

综述

蛋白质“液-液相分离”的理论和计算方法进展*

张鹏程¹⁾²⁾ 方文玉¹⁾²⁾ 鲍磊¹⁾²⁾ 康文斌^{1)2)†}

1) (湖北医药学院, 公共卫生与管理学院, 十堰 442000)

2) (湖北省南水北调水源区生物医药研发检测共享平台, 十堰 442000)

(2020年3月24日收到; 2020年4月10日收到修改稿)

蛋白质分子的“液-液相分离”是近几年生物物理学领域迅速发展起来的研究热点. 蛋白质相分离在一系列生物学过程中发挥着重要的作用. 蛋白质分子序列和构象的多样性和复杂性, 给蛋白质分子的理论研究、计算机模拟和实验研究都带来了巨大的挑战. 当前, 多尺度理论模型和多分辨率计算方法被广泛地用于蛋白质分子的“液-液相分离”的研究中. 本文将对蛋白质分子“液-液相分离”的理论基础和计算机模拟方法进行简要的综述, 并对这些理论和方法未来的发展趋势进行了初步的探讨和展望. 期望为进一步研究蛋白质“液-液相分离”的物理化学机制和过程提供理论基础和方法借鉴.

关键词: “液-液相分离”, 内禀无序蛋白质, 多价相互作用, 多尺度模拟

PACS: 87.15. A-, 87.15.Cc, 87.18.-h, 47.11.St

DOI: 10.7498/aps.69.20200438

1 引言

近些年, 由多价相互作用驱使的“液-液相分离”(liquid-liquid phase separation, LLPS) 而形成的无膜区室在很多生物过程中发挥着重要的作用^[1-7]. 虽然“液-液相分离”的概念已被用于许多物理化学过程^[8], 但是在生物系统中“液-液相分离”仍是一个新兴领域^[9]. 细胞利用相分离来完成许多不同的功能, 这些功能不一定是由脂质分子组成的细胞膜完成^[10], 也可能由特定的生物大分子通过相互作用形成的无膜区室完成^[11]. 这种相分离产生了高度浓缩的组装体, 通常称其为生物分子凝聚体或无膜细胞器^[12]. 它们发挥着如下许多重要的功能, 如加速反应速率、环境压力应激、基因转录调控、信号转导、DNA 修复、RNA 剪接与转运等^[13-15].

无膜细胞器中通常含有大量的内禀无序蛋白

质或者固有无序区域 (intrinsically disordered proteins or regions, IDPs/IDRs)^[4,12,16-25], 因此目前对蛋白质相分离的研究大多数主要集中在内禀无序蛋白及其相分离的过程^[26,27]. 通过简单地突变/修饰内禀无序蛋白质或者无序多肽中的氨基酸序列, 就可以完全打乱蛋白质的热响应特性^[28]. 然而, 蛋白质序列的复杂性和构象的多样性, 给内禀无序蛋白质的计算和实验研究带来了重大挑战. 虽然内禀无序蛋白质有很多种类, 但是仍可以将其归类为低复杂度序列 (low complexity sequences, LCS)^[29]. 这种序列仅由二十种天然氨基酸中的小部分构成. 对这样的序列构成的蛋白质而言, 其能量面形貌通常有很多个极小点并且各点的势垒很低^[30,31]. 此外, 这一类分子内部相互作用的简并表现为分子间相互作用的多价性而导致自组装. 由于序列的低复杂度、单分子水平上状态的简并性以及聚集和/或相分离的倾向性, 使得揭示这些序列对应的分子的详细结构和机理的实验研究变得十分困难^[32-35].

* 国家自然科学基金 (批准号: 11947006)、湖北医药学院人才启动金项目 (批准号: 2019QDJZR12, 2018QDJZR22) 和湖北省教育厅科学技术研究项目 (批准号: B2018434).

† 通信作者. E-mail: wbkang@hbmh.edu.cn

理论研究和计算机模拟方法能够提供理论预测和显示分子细节,这不但有助于解释实验现象^[13,36],同时可以用来补充和指导蛋白质的实验研究而展现出其特有的优势.

关于蛋白质分子的“液-液相分离”的多尺度理论模型和多分辨率计算方法很多^[32,37,38](图1).由于蛋白质分子序列和构象的多样性和复杂性,给蛋白质分子的理论研究、计算机模拟和实验研究带来了巨大的挑战^[39,40].本文将对蛋白质分子“液-液相分离”的理论基础和计算机模拟方法进行简要的综述,目的是为进一步研究蛋白质等生物大分子的“液-液相分离”的物理化学机制和过程提供理论和方法借鉴.

2 蛋白质分子“液-液相分离”理论和计算方法简介

2.1 理论研究

2.1.1 Flory-Huggins 理论

从热力学角度看,相分离的发生降低了体系整体的自由能^[11].体系的自由能包括焓与熵变化的贡献,其中焓变包含蛋白质分子与溶剂之间、蛋白质分子之间以及溶剂分子之间势能的变化,而熵变

衡量了体系自由度的变化. Flory^[41]和 Huggins^[42]在20世纪40年代提出了高分子溶液的似晶格模型,用高分子的体积分数 ϕ 作为参数,推导出了理想高分子溶液混合熵变的计算公式,以及用平均作用场假设^[43],引入参数 χ 描述相互作用能的变化,得到了混合焓变的理论计算公式,混合自由能变 F 的计算公式即为

$$\frac{F}{k_B} = \frac{\phi}{N} \ln \phi + (1 - \phi) \ln (1 - \phi) + \chi \phi (1 - \phi), \quad (1)$$

$$\chi = \frac{z}{k_B T} \left[u_{ps} - \frac{1}{2} (u_{pp} + u_{ss}) \right]. \quad (2)$$

式中 k_B 为玻尔兹曼常数, T 为温度, N 为高分子链的长度, z 为晶格的配位数, u_{ps} , u_{pp} 和 u_{ss} 分别为高分子的一个链段与一单位溶剂之间、一对链段之间、一对溶剂之间的结合能. Flory和 Huggins的理论以十分简洁、粗略的形式解释了包含蛋白质分子在内的高分子相分离过程的热力学机制. Flory和 Huggins的理论中体系由于混合熵变总为负,当 χ 取较大的正值时,混合自由能会取到正值,也就是体系趋向于两相,而不是混合,这时混合自由能对浓度(体积分数)的函数存在两个极小值点,即为相分离的临界浓度值^[44].尽管这种方法粗略而十分简洁,但是它给高分子聚合物以及细胞内相分离等的研究提供了基本的理论框架.

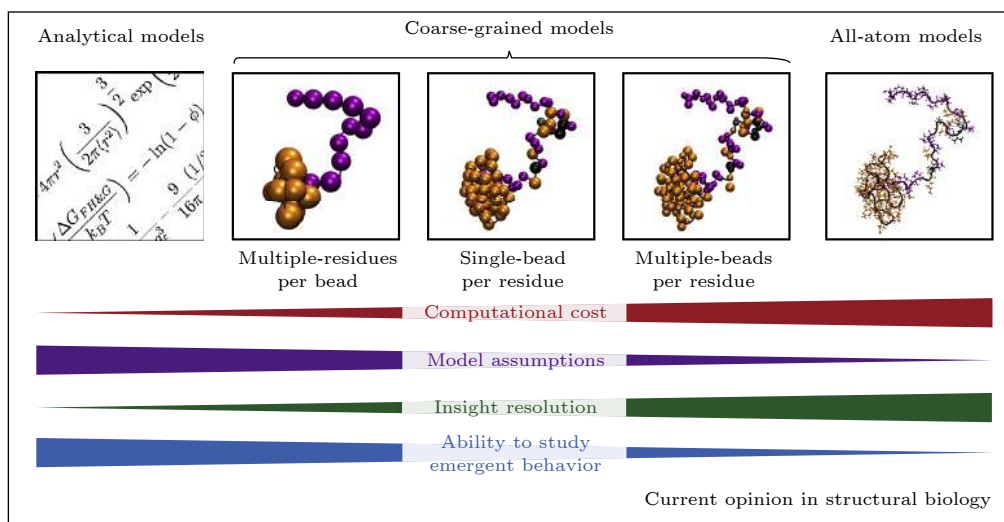


图1 研究蛋白质“液-液相分离”的计算方法:从左至右分别是解析模型、粗粒化模型(包括多个残基单个粒子、单个残基单个粒子、单个残基多个粒子的模型)和全原子模型,分辨率的提升对计算资源的耗费程度急剧增高,而模型假设数的减少会减弱研究涌现行为的能力^[35]

Fig. 1. Computational approaches utilized to study protein liquid-liquid phase separation (LLPS). From left to right are the analytical model, the coarse-grained model (multiple-residues per bead, single-bead per residue and multiple-beads per residue coarse-grained models), and all-atom model. High-resolution descriptions increase the computational cost of the simulations. All-atom model is often impractical for the study of emergent behavior of LLPS^[35].

2.1.2 其他统计热力学理论

在 Flory 和 Huggins 的焓变公式推导过程中只考虑了近程的相互作用, Overbeek 和 Voorn 在他们的工作基础上考虑了长程静电相互作用, 应用于带 Z 个正电和 Z 个负电而整体不带电的聚电解质高分子溶液. 他们假设这些电荷随机的分布于溶液中, 与盐溶液类似可以用 Debye-Hückel 理论假设计算体系中的电荷间相互作用的能量. 虽然他们的平均场理论模型在分子聚合物及其相分离领域提供了十分简洁的形式, 但是该种模型对内禀无序蛋白内部电荷排列及关联性没有成功的解决. 随后, 随机相近似 (random phase approximation, RPA) 方法被应用于的聚电解质高分子溶液静电作用的计算^[45], 该方法可用于任意电荷分布样式的聚电解质高分子溶液体系. 虽然 RPA 理论模型可以给任何形式电荷排列的无序多肽的相分离研究提供了基本的理论框架, 但是在涨落很大的情况下, RPA 模型在 IDPs 构象特性和相分离行为的预测可能会失败. Lin 等^[21,45] 将该方法应用于蛋白质分子相分离体系的计算, 解释了同种电荷在序列上集中分布对于相分离的驱动作用, 以及离子强度对于静电作用为主的相分离体系的影响. 场理论模拟 (field theory simulation, FTS) 方法是一种更细致处理聚电解质高分子溶液体系的方法^[46,47], 通过对场理论中的共轭变量进行模拟采样, 来获得体系的相图. McCarty 等^[1] 用离散的链模型研究了不同谷氨酸 (E) 和赖氨酸 (K) 二组分构成的多肽链, 用场理论模拟方法获得了它们的盐浓度-蛋白浓度相图, 并得到跟 RPA 理论相近的结果.

2.2 计算机模拟方法

蛋白质分子“液-液相分离”的研究中用到最多的是粗粒化模拟方法, 全原子模型的计算机模拟方法偶尔也会用到. 这里的粗粒化模拟方法是指将多个原子简化为单个粒子来降低分辨率提高计算效率的计算机模拟方法^[35]. 该方法是通过显著地减少粒子的数量的模型, 并为理论研究和全原子模拟之间的沟通提供了一个桥梁. 下面将重点综述近几年发展起来的粗粒化模型.

2.2.1 非格点模型的粗粒化模拟方法

1) 多个残基简化为单个球的粗粒化模型

根据蛋白质中氨基酸序列的特征, 可将一条多

肽简化为一个粗粒化的粒子或将一个重复的短肽片段处理成一个粗粒化的粒子. 对于低复杂度序列组成的无序蛋白质, 其序列的性质有利于粗粒化建模, 对同聚蛋白质尤为方便^[13,48,49]. 该种类型的蛋白质指的是由单种氨基酸组成的短肽或是短肽重复片段组成的蛋白质. 为了研究重复序列长度和序列关联性对相分离的影响, 可以将重复序列构成的一条多肽简化为一个粒子或将多个重复的片段简化为一个粒子. 例如, 将 huntingtin 蛋白可视为二嵌段聚合物而得到了广泛的粗粒化研究^[50]. huntingtin 蛋白的外显子包含一个多聚谷氨酰胺束, 可以看作是一个仅由多聚谷氨酰胺和脯氨酸构成的二嵌段共聚物^[50]. Burke 等^[48] 利用多个氨基酸残基简化为一个粒子的粗粒化模型探索了多聚谷氨酰胺束长度和构象倾向性对其自组装过程的影响. 研究表明, 其自组装的驱动力主要取决于多聚谷氨酰胺束的长度、蛋白质浓度和两个区块的相对疏水性, 而不是多聚谷氨酰胺束的构象偏好^[48]. Condon 等^[51] 研究了 (elastin-like polypeptide, ELP) 弹性蛋白, 通过用一个粒子表示 1~2 个重复单元, 这些模拟能够达到与实验研究相当的长度和时间尺度, 并可重现与实验测量的涌现行为一致的特征.

2) 单个残基简化为单个球的粗粒化模型

根据蛋白质中氨基酸序列的特征, 可以将单个氨基酸简化为包含疏水性和电荷性质特征的球, 也就是通常所说的残基化水平的粗粒化模型. 残基化水平的粗粒化模型可以用来理解驱动分子间相互作用和相分离行为序列的重要特征^[36,52,53]. 单个氨基酸被简化为单个粒子, 通常用疏水性指数来参数化成对氨基酸之间的相互作用, 并且重复出了实验上观察到的构象. 例如, Dignon 等^[17,52] 基于 Kapcha 和 Rosky^[54] 氨基酸的疏水性指数和 Ashbaugh-Hatch^[55] 势函数的形式 开发出一种将单个残基简化为单个球的粗粒化模型来研究蛋白质的相分离的计算方法. 在这种相互作用势函数形式中, 通过比较粗粒化模拟与实验测定得到的回转半径, 确定了成对氨基酸之间短程相互作用的最佳数值. Ghavami 等^[56] 利用类似的方法研究了酵母核孔 (nuclear pore complex, NPC) 中苯丙氨酸-甘氨酸序列组成多肽的自组装结构. 通过实验测定的斯托克斯半径和疏水性指数来确定粗粒化模拟参数, 粗粒化模拟表明, 在 NPC 的中心区域重复序

列形成圆饼状的结构. 序列突变研究发现, 这种组织结构依赖于带电残基的组分和排列. Borgia 等^[57]使用同样的方法研究了相同数目的谷氨酸和天冬氨酸以及赖氨酸和精氨酸的短肽. 模拟结果与核磁共振光谱数据吻合得很好, 表明相互作用纯粹是静电驱动的, 不同类型带电氨基酸之间的结构差异对于这个系统来说并不显著^[58].

单个残基水平的粗粒化模型不仅可以提供相分离过程中蛋白质分子间何种相互作用驱动相分离和序列特征, 而且弥补了理论模型的局限性. 例如, 基于 RPA 理论^[45]和粗粒化模拟研究表明, (EK)₅₀ 序列中不同电荷排列模式导致相分离结果定性相似, 但定量不同的现象. 同样地, 通过使用一个氨基酸一个粒子的显式粗粒化模型, Song 等^[59]的研究表明, (EK)₅₀ 序列的回转半径和链末端距离由于构象异质性而不相关, 这是均聚物模型通常缺少的特征. 综合起来, 单个残基水平的粗粒化模型提供了一个可以与解析理论进行比较的途径, 有助于识别系统的局限性, 同时也为模拟低复杂度序列提供了一个简单而强大的模拟框架.

3) 单个残基简化为多个球的粗粒化模型

单个残基简化为多个球的粗粒化模型是更精细的粗粒化模型. 这里每个残基可视为由多个粒子组成的. 该模型有两种类型: 可移植的模型和系统特定模型 (在所有被研究的系统中永久化是固定的)^[60]. 常见的可移植多个粒子组成单个残基的粗粒化模型有 PRIME^[61–65], PLUM^[66–68], AWSEM^[69]和 MARTINI^[70–73]模型. 这些模型被用于模拟同源蛋白的原纤化形成和重复序列短肽的聚集过程^[62–64], 且可用来研究 NPC 蛋白的门控和运输机制和类蚕丝弹性蛋白热应答行为的分子机制^[71]. 其中 PLUM 模型被用来研究类蚕丝弹性蛋白单体的转变温度, 与实验研究结果是一致的^[66].

尽管上述的多粒子的粗粒化模型具有明显的普遍性, 可移植模型并不总是适合于研究低复杂度的蛋白质序列. 一个给定模型的相关性通常取决于模型最初参数化. 例如, 原始的 PLUM 模型可能导致一些 IDPs 的二级结构过度稳定, 而 MARTINI 模型, 可能不会重现出蛋白质的结构特征并且禁止了结构转变^[70]. 由于可移植模型的局限性, 系统特定的粗粒化模型仅仅被频繁地用于研究低复杂度序列. 这些模型经常使用从全原子模拟收集的信息来参数化特定的研究对象. 从全原子模拟到粗粒化

模型的建立有两种主要方法: 基于结构的方法和力匹配方法^[74]. 在基于结构的方法中, 目标是从全原子模拟中再现特定的构象分布. 而在力匹配方法从全原子模拟映射到粗粒化模型中更简单的经验力场 (已经被用来研究与疾病相关的同源蛋白的聚集). 由 Hills 和 Voth^[75] 开创的力匹配方法, 该方法被用来研究了酵母菌核孔复合体中的由苯丙氨酸和甘氨酸组成的低复杂度序列 (phenylalanine-glycine LCS within the yeast nuclear pore complex, FG-nups) 的构象特性. 这种方法捕捉到了几个较长的 FG-nups 的两相无序态, 并且表明序列组分以及序列特异性对纤维蛋白核糖核酸酶的构象偏差有强烈的影响^[76].

4) 长条形盒子动力学模拟 (slab simulation) 的粗粒化方法

该模拟方法将所有分子放在长条形的周期性盒子 (三个方向边长: $x = y \ll z$) 中, 在 NPT 系综的隐式溶剂分子动力学模拟中, 做只沿 z 轴方向平衡的各向异性压力耦合 (pressure coupling). 这样分子模拟中相分离完成后蛋白分子聚集于凝聚相, 在 z 轴的很窄的区域分布. Mittal 实验室^[4,5,17,52]开发了用于这一模拟的蛋白质粗粒化方法. 蛋白质链的每个残基看作单个粒子, 相邻粒子间用弹簧相连. 不相邻粒子间包含两种作用势函数: Debye-Hückel 静电屏蔽的静电作用和短程的成对势. 短程的成对势从 Lenard-Jones (LJ) 势改造而来, Mittal 实验室发展了 (hydrophobicity scale, HPS)^[55]和 Kim-Hummer (KH)^[77]两种模型. 其中 HPS 模型是基于残基对疏水性调整 LJ 势的模型, KH 模型是用于研究蛋白-蛋白相互作用的修正 LJ 势的势阱深度参数的模型. 这两个模型的未定参数都通过拟合天然无序蛋白质的回转半径 (来自小角 X 射线散射实验) 得到. 随后他们用这个粗粒化动力学模拟方法计算了 (fused in sarcoma, FUS) FUS 中的 DNA 结合蛋白和 LAF 中的解螺旋蛋白 LAF-1 片段这两个相分离体系的温度-浓度相图, 其计算表明 FUS 的模拟磷酸化突变使最高临界共溶温度降低 (即不利于相分离), 将 FUS 序列拷贝延长将增加最高临界共溶温度, LAF 全长的最高临界共溶温度比只有 IDRs 时的值高. 这些结果都与实验结果相吻合. 该研究组还用这个模型计算了临界温度值, 以及天然无序蛋白质单分子的性质, 即天然无序线团到蜷缩球转变温度 (T_θ) 高

度相关, 因此他们认为天然无序蛋白自身的性质可以用于预估其相分离发生的条件^[17]. 他们也与 Fawzi 研究组合作, 用此模拟方法研究了 hnRNPA2 低复杂度结构域的单点突变对其相分离条件的影响. 他们发现天冬氨酸 (D) 到缬氨酸 (V) 的突变体 (D290 V) 比野生型具有更低的临界共溶温度, 相同温度下, 此突变体的临界浓度比野生型高^[78].

粗粒化模型对多肽早期聚集事件和聚集形态可以提供有价值的信息. 这些模型可以明确地捕获序列特异性, 这是解析模型中通常缺乏的特征. 然而粗粒化模型有其自身的局限性. 粗粒化模型的参量化往往基于全原子模拟的模型, 因此高度依赖于全原子模拟的准确性, 并且多体相互作用项可以通过对较小数量分子的模拟得到. 此外, 粗粒化模型的相关性依赖于这样的假设: 模型能够恰当地捕获到感兴趣的区域的特性. 对于能够在高阶自组装过程中经历涌现构象转变的低复杂度序列, 粗粒化模型在其捕获自联结的热力学和动力学的方面可能会从根本上受到限制. 综上所述, 虽然粗粒化模型提供了一条重要的途径来评估个别分子的序列依赖性偏差, 但在可以得出的结论的确定性方面, 应始终考虑到它们的局限性.

2.2.2 格点模型的粗粒化模拟

格点模型也可以被用来研究生物大分子的“液-液相分离”过程. 例如, Pappu 课题组和 Chan 课题组分别开发了不同的格点模型用来研究无序多肽或者生物大分子的“液-液相分离”. 其中美国华盛顿大学的 Pappu 实验组专注于该方法的研究, 他们开发了 LASSI (Lattice simulation of Sticker and Spacer Interactions, LASSI) 具体实现了格点模型的聚集模拟^[79]. 在格点模型模拟过程中, 一般将生物大分子的结构域或者功能模块整体作为单个粗粒化的粒子, 粒子只能处于离散立方网格的格点中, 不同粒子所在格点不能重合, 一般假设只有相邻格点粒子有相互作用并且粒子对的相互作用能参数化为一个确定值, 体系用蒙特卡洛方法演化采样. Pappu 实验室用此模拟方法研究了核仁 FIB1-NPM1-RNA 三元体系中结构域相互作用模式影响形成分层液滴的机制, 其中用一个分枝的结构模拟 NPM1 分子五聚体的结构^[13]. 他们也用此模拟方法做了 $(SH3)_m + (PRM)_n$ 体系中无序连接链对相分离的影响的研究^[18]. 模拟结果显示

连接链的有效溶剂体积影响该体系聚集态的性质, 连接链有效溶剂体积大、主要为较刚性的伸展构象时体系倾向于聚集成凝胶, 相反的, 连接链以压缩状态为主时体系倾向于聚集成液滴.

Das 等^[53]应用格点模型上的粗粒化模拟方法研究了 $(EK)_{25}$ 序列的不同电荷排布情况下的多肽序列的相分离特点. 在他们的研究中, 重点考察了正负电荷混合均匀和完全分离的两种典型的序列. 他们认为每一条多肽链以自回避的形式占据三维空间点. 将每一个氨基酸残基简化为单个粒子并只占有一个格点, 两个相邻的粒子之间通过共价键来链接, 此外所有的粒子静电作用为屏蔽的库伦静电相互作用. 通过大量的蒙特卡洛模拟研究发现, 可以发现两种不同的序列相分离的临界温度有明显的差异, 对于正负电荷分离形成大的电荷模块的序列来说, 这一类序列显示出明显的相分离特性. 他们在格点模型上的大规模蒙特卡洛模拟研究与 RPA 理论研究的结果是一致的.

2.2.3 全原子模型

虽然粗粒化模拟已被用于研究低复杂度序列的聚集和组装, 全原子模拟对于在单个多肽水平上提取详细的构象特征具有更高的价值. 鉴于其普遍的序列重复性和高聚集倾向性, 这些蛋白质的研究往往给实验带来了巨大的挑战性. 全原子模拟提供了一个方便的解决方案, 提供了无限稀释下的“无限分辨率”^[80-87]. 对于易于聚集的同源蛋白质 (例如, 聚谷氨酰胺、聚甘氨酸、聚丙烯酸) 的全原子模拟已经对单体的构象偏差有了相当深入的了解. 单分子荧光共振实验补充了全原子模拟的研究. 例如, Warner 等他们能够提取 huntingtin 外显子 1 单体的构象系综. 全原子模拟也可以与其它实验相结合技术, 如核磁共振波谱法和小角度 X 射线散射产生互补信息. 例如 TDP-43 的 LCS 结构域、hnRNPA2 的 LCS 结构域和 RNA 聚合酶 II 的碳末端等系统得到了全原子模拟和上述实验技术的研究.

尽管全原子模拟提供了高分辨率的解析精度, 但它们并非没有局限性. 考虑到它们的计算成本, 研究大量分子的分子间结合和聚集仍然令人望而却步, 得出的结论的准确性依赖于力场的可靠性. 在低复杂度序列的研究中力场的局限性往往被放大. 特别是在全原子分子动力学的框架下, 构型状

态的高度简并性使得构型抽样成为一个主要的挑战. 因此, 尽管全原子模拟已经为理解 LCS 序列提供了较高的精度, 但是依然要像粗粒化方法一样要谨慎地考虑全原子模拟方法的局限性.

2.2.4 各种模型的优缺点概述

上述不同的理论方法和计算模型有着各自的优势和用处, 也有着不足和缺陷, 在使用的时候不能全盘否定. 例如, 平均场理论十分简洁地给出了高分子聚合物及其相分离的理论表达式, 但是在针对具体体系和涉及电荷关联性较强的情况下, 理论预测可能显得不是太准确. RPA 理论模型的优点在于能够给出电荷相互作用如何调控 IDPs 相分离过程的清晰的物理图像, 缺点在于针对具体体系常常预测能力显得不是十分充足. 格点模型的优势就是抓住了整个问题中的最核心要素, 简化了构象空间, 使得计算简单并且快速. 而非格点模型描述构象相对比较准确, 各种相互作用也可以考虑得更精细, 粗粒化模型的参量化往往是基于全原子模拟的模型, 因此高度依赖于全原子模拟的准确性. 全原子模型的优点在于计算精确, 能够给出具体的相互作用细节, 能够更为精细地描述结构及其特征, 但是由于原子数目的显著增加, 导致计算太耗费时间, 目前只能研究小体系或者相分离的早期聚集等行为特征. 总之, 在实际的研究过程中, 要根据所研究对象的需求和具体的特征, 来选择合适的方法恰当的描述系统的相互作用及其集体运动行为.

3 总结与展望

当前, 蛋白质分子的构象特点、相分离过程及其序列特征等引起了众多科学家的广泛兴趣. 而多尺度和多分辨率的理论和模型在生物分子的相分离的研究中得到了迅速的发展和改进. 本文综述了关于蛋白质分子“液-液相分离”的多种先进的理论和计算机模拟方法.

生物大分子相分离过程的研究当前处于起步和发展阶段. 研究者可以将上述的各种方法巧妙地应用于生物系统的相分离中来. 也可以发展新的理论方法以及模拟手段研究生物体系的相分离. 最终为深入地理解蛋白质等生物大分子的序列、结构以及相分离产生的临界条件之间的基本关系提供恰当而快速的理论方法和模拟手段. 此外, 也可以对

上述的研究方法进行混合来解决生物大分子的相分离问题. 相信在未来, 生物大分子涉及的相分离将会吸引越来越多的科学研究人员的关注, 并且应用上述的研究手段或者发展新的方法来解决相分离相关的问题.

参考文献

- [1] McCarty J, Delaney K T, Danielsen S P, Fredrickson G H, Shea J E 2019 *J. Phys. Chem. Lett.* **10** 1644
- [2] Lin Y H, Song J H, Forman-Kay J D, Chan H S 2017 *J. Mol. Liq.* **228** 176
- [3] Lin Y H, Song J H, Forman-Kay J D, Chan H S 2019 *J. Mol. Liq.* **273** 676
- [4] Dignon G L, Zheng W W, Kim Y C, Mittal J, Best R B 2018 *Biophys. J.* **114** 431a
- [5] Dignon G L, Zheng W W, Best R B, Mittal J 2017 *Biophys. J.* **112** 200a
- [6] Choi J M, Mitrea D M, Stanley C B, Ruff K M, Holehouse A S, Kriwacki R W, Pappu R V 2019 *Biophys. J.* **116** 349a
- [7] Chin A F, Hilser V J, Zheng Y X 2019 *Biophys. J.* **116** 179a
- [8] Zhou F M, Sun D K, Zhu M F 2010 *Acta Phys. Sin.* **59** 3394 (in Chinese) [周丰茂, 孙东科, 朱鸣芳 2010 物理学报 **59** 3394]
- [9] Zhang C S, Lai L H 2010 *Acta Phys. Chim. Sin.* **36** 1907053 (in Chinese) [张长胜, 来鲁华 2010 物理化学学报 **36** 1907053]
- [10] Burke K A, Janke A M, Rhine C L, Fawzi N L 2015 *Mol. Cell* **60** 231
- [11] Brangwynne C P, Tompa P, Pappu R V 2015 *Nat. Phys.* **11** 899
- [12] Uversky V N 2017 *Curr. Opin. Struct. Biol.* **44** 18
- [13] Feric M, Vaidya N, Harmon T S, Mitrea D M, Zhu L, Richardson T M, Kriwacki R W, Pappu R V, Brangwynne C P 2016 *Cell* **165** 1686
- [14] Falahati H, Wieschaus E 2017 *Proc. Natl. Acad. Sci. U.S.A.* **114** 1335
- [15] Sabari B R, Dall'Agnese A, Boija A, Klein I A, Coffey E L, Shrinivas K, Abraham B J, Hannett N M, Zamudio A V, Manteiga J C, Li C H, Guo Y E, Day D S, Schuijers J, Vasile E, Malik S, Hnisz D, Lee T I, Cisse, I I, Roeder R G, Sharp P A, Chakraborty A K, Young R A 2018 *Science* **361** eaar3958
- [16] Das R K, Pappu R V 2013 *Proc. Natl. Acad. Sci. U.S.A.* **110** 13392
- [17] Dignon G L, Zheng W W, Best R B, Kim Y C, Mittal J 2018 *Proc. Natl. Acad. Sci. U.S.A.* **115** 9929
- [18] Harmon T S, Holehouse A S, Rosen M K, Pappu R V 2017 *ELIFE* **6** e30294
- [19] Kang W B, He C, Liu Z X, Wang J, Wang W 2019 *J. Biomol. Struct. Dyn.* **37** 1956
- [20] Lin Y, