

# 两类蛋白质粗粒化模型

史绍康<sup>1</sup>, 赵莉<sup>2</sup>, 吕中元<sup>1</sup>

(1. 吉林大学 化学学院, 长春 130012; 2. 吉林大学 生命科学学院, 长春 130012)

**摘要:** 基于在分子层面探究蛋白质折叠、组装和相分离等行为的动力学细节特征是目前该领域研究的重点和难点, 而粗粒化模型已成为应对该问题的关键策略. 综述蛋白质粗粒化模型的发展历程, 介绍两种常用的粗粒化模型, 并阐述其建模方法、势能函数以及在实际生物体系中的应用. 通过展示这些模型在复杂蛋白质体系模拟中的应用优势, 揭示粗粒化模型在显著减少计算资源消耗方面的独特价值, 及其在推进大规模蛋白质动力学过程研究中的潜力与重要意义.

**关键词:** 蛋白质; 粗粒化; 多尺度; 分子动力学模拟

**中图分类号:** O641.3 **文献标志码:** A **文章编号:** 1671-5489(2025)01-0182-09

## Coarse-Grained Models for Two Types of Proteins

SHI Shaokang<sup>1</sup>, ZHAO Li<sup>2</sup>, LÜ Zhongyuan<sup>1</sup>

(1. College of Chemistry, Jilin University, Changchun 130012, China;

2. School of Life Sciences, Jilin University, Changchun 130012, China)

**Abstract:** Exploring the dynamic detail characteristics of protein folding, assembly, and phase separation at the molecular level is currently the focus and difficulty of research in this field, and coarse-grained model have become a key strategy to address these issues. We review the development history of coarse-grained model of protein, introduce two commonly used coarse-grained models, and explain their modeling methods, potential energy functions, and applications in practical biological systems. By demonstrating the application advantages of these models in simulating complex protein systems, we review the unique value of coarse-grained model in significantly reducing computational resource consumption, as well as their potential and significance in advancing the study of large-scale protein dynamic processes.

**Keywords:** protein; coarse-grained; multiscale; molecular dynamics simulation

蛋白质在生命活动中发挥重要作用, 广泛参与催化、识别、免疫、运输、信号转导和能量储存等多种生物过程. 蛋白质由 20 种氨基酸按特定序列组合而成, 通常情况下通过折叠形成独特的三维结构, 从而实现其多样的生物功能. 目前, 研究人员主要利用 X 射线衍射、冷冻电镜和核磁共振等实验方法解析蛋白质的三维结构. 随着计算机技术的迅猛发展, 人们已实现对氨基酸序列与其空间结构间关系的精准预测. 其中, AlphaFold<sup>[1]</sup> 软件可高精度预测包括蛋白质数据库 (protein data bank, PDB) 在内

收稿日期: 2024-12-02.

**第一作者简介:** 史绍康(2002—), 男, 汉族, 博士研究生, 从事生物大分子液-液相分离的分子动力学模拟的研究, E-mail: shick24@mails.jlu.edu.cn. **通信作者简介:** 吕中元(1973—), 男, 汉族, 博士, 教授, 博士生导师, 从事大分子自组装的理论模拟和聚合物多尺度模拟方法的研究, E-mail: luzhy@jlu.edu.cn.

**基金项目:** 国家自然科学基金(批准号: 22473048).

的超过 2 亿种蛋白质结构, 极大推动了结构生物学的发展. 该重大突破不仅颠覆了传统的蛋白质结构研究方法, 其开发团队还与 Baker 共同分享了 2024 年诺贝尔化学奖.

然而, 蛋白质的生物学功能除依赖于三维结构外, 还受其动力学行为及分子间相互作用的影响, 且通过实验方法捕捉蛋白质在溶液中的动力学细节特征仍是一项巨大挑战. 近年来, 人们发现一类在溶液中缺乏固定三维结构的蛋白质, 即固有无序蛋白(intrinsically disordered proteins, IDP). 由于无法结晶, 因此这类蛋白质难以通过传统的 X 射线衍射等方法解析结构. 此外, 当固有无序蛋白通过液-液相分离(liquid-liquid phase separation, LLPS), 在细胞或溶液环境中形成蛋白质凝聚体时<sup>[2]</sup>, 其复杂的相互作用同样难以通过传统技术全面表征. 因此, 单靠实验方法无法深入揭示蛋白质凝聚体动力学行为及其结构与生物学功能之间的联系.

分子动力学(molecular dynamics, MD)模拟逐渐成为研究蛋白质动力学的重要工具. 其优势在于 MD 能在原子分辨率上捕捉蛋白质的动态变化, 为揭示蛋白质的折叠机制及动力学特性提供了理论支持<sup>[3]</sup>. 近年来, 随着计算能力的快速提升, MD 模拟在研究大分子动态行为时表现出独特优势<sup>[4]</sup>. 此外, 随着模拟算法和技术的进步, 研究人员能对更长的时间尺度进行模拟, 从而捕捉蛋白质折叠过程中的关键动态事件<sup>[5]</sup>. 这些突破使 MD 模拟成为探索蛋白质折叠与功能机制的核心工具, 为生物大分子的研究提供了全新视角和方法<sup>[6]</sup>.

