578–586.
- [73] Braig K, Simon M, Furuya F, et al. A polypeptide bound by the chaperonin GroEL is localized within a central cavity [J]. Proc. Natl. Acad. Sci. U.S.A., 1993, 90(9): 3978–3982.
- [74] Steinbacher S, Ditzel L. Review: nucleotide binding to the thermoplasma thermosome: implications for the functional cycle of group II chaperonins [J]. J. Struct. Biol., 2001, 135(2): 147–156.
- [75] Tang Y C, Chang H C, Roeben A, et al. Structural features of the GroEL–GroES nano-cage required for rapid folding of encapsulated protein [J]. Cell, 2006, 125(5): 903–914.
- [76] van der Vaart A, Ma J, Karplus M. The unfolding action of GroEL on a protein substrate [J]. Biophys. J., 2004, 87(1): 562–573.
- [77] Takagi F, Koga N, Takada S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations [J]. Proc. Natl. Acad. Sci. U.S.A., 2003, 100(20): 11367–11372.
- [78] Ren Y, Gao J, Xu J, et al. Explicit solvent molecular dynamics simulations of chaperonin-assisted rhodanese folding [J]. Particuology, 2009, 7(3): 220–224.
- [79] 潘章, 陈静, 耿铁钊, 等. 驱动蛋白的研究进展 [J]. 生命科学, 2012, 16(4): 350–356.
- Pan Z, Chen J, Geng Y Z, et al. Progresses on kinesin superfamily [J]. Life Science Research, 2012, 16(4): 350–356.
- [80] 刘梅, 徐娜, 阮世龙, 等. 驱动蛋白及其作用研究进展 [J]. 杭州师范大学学报(自然科学版), 2013, 12(1): 40–44.
- Liu M, Xu N, Ruan S L, et al. Research progress of kinesin and its functions [J]. Journal of Hangzhou Normal University (Natural Science Edition), 2013, 12(1): 40–44.
- [81] 曹添亮, 韩孟之, 徐骥, 等. 驱动蛋白结构与运动机制 [J]. 中国生物化学与分子生物学报, 2016, 32(7): 734–744.
- Cao T L, Han M Z, Xu J, et al. Structures and mechanisms of kinesin's walking [J]. Chinese Journal of Biochemistry and Molecular Biology, 2016, 32(7): 734–744.
- [82] 曹添亮. 驱动蛋白动态结构模拟与控制机制研究 [D]. 北京: 中国科学院大学, 2016: 23–52.
- Cao T L. Molecular dynamics simulations of the dynamic structures of kinesin and analysis of the underlying mechanisms [D]. Beijing: University of Chinese Academy of Sciences, 2016: 23–52.
- [83] Freddolino P L, Arkhipov A S, FLarson S B, et al. Molecular dynamics simulations of the complete satellite tobacco mosaic virus [J]. Structure, 2006, 14(3): 437–449.
- [84] Zhao G, Perilla J R, Yufenyuy E L, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics [J]. Nature, 2013, 497(7451): 643–646.
- [85] Xu J, Wang X W, He X F, et al. Application of the Mole-8.5 supercomputer: probing the whole influenza virion at the atomic level [J]. Chin. Sci. Bull., 2011, 56(20): 2114–2118.
- [86] Rapaport D C. Self-assembly of polyhedral shells: a molecular dynamics study [J]. Phys. Rev. E, 2004, 70(5): 1539–1555.
- [87] Rapaport D C. Role of Reversibility in viral capsid growth: a paradigm for self-assembly [J]. Phys. Rev. Lett., 2008, 101(18): 186101(1–4).
- [88] Arkhipov A, Roos W H, Wuite G J, et al. Elucidating the mechanism behind irreversible deformation of viral capsids [J]. Biophys. J., 2009, 97(7): 2061–2069.
- [89] Heller H, Grubmüller H, Schulten K. Molecular dynamics simulation on a parallel computer [J]. Mol. Simul., 1990, 5(3/4): 133–165.
- [90] Brown D, Clarke J H, Okuda M, et al. A domain decomposition parallelization strategy for molecular dynamics simulations on distributed memory machines [J]. Comput. Phys. Commun., 1993, 74(1): 67–80.
- [91] Bhandarkar M, Kalé L V, de Sturler E, et al. Adaptive load balancing for MPI programs [C]//Alexandrov E V, Dongarra J J, Juliano B A, et al. Computational Science-ICCS 2001. Heidelberg: Springer, 2001: 108–117.
- [92] Hess B, Kutzner C, van der Spoel D, et al. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation [J]. J. Chem. Theory Comput., 2008, 4(3): 435–447.
- [93] Phillips J C, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD [J]. J. Comput. Chem., 2005, 26(16): 1781–1802.
- [94] Plimpton S. Fast parallel algorithms for short-range molecular dynamics [J]. J. Comput. Phys., 1995, 117(1): 1–19.
- [95] Bowers K J, Chow E, Xu H F, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters [C]//Proceedings of the ACM/IEEE Conference on Supercomputing (SC06). 2006: 43.
- [96] Larson S M, Snow C D, Shirts M, et al. Folding@home and genome@home: using distributed computing to tackle previously intractable problems in computational biology [C]//Grant E R. Computational Genomics: Horizon Press, 2002: 634–654.
- [97] Xu J, Ren Y, Ge W, et al. Molecular dynamics simulation of

- macromolecules using graphics processing unit [J]. *Mol. Simul.*, 2010, 36(14): 1131–1140.
- [98] Xu J, Ge W, Ren Y, et al. Implementation of particle-mesh ewald (PME) on graphics processing units [J]. *Chin. J. Comput. Phys.*, 2009, 27(4): 548–554.
- [99] Chen F G, Ge W, Guo L, et al. Multi-scale HPC system for multi-scalediscrete simulation—development and application of a supercomputer with 1 petaflops peak performance in single precision [J]. *Particuology*, 2009, 7(4): 332–335.
- [100] Shaw D E, Deneroff M M, Dror R O, et al. Anton, a special-purpose machine for molecular dynamics simulation [J]. *Commun. ACM*, 2008, 51(7): 91–97.
- [101] Shaw D E, Maragakis P, Lindorff-Larsen K, et al. Atomic-level characterization of the structural dynamics of proteins [J]. *Science*, 2010, 330(6002): 341–346.
- [102] Shaw D E, Grossman J P, Bank J A, et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. SC14 [C]//International Conference for High Performance Computing, Networking, Storage and Analysis. Piscataway: IEEE, 2014: 41–53.
- [103] Hu X, Hong L, Dean Smith M, et al. The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time [J]. *Nat. Phys.*, 2015, 12: 171–174.
- [104] Li J H. Exploring the logic and landscape of the knowledge system: multilevel structures, each multiscaled with complexity at the mesoscale [J]. *Engineering Science*, 2016, 2(3): 276–285.
- [105] 李静海, 胡英, 袁权. 探索介尺度科学: 从新角度审视老问题 [J]. *中国科学·B辑: 化学*, 2014, 44(3): 277–281.
- Li J H, Hu Y, Yuan Q. Mesoscience: exploring old problems from a new angle [J]. *Scientia Sinica (Chimica)*, 2014, 44(3): 277–281.
- [106] Li J H, Huang W L. From multiscale to mesoscience: addressing mesoscales in mesoregimes of different levels [J]. *Annual Review of Chemical & Biomolecular Engineering*, 2018, 9(1): 41–60.