近年来, 蛋白质复合体、膜蛋白以及蛋白质凝聚体等更复杂的体系逐渐成为研究重点. 这些体系通常由数十甚至上百个蛋白质分子和磷脂等成分组成. 传统的全原子分子动力学(all-atom molecular dynamics, AA-MD)模拟虽然在分辨率上具有优势, 但在处理大体系或实现长时间尺度的动力学模拟时存在局限性. 因此, 研究人员开发了一系列粗粒化(coarse-grained, CG)模型, 通过适当简化体系的细节提升计算效率, 从而能模拟研究更大尺度、更复杂的分子聚集体体系. 相比全原子模型, 粗粒化模型能覆盖数十至数百纳米的空间尺度以及毫秒级的时间尺度, 使得粗粒化分子动力学(coarse-grained molecular dynamics, CG-MD)成为研究纳米和介观尺度分子现象以及解决传统建模方法难以处理问题的关键工具<sup>[7-8]</sup>. 本文讨论两种具有代表性的蛋白质粗粒化模拟模型: Martini 模型和基于疏水标度(hydrophobic scale, HPS)的模型, 总结两种模型在不同研究领域中的典型应用, 并讨论其各自的优势和局限性, 为更高效研究复杂蛋白质体系提供新的视角和思路.

## 1 粗粒化模型

粗粒化模型构建目标是保留目标体系的关键特征, 同时显著减少需显式处理的自由度. 在粗粒化模型中, 通常将多个原子映射为一个粗粒化粒子. 根据模型分辨率的不同, 粗粒化模型可从超粗粒化级别(如用一个粒子代表整个蛋白质)到近原子级别(保留大多数化学特征)<sup>[8-9]</sup>. 在近原子级别的粗粒化模型中, 通常将空间上相近的 3~6 个比氢原子更重的原子映射为一个粗粒化粒子, 从而在简化的同时保留分子内的基本结构信息. 针对蛋白质, 大多数粗粒化模型将蛋白质的单个氨基酸残基分成两部分处理: 用 1~3 个粒子表示氨基酸的骨架, 用 1~5 个粒子表示其侧基. Levitt 等<sup>[10]</sup>因开发了蛋白质粗粒化模型, 获得了 2013 年的诺贝尔化学奖. 该模型将每个氨基酸的骨架部分和对应的侧基简化为两个粗粒化粒子, 分别定位于氨基酸骨架的  $\alpha$ -碳( $C_\alpha$ )位置和侧基的构象中心. 通过从扩展构象出发, 结合局部能量最小化和模拟热扰动, 成功获得了具有部分天然特征的蛋白质结构, 为分子模拟研究提供了重要工具.

最常用粗粒化分子动力学力场的势函数( $U$ )主要由两部分组成: 第一部分( $U_{\text{成键}}$ )描述键合粒子的相互作用, 包括成键、键角和二面角相互作用; 第二部分( $U_{\text{未成键}}$ )描述非键相互作用, 包括短程相互作用和静电相互作用, 一般使用 Lennard-Jones(LJ)势和库仑势表示:

$$U = U_{\text{成键}} + U_{\text{未成键}}, \quad (1)$$

$$U_{\text{成键}} = \frac{1}{2} \sum_{\text{键}} k_r (r - r_0)^2 + \frac{1}{2} \sum_{\text{角}} k_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{\text{二面角}} k_\varphi [1 + \cos(n\varphi - \delta)], \quad (2)$$

$$U_{\text{未成键}} = U_{\text{LJ}} + U_{\text{elec}}, \quad (3)$$

$$U_{LJ} = \sum_{i,j} 4 \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (4)$$

$$U_{elec} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (5)$$

势函数中一般包含较多参数, 确定这些参数数值的过程称为参数化. 参数化通常采用两种策略: 1) 自上而下的方法, 即通过重现实验测得的热力学或力学性质对模型进行参数化. 这种方法旨在捕捉模拟体系的整体表现. 2) 自下而上的方法, 即以再现分辨率更高的模型(如全原子模型)的结构特性为主要目标. 这种方法注重再现分子间的结构细节和局部相互作用特性<sup>[11]</sup>. 此外, 许多粗粒化模型结合了自上而下和自下而上的参数化策略. 如成键相互作用的参数通常基于微观结构特性确定, 非键相互作用通过实验数据进行优化. 这种混合方法可较好地平衡准确性和通用性. 由于自下而上的方法能捕捉到精细的相互作用细节, 通常更准确. 而自上而下及混合方法在定义复杂体系(如包含蛋白质的多组分体系)的参数时更具通用性, 因而被广泛应用. 近年来, 基于人工智能(artificial intelligence, AI)方法在构建蛋白质粗粒化模型中的作用日益突出, 即使在温度或 pH 值变化等复杂条件下, AI 方法也可利用结构和热力学数据得到粗粒化力场的参数<sup>[12]</sup>.

下面介绍 Martini 蛋白质粗粒化模型和基于疏水标度(HPS)的蛋白质粗粒化模型(图 1).

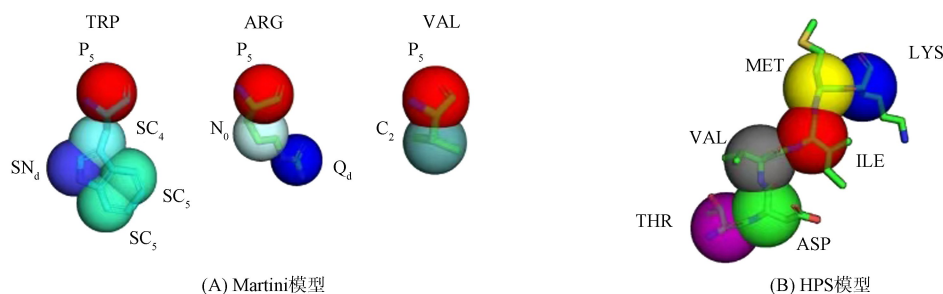


图 1 Martini 粗粒化模型和 HPS 粗粒化模型的映射方式

Fig. 1 Mapping strategy of Martini coarse-grained model and HPS coarse-grained model

### 1.1 Martini 粗粒化模型

Martini 粗粒化模型的核心假设是将小分子片段简化为粗粒化粒子, 这些粒子在疏水性、相互作用模式及尺寸等方面遵循特定规律. 通过特定的组合方式, 这些粒子可组装成各种分子<sup>[13]</sup>. Martini 粗粒化模型最初专为模拟脂质体系设计<sup>[14]</sup>, 之后经不断优化, 已扩展应用至更广泛的生物分子体系, 包括碳水化合物<sup>[15]</sup>、固醇<sup>[16]</sup>、核酸<sup>[17]</sup>和蛋白质<sup>[18-19]</sup>等, 并且进一步拓展应用到材料科学领域<sup>[20]</sup>. Martini 粗粒化力场主要应用在 GROMACS<sup>[21]</sup> 软件中, 但也支持其他分子动力学模拟软件, 如 NAMD<sup>[22]</sup>, LAMMPS<sup>[23]</sup> 和 OpenMM<sup>[24]</sup> 等软件. Martini 粗粒化模型有一个活跃的开发社区, 提供多种辅助工具, 使研究人员的工作效率显著提升. 这些工具包括构建拓扑文件<sup>[25]</sup>、生成初始配置<sup>[26]</sup>及粗粒化模型与全原子模型之间的转换等<sup>[27]</sup>, 进一步丰富了 Martini 粗粒化模型在科学研究中的应用潜力.

Martini 蛋白质粗粒化模型于 2008 年首次推出<sup>[18]</sup>, 并在 2013 年进行了优化升级<sup>[19]</sup>. 在该模型中, 蛋白质主链的粗粒化粒子用位于氨基酸骨架质心的单个粒子表示, 氨基酸的侧基根据其化学性质由 1~5 个粒子表示(图 1(A)). 为稳定蛋白质的二级结构, Martini 粗粒化模型通过调整主链的二级结构键参数实现约束; 为维持三级结构, 需引入基于结构涨落的控制方法, 如弹性网络模型(elastic network model, ENM)和 Gö 模型<sup>[28]</sup>. 弹性网络模型通过在主链粒子之间施加弹簧势, 并限制在一定截距范围内的相互作用, 维持蛋白质三级结构的稳定性. 由于依赖大量弹簧势, ENM 在模拟中可能会抑制蛋白质应有的构象变化. Gö 模型通过使用 Lennard-Jones 势替代弹簧势, 并基于接触图(contact map)选择相互作用粒子对, 从而克服了 ENM 采用大量弹簧势的限制, 可对未折叠状态的蛋白质构象采样<sup>[29]</sup>. Martini 模型构建可通过专用的 Martinize2 程序完成, 该程序支持从初始全原子结构生成粗粒化模型及相应的拓扑文件<sup>[25]</sup>. 该流程简化了粗粒化建模操作, 使研究人员能高效构建适用

于复杂蛋白质体系的模拟模型.

Martini 是一个完全开源的粗粒化模型, 其官方网站(<http://cgmartini.nl>)汇集了论坛、工具、力场文件以及与该模型相关的最新研究报道. 此外, Martini 数据库(<https://mad.ibcp.fr/>)为用户提供了丰富的拓扑文件和参数资源. Martini 力场计划的 GitHub 页面(<https://github.com/Martini-Force-Field-Initiative>)致力于构建一个详细的 Martini 拓扑存储库, 进一步促进了该模型的共享和应用.

## 1.2 基于 HPS 的蛋白质粗粒化模型

通常 Martini 粗粒化模型采用“四对一”的粗粒化映射策略, 即将 4 个 C, N, O 等原子映射为一个粗粒化粒子. 这种方法虽有效简化了模拟体系, 但在研究大量固有无序蛋白参与的 LLPS 过程中受体系尺度限制. 基于 HPS 的粗粒化模型可较好解决该问题: HPS 模型开发的宗旨是在考虑蛋白质序列的同时, 实现对更大体系和更长时间的高效模拟. HPS 蛋白质粗粒化模型将每个氨基酸残基表示为一个粗粒化粒子, 忽略了氨基酸内部的结构细节(图 1(B)). 该映射方案大幅度简化了粗粒化模拟的复杂度, 同时保留了蛋白质序列相关信息, 这对捕获序列特异性相互作用和行为至关重要.

HPS 蛋白质粗粒化模型的理论基础源于 Kapcha 等<sup>[30]</sup>提出的残基级别疏水标度模型. 该模型根据全原子力场中原子电荷分布评估原子的疏水性, 并通过累加得到氨基酸残基的整体疏水标度值. 在此基础上, Dignon 等<sup>[31]</sup>进一步开发了 HPS 粗粒化力场, 引入基于疏水标度的短程相互作用参数  $\lambda_{ij}^{\text{KR}}$ , 并采用 Debye-Hückel 势描述静电相互作用:

$$U_{\text{未成键}} = U_{\text{AH}} + U_{\text{elec}}, \quad (6)$$

$$U_{\text{AH}} = \sum_{i,j} \begin{cases} U_{\text{LJ}} + \epsilon_{\text{AH}}(1 - \lambda_{ij}^{\text{KR}}), & r \leq 2^{\frac{1}{6}}\sigma, \\ \lambda_{ij}^{\text{KR}} U_{\text{LJ}}, & r > 2^{\frac{1}{6}}\sigma, \end{cases} \quad (7)$$

$$U_{\text{elec}} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} e^{-r/D}, \quad (8)$$

其中  $\lambda_{ij}^{\text{KR}}$  是氨基酸  $i$  和  $j$  疏水标度的算术平均值,  $\epsilon_{\text{AH}}$  用于拟合相互作用能量. 这种方法不仅能有效再现 IDP 在溶液中的行为, 还为研究序列特异性驱动的相分离现象提供了重要工具支持. 为进一步提升 HPS 力场对固有无序蛋白的描述能力, 在 HPS-Urry 力场中引入两个自由参数  $\mu$  和  $\Delta$  微调疏水参数<sup>[32]</sup>:

$$\lambda_{ij}^{\text{Urry}} = \mu\lambda_{ij}^{\text{KR}} - \Delta. \quad (9)$$

HPS 和 HPS-Urry 是最早用于描述 IDP 形成凝聚体的蛋白质粗粒化模型, 人们在这两个粗粒化模型的基础上进行了多项改进.

Dannenhoffer-Lafage 等<sup>[33]</sup>利用机器学习算法对 HPS 粗粒化力场的疏水标度进行了优化, 开发了 FB-HPS 粗粒化力场, 从而能更准确描述疏水相互作用. 在此基础上, Das 等<sup>[34]</sup>提出了 HPS+阳离子- $\pi$  粗粒化力场, 通过显式考虑阳离子氨基酸(如 Arg 和 Lys)与芳香族残基之间的阳离子- $\pi$  相互作用, 进一步提升了对某些蛋白质相互作用的精细刻画能力. Tesei 等<sup>[35]</sup>通过 Bayes 参数学习方法优化了疏水标度, 开发了 M1-3 蛋白质粗粒化模型, 该方法有效消除了早期疏水参数分布中的偏差, 为 IDP 的粗粒化分子动力学模拟提供了更准确且可靠的力场参数, 该优化方法对疏水参数的初始值无依赖性. 因此, 即使在缺乏具体 IDP 特性信息的情况下, M1 力场仍能给出可靠的模拟结果, 因而 M1 粗粒化力场成为研究 IDP 动力学行为的重要工具. Cao 等<sup>[36]</sup>在此基础上进一步优化, 将训练集中加入具有多个结构域的蛋白, 同时在模拟过程中引入了弹性网络模型(ENM). 通过这一改进, 他们给出了新的基于 HPS 模型的氨基酸疏水参数, 发展了 CAVADOS3 蛋白质粗粒化模型. 该模型显著提高了对复杂蛋白质体系的描述能力, 为研究大量 IDP 的聚集行为提供了更高效且精确的工具.

在上述 HPS 系列蛋白质粗粒化模型中, 氨基酸残基之间的短程相互作用通常由 Ashbaugh-Hatch 势(式(7))描述. Joseph 等<sup>[37]</sup>通过以 150 mmol/L NaCl 浓度下的平均力势(potential-of-mean-force, PMF)为目标函数对势函数进行参数化, 结合生物信息学数据开发了 Mpipi 力场. 在 Mpipi 力场中, 采用 Wang-Frenkel 势作为相互作用函数, 用以更准确描述氨基酸残基间的相互作用:

$$U_{WF} = \epsilon_{ij} \alpha_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{2\nu_{ij}} - 1 \right] \left[ \left( \frac{R_{ij}}{r} \right)^{2\mu_{ij}} - 1 \right]^{2\nu_{ij}}, \quad (10)$$

其中  $\epsilon_{ij}$  用于拟合相互作用能量,  $\alpha_{ij}$  用于拟合 PMF 曲线,  $\mu_{ij}$  和  $\nu_{ij}$  为参数,  $R_{ij}$  为截止距离. 该模型强调  $\pi$ - $\pi$  相互作用和阳离子- $\pi$  相互作用的贡献, 对于富含芳香族残基和正电荷氨基酸的 IDP 提供了很好的粗粒化力场支持. Garaizar 等<sup>[38]</sup> 利用该模型研究了单组分蛋白质凝聚体中多相结构的形成过程, 深入解析了这些蛋白质凝聚体的复杂相行为及其内部结构排列. 通过模拟, 该模型揭示了蛋白质凝聚体在细胞形成过程中潜在的生物学功能和作用机制, 为理解 IDP 驱动的 LLPS 提供了重要理论依据.

## 2 粗粒化模型的应用

### 2.1 Martini 蛋白质粗粒化模型的应用

Martini 蛋白质粗粒化模型主要用于研究蛋白质和生物膜之间的相互作用. 这类研究对蛋白质构象变化的需求较低, 避免了 Martini 粗粒化模型的短板. Martini 粗粒化模型已广泛用于探索蛋白质-脂质的特异性相互作用, 特别是蛋白质插入复杂生物膜结构时周围脂质环境的“指纹”特征<sup>[39]</sup>.

#### 2.1.1 膜蛋白的研究

借助 Martini 力场, 人们深入研究了多种膜蛋白的功能机制, 包括 G 蛋白偶联受体 (G protein-coupled receptors, GPCR)<sup>[40]</sup>、多孔蛋白和跨膜通道<sup>[41]</sup> 以及外周膜蛋白<sup>[42]</sup> 等. Valério 等<sup>[43]</sup> 研究了抗菌肽和病毒融合肽诱导的机制及其对膜生物物理特性的影响. Li 等<sup>[44]</sup> 使用 Martini 模型探索了蛋白质插入酶的脂质扰动特性, 证实了 Martini 模型在评估脂质扰动特性方面的强大能力. 此外, Martini 模型还被用来研究阿尔茨海默症相关淀粉样蛋白在膜环境中的寡聚化过程以及特定脂质的相互作用和分布<sup>[45]</sup>.

#### 2.1.2 多尺度模拟研究进展

Martini 力场能支持更大的空间尺度和更长时间的模拟, 为多尺度模拟提供了新的可能性. Zhang 等<sup>[46]</sup> 使用 Martini 模型模拟研究了一个大型的 900 000 多组分蛋白酶体-纳米孔. Mosalaganti 等<sup>[47]</sup> 对一个 120 000 000 的人类核孔复合物进行了超过 1  $\mu$ s 的模拟. 此外, Martini 粗粒化模型还被用于模拟 SARS-CoV 和 SARS-CoV-2 的包膜结构<sup>[48]</sup>.

#### 2.1.3 在 IDP 与生物大分子凝聚体研究中的应用

Martini 粗粒化模型在 IDP 研究中的应用有一定的局限性. Larsen 等<sup>[49]</sup> 发现其对 IDP 的整体尺寸存在低估, 但通过增强蛋白质-水相互作用强度, 模型给出的构象集合与实验数据一致. 基于此优化, Martini 模型已成功用于研究人类生长激素受体无序区域的动力学<sup>[50]</sup>、RNA 结合蛋白 hnRNPA1 的结构域和固有无序区域 (intrinsically disordered region, IDR) 之间的相互作用<sup>[51]</sup> 以及 IDP 与脂质双层的相互作用<sup>[52]</sup>.

Tsanai 等<sup>[53]</sup> 利用 Martini 粗粒化模型描述了聚赖氨酸/聚谷氨酸体系的盐依赖性凝聚现象, 并观察到 RNA 分子分配到凝聚体中. Liu 等<sup>[54]</sup> 对脂质囊泡中的聚赖氨酸/聚天冬氨酸和聚赖氨酸/聚谷氨酸体系进行了模拟, 发现凝聚体形成会影响囊泡的形状. 这些研究展示了 Martini 粗粒化模型在揭示驱动生物分子凝聚体形成的物理机制方面的潜力.

#### 2.1.4 模拟复杂多组分体系

Martini 的应用潜力在描述复杂多组分体系时得到了充分展示, 且目前已成功应用于构建完整细胞的粗粒化模型中. 如使用 Martini 粗粒化模型构建了最小细胞 JCVI-syn3A 模型, 其中包含超过 60 000 种可溶性蛋白和 2 200 种膜蛋白<sup>[55]</sup>. 然而, 现有分子动力学模拟软件受计算能力的限制, 尚无法实现全细胞尺度的粗粒化模拟. 尽管仍面临挑战, 但该例子仍展示了全细胞级别动力学模拟的未来发展方向, 为生物分子行为的多尺度解析提供了重要依据.

### 2.2 HPS 蛋白质粗粒化模型的应用

HPS 蛋白质粗粒化模型可预测 IDP 发生 LLPS 的临界温度, 以及浓相和稀相中的蛋白质浓度, 且其结果与 Flory-Huggins 理论中关于高分子的相分离理论描述一致. 人们结合大量实验数据与分子动力学模拟结果对模型进行了优化<sup>[32-37]</sup>. 目前, HPS 模型可定量预测蛋白质的饱和浓度, 并解释了温度

和盐浓度对 IDP 相分离体系的影响, 已广泛应用于蛋白质的 LLPS 相关研究中。

### 2.2.1 蛋白质液-液相分离机制的研究

液-液相分离是活细胞中生物大分子凝聚体形成的重要基础, 因此揭示其背后的驱动力对理解生物功能及失调机制至关重要. Murthy 等<sup>[56]</sup>结合 NMR、Raman 光谱和动力学模拟, 研究了 FUS 蛋白中 IDR 部分的聚集, 发现氨基酸之间的异质相互作用是 FUS 蛋白液-液相分离的基础. 研究表明, FUS 蛋白在凝聚体中保留了构象异质性<sup>[56]</sup>. Krainer 等<sup>[57]</sup>通过研究多种 IDP 在不同盐浓度下的相分离行为, 发现低盐浓度下形成凝聚体的蛋白质在中等盐浓度下会重新溶解, 并可在高盐浓度 ( $>1.5$  mol/L NaCl) 下重新进入相分离状态. 模拟结果表明, 这种转变由疏水和非静电相互作用(如 Ala-Ala, Pro-Pro, Tyr-Tyr, Ser-Ser, Arg-Tyr, Arg-Arg 之间的相互作用, Arg 在高盐浓度下变得疏水并转变为  $\pi$ - $\pi$  相互作用)驱动, 即这些相互作用是液-液相分离过程中的主要驱动力. 此外, HPS 粗粒化模型还被用于探索 IDP 凝聚体的动力学行为, 为理解生物凝聚体的形成和功能提供了重要线索.

### 2.2.2 基于 HPS 的模型拓展

在 HPS 蛋白质粗粒化模型的基础上, 人们进一步开发了包括核酸的粗粒化模型, 使其能描述蛋白质和 RNA 以及蛋白质和 DNA 的相互作用. Tejedor 等<sup>[58]</sup>研究了几种 RNA 结合蛋白(如 FUS, hnRNPA1 和 TDP-43)单独或与 RNA 共存时的相行为, 发现 RNA 会引发双重效应: 当 RNA 的回转半径( $R_g$ )和蛋白质相近时, RNA 在低浓度下增强蛋白质的相分离, 在高浓度下抑制其相分离能力. 此外, 短链 RNA 在高浓度下显著降低凝聚体的黏度, 长链 RNA 会增加凝聚体黏度. Lebold 等<sup>[59]</sup>模拟研究了 DNA 和组蛋白 H1 的多阳离子 C 端形成的凝聚体, 发现蛋白质中带电残基的分布会调节相分离行为. 研究表明, 这些序列特性与某些天然蛋白质(如精子细胞中的鱼精蛋白)可使 DNA 通过平行堆积形成非常紧凑的结构密切相关<sup>[59]</sup>. 这些研究不仅扩展了 HPS 模型在蛋白质-核酸体系中的应用范围, 还揭示了 RNA 和 DNA 在调控生物分子凝聚体形成中的关键作用, 推动了液-液相分离领域的深入发展.

## 3 讨 论

粗粒化模型作为一种重要策略, 通过简化蛋白质体系的复杂性, 显著降低计算资源的需求, 使人们能深入探索复杂蛋白质体系的聚集结构与动力学特性. 本文主要介绍了两种常用的蛋白质粗粒化模型: Martini 模型和基于疏水标度的 HPS 模型, 展示了它们的建模方法、势能函数以及在实际体系中的应用.

粗粒化模型通过将多个原子映射为一个粗粒化粒子, 最大限度地保留了目标体系的关键特征, 同时大幅度减少体系自由度. 这种粗粒化方法在蛋白质-脂质相互作用以及 IDP 相分离等研究领域展现了巨大的应用潜力. Martini 模型因其广泛的参数化数据库和跨平台的适用性, 已成为研究蛋白质和生物膜相互作用的主流工具. HPS 模型以其高效的粗粒化方案、简单的势能函数形式以及对蛋白质序列信息的保留, 成为研究 IDP 相分离行为的有效工具, 为深入理解生物大分子凝聚体的形成机制提供了帮助.

尽管粗粒化模型在蛋白质体系模拟研究中已取得了显著进展, 但仍存在一定的局限性: Martini 模型在描述蛋白质的构象变化和折叠动力学方面表现不足, 限制了其在特定场景下的应用; HPS 模型对特定蛋白质行为的预测精度, 尤其是在捕捉复杂序列效应和特殊相互作用时尚需进一步优化; 有必要通过引入更多高质量的实验数据和先进算法, 进一步提升粗粒化模型的物理化学描述准确性, 同时结合机器学习算法, 对粗粒化力场的参数化过程进行自动化优化, 有效增强模型的预测能力和适用范围. 此外, 还需开发更具普适性的粗粒化模型, 以适应对复杂多组分生物体系(如细胞膜、大型蛋白质复合物以及核酸-蛋白质相互作用等体系)的研究需求. 这些方面的进展将为粗粒化模型在生物学领域的应用提供更好的理论依据.

随着计算能力的不断提高以及算法的持续完善, 粗粒化模型在未来的生物分子研究中有望发挥更关键的作用, 为揭示生命过程的分子机制提供更强有力的支持.

## 参 考 文 献

- [1] JUMPER J, EVANS R, PRITZEL A, et al. Highly Accurate Protein Structure Prediction with AlphaFold [J]. *Nature*, 2021, 596: 583-589.
- [2] BANANI S F, LEE H O, HYMAN A A, et al. Biomolecular Condensates: Organizers of Cellular Biochemistry [J]. *Nature Reviews Molecular Cell Biology*, 2017, 18(5): 285-298.
- [3] DILL K A, MACCALLUM J L. The Protein-Folding Problem, 50 Years on [J]. *Science*, 2012, 338: 1042-1046.
- [4] BENSON N C, DAGGETT V. A Comparison of Multiscale Methods for the Analysis of Molecular Dynamics Simulations [J]. *The Journal of Physical Chemistry B*, 2012, 116: 8722-8731.
- [5] SAHOO A, XU H C, MATYSIAK S. Pathways of Amyloid-Beta Absorption and Aggregation in a Membranous Environment [J]. *Physical Chemistry Chemical Physics*, 2019, 21: 8559-8568.
- [6] WHITFORD P C, ONUCHIC J N. What Protein Folding Teaches Us about Biological Function and Molecular Machines [J]. *Current Opinion in Structural Biology*, 2015, 30: 57-62.
- [7] INGÓLFSSON H I, LOPEZ C A, UUSITALO J J, et al. The Power of Coarse Graining in Biomolecular Simulations [J]. *WIREs Computational Molecular Science*, 2014, 4(3): 225-248.
- [8] MARRINK S J, CORRADI V, SOUZA P C T, et al. Computational Modeling of Realistic Cell Membranes [J]. *Chemical Reviews*, 2019, 119(9): 6184-6226.
- [9] PAK A J, VOTH G A. Advances in Coarse-Grained Modeling of Macromolecular Complexes [J]. *Current Opinion in Structural Biology*, 2018, 52: 119-126.
- [10] LEVITT M, WARSHEL A. Computer Simulation of Protein Folding [J]. *Nature*, 1975, 253: 694-698.
- [11] JIN J, PAK A J, DURUMERIC A E P, et al. Bottom-Up Coarse-Graining: Principles and Perspectives [J]. *Journal of Chemical Theory and Computation*, 2022, 18(10): 5759-5791.
- [12] MAJEWSKI M, PÉREZ A, THÖLKE P, et al. Machine Learning Coarse-Grained Potentials of Protein Thermodynamics [J]. *Nature Communications*, 2023, 14(1): 5739-1-5739-13.
- [13] MARRINK S J, MONTICELLI L, MELO M N, et al. Two Decades of Martini: Better Beads, Broader Scope [J]. *WIREs Computational Molecular Science*, 2023, 13(1): e1620-1-e1620-42.
- [14] MARRINK S J, DE VRIES A H, MARK A E. Coarse Grained Model for Semiquantitative Lipid Simulations [J]. *The Journal of Physical Chemistry B*, 2004, 108(2): 750-760.
- [15] GRÜNEWALD F, PUNT M H, JEFFERYS E E, et al. Martini 3 Coarse-Grained Force Field for Carbohydrates [J]. *Journal of Chemical Theory and Computation*, 2022, 18(12): 7555-7569.
- [16] BORGES-ARAÚJO L, BORGES-ARAÚJO A C, OZTURK T N, et al. Martini 3 Coarse-Grained Force Field for Cholesterol [J]. *Journal of Chemical Theory and Computation*, 2023, 19(20): 7387-7404.
- [17] UUSITALO J J, INGÓLFSSON H I, MARRINK S J, et al. Martini Coarse-Grained Force Field: Extension to RNA [J]. *Biophysical Journal*, 2017, 113(2): 246-256.
- [18] MONTICELLI L, KANDASAMY S K, PERIOLE X, et al. The Martini Coarse-Grained Force Field: Extension to Proteins [J]. *Journal of Chemical Theory and Computation*, 2008, 4(5): 819-834.
- [19] DE JONG D H, SINGH G, BENNETT W F D, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field [J]. *Journal of Chemical Theory and Computation*, 2013, 9(1): 687-697.
- [20] ALESSANDRI R, GRÜNEWALD F, MARRINK S J. The Martini Model in Materials Science [J]. *Advanced Materials*, 2021, 33(24): 2008635-1-2008635-20.
- [21] ABRAHAM M J, MURTOLA T, SCHULZ R, et al. GROMACS: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers [J]. *SoftwareX*, 2015, 1/2: 19-25.
- [22] PHILLIPS J C, HARDY D J, MAIA J D C, et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD [J]. *The Journal of Chemical Physics*, 2020, 153(4): 044130-1-044130-33.
- [23] JEWETT A I, STELTER D, LAMBERT J, et al. Moltemplate: A Tool for Coarse-Grained Modeling of Complex Biological Matter and Soft Condensed Matter Physics [J]. *Journal of Molecular Biology*, 2021, 433(11): 166841-1-166841-9.

- [24] MACCALLUM J L, HU S N, LENZ S, et al. An Implementation of the Martini Coarse-Grained Force Field in OpenMM [J]. *Biophysical Journal*, 2023, 122(14): 2864-2870.
- [25] KROON P C, GRUNEWALD F, BARNOUD J, et al. Martinize2 and Vermouth: Unified Framework for Topology Generation [J]. *eLife*, 2024, 12: RP90627-1-RP90627-39.
- [26] HSU P C, BRUININKS B M H, JEFFERIES D, et al. Charmm-GUI Martini Maker for Modeling and Simulation of Complex Bacterial Membranes with Lipopolysaccharides [J]. *Journal of Computational Chemistry*, 2017, 38(27): 2354-2363.
- [27] VICKERY O N, STANSFELD P J. CG2AT2: An Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations [J]. *Journal of Chemical Theory and Computation*, 2021, 17(10): 6472-6482.
- [28] POMA A B, CIEPLAK M, THEODORAKIS P E. Combining the Martini and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins [J]. *Journal of Chemical Theory and Computation*, 2017, 13(3): 1366-1374.
- [29] WOŁEK K, GÓMEZ-SICILIA À, CIEPLAK M. Determination of Contact Maps in Proteins: A Combination of Structural and Chemical Approaches [J]. *The Journal of Chemical Physics*, 2015, 143(24): 243105-1-243105-14.
- [30] KAPCHA L H, ROSSKY P J. A Simple Atomic-Level Hydrophobicity Scale Reveals Protein Interfacial Structure [J]. *Journal of Molecular Biology*, 2014, 426(2): 484-498.
- [31] DIGNON G L, ZHENG W W, KIM Y C, et al. Sequence Determinants of Protein Phase Behavior from a Coarse-Grained Model [J]. *PLOS Computational Biology*, 2018, 14(1): e1005941-1-e1005941-23.
- [32] REGY R M, THOMPSON J, KIM Y C, et al. Improved Coarse-Grained Model for Studying Sequence Dependent Phase Separation of Disordered Proteins [J]. *Protein Science*, 2021, 30(7): 1371-1379.
- [33] DANNENHOFFER-LAFAGE T, BEST R B. A Data-Driven Hydrophobicity Scale for Predicting Liquid-Liquid Phase Separation of Proteins [J]. *The Journal of Physical Chemistry B*, 2021, 125(16): 4046-4056.
- [34] DAS S, LIN Y H, VERNON R M, et al. Comparative Roles of Charge,  $\pi$ , and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins [J]. *Proceedings of the National Academy of Sciences*, 2020, 117(46): 28795-28805.
- [35] TESEI G, SCHULZE T K, CREHUET R, et al. Accurate Model of Liquid-Liquid Phase Behavior of Intrinsically Disordered Proteins from Optimization of Single-Chain Properties [J]. *Proceedings of the National Academy of Sciences*, 2021, 118(44): e2111696118-1-e2111696118-10.
- [36] CAO F, VON BÜLOW S, TESEI G, et al. A Coarse-Grained Model for Disordered and Multi-domain Proteins [J]. *Protein Science*, 2024, 33(11): e5172-1-e5172-14.
- [37] JOSEPH J A, REINHARDT A, AGUIRRE A, et al. Physics-Driven Coarse-Grained Model for Biomolecular Phase Separation with Near-Quantitative Accuracy [J]. *Nature Computational Science*, 2021, 1(11): 732-743.
- [38] GARAIZAR A, ESPINOSA J R, JOSEPH J A, et al. Aging Can Transform Single-Component Protein Condensates into Multiphase Architectures [J]. *Proceedings of the National Academy of Sciences*, 2022, 119(26): e2119800119-1-e2119800119-11.
- [39] CORRADI V, MENDEZ-VILLUENDAS E, INGÓLFSSON H I, et al. Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins [J]. *ACS Central Science*, 2018, 4(6): 709-717.
- [40] KJØLBYE L R, SØRENSEN L, YAN J, et al. Lipid Modulation of a Class B GPCR: Elucidating the Modulatory Role of Pi(4,5)P2 Lipids [J]. *Journal of Chemical Information and Modeling*, 2022, 62(24): 6788-6802.
- [41] BORGES-ARAÚJO L, SOUZA P C T, FERNANDES F, et al. Improved Parameterization of Phosphatidylinositol Lipid Headgroups for the Martini 3 Coarse-Grain Force Field [J]. *Journal of Chemical Theory and Computation*, 2022, 18(1): 353-373.
- [42] THALLMAIR V, SCHULTZ L, ZHAO W C, et al. Two Cooperative Binding Sites Sensitize Pi(4,5)P2 Recognition by the Tubby Domain [J]. *Science Advances*, 2022, 8(36): eabp9471-1-eabp9471-12.
- [43] VALÉRIO M, MENDONÇA D A, MORAIS J, et al. Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structures [J]. *ACS Chemical Biology*, 2022, 17(7): 1831-1843.
- [44] LI D Z, ROCHA-ROA C, SCHILLING M A, et al. Lipid Scrambling Is a General Feature of Protein Insertases [J]. *Proceedings of the National Academy of Sciences*, 2024, 121(17): e2319476121-1-e2319476121-11.

- [45] CHENG S Y, CAO Y Y, ROUZBEHANI M, et al. Coarse-Grained Md Simulations Reveal Beta-Amyloid Fibrils of Various Sizes Bind to Interfacial Liquid-Ordered and Liquid-Disordered Regions in Phase Separated Lipid Rafts with Diverse Membrane-Bound Conformational States [J]. *Biophysical Chemistry*, 2020, 260: 106355-1-106355-18.
- [46] ZHANG S L, HUANG G, VERSLOOT R C A, et al. Bottom-Up Fabrication of a Proteasome-Nanopore That Unravels and Processes Single Proteins [J]. *Nature Chemistry*, 2021, 13(12): 1192-1199.
- [47] MOSALAGANTI S, OBARSKA-KOSINSKA A, SIGGEL M, et al. AI-Based Structure Prediction Empowers Integrative Structural Analysis of Human Nuclear Pores [J]. *Science*, 2022, 376: eabm9506-1-eabm9506-14.
- [48] WANG B B, ZHONG C Q, TIELEMAN D P. Supramolecular Organization of Sars-Cov and Sars-Cov-2 Virions Revealed by Coarse-Grained Models of Intact Virus Envelopes [J]. *Journal of Chemical Information and Modeling*, 2022, 62(1): 176-186.
- [49] LARSEN A H, WANG Y, BOTTARO S, et al. Combining Molecular Dynamics Simulations with Small-Angle X-Ray and Neutron Scattering Data to Study Multi-domain Proteins in Solution [J]. *PLOS Computational Biology*, 2020, 16(4): e1007870-1-e1007870-29.
- [50] KASSEM N, ARAYA-SECCHI R, BUGGE K, et al. Order and Disorder: An Integrative Structure of the Full-Length Human Growth Hormone Receptor [J]. *Science Advances*, 2021, 7(27): eabh3805-1-eabh3805-19.
- [51] MARTIN E W, THOMASEN F E, MILKOVIC N M, et al. Interplay of Folded Domains and the Disordered Low-Complexity Domain in Mediating Hnrnpa1 Phase Separation [J]. *Nucleic Acids Research*, 2021, 49(5): 2931-2945.
- [52] THOMASEN F E, SKAALUM T, KUMAR A, et al. Rescaling Protein-Protein Interactions Improves Martini 3 for Flexible Proteins in Solution [J]. *Nature Communications*, 2024, 15: 6645-1-6645-17.
- [53] TSANAI M, FREDERIX P W J M, SCHROER C F E, et al. Coacervate Formation Studied by Explicit Solvent Coarse-Grain Molecular Dynamics with the Martini Model [J]. *Chemical Science*, 2021, 12(24): 8521-8530.
- [54] LIU Y, WANG X Y, WAN Z L, et al. Capturing Coacervate Formation and Protein Partition by Molecular Dynamics Simulation [J]. *Chemical Science*, 2023, 14(5): 1168-1175.
- [55] STEVENS J A, GRÜNEWALD F, VAN TILBURG P A M, et al. Molecular Dynamics Simulation of an Entire Cell [J]. *Frontiers in Chemistry*, 2023, 11: 1106495-1-1106495-9.
- [56] MURTHY A C, DIGNON G L, KAN Y, et al. Molecular Interactions Underlying Liquid-Liquid Phase Separation of the Fus Low-Complexity Domain [J]. *Nature Structural and Molecular Biology*, 2019, 26(7): 637-648.
- [57] KRAINER G, WELSH T J, JOSEPH J A, et al. Reentrant Liquid Condensate Phase of Proteins Is Stabilized by Hydrophobic and Non-ionic Interactions [J]. *Nature Communications*, 2021, 12(1): 1085-1-1085-14.
- [58] TEJEDOR A R, GARAIZAR A, RAMÍREZ J, et al. RNA Modulation of Transport Properties and Stability in Phase-Separated Condensates [J]. *Biophysical Journal*, 2021, 120(23): 5169-5186.
- [59] LEBOLD K M, BEST R B. Tuning Formation of Protein-DNA Coacervates by Sequence and Environment [J]. *The Journal of Physical Chemistry B*, 2022, 126: 2407-2419.

(责任编辑:单凝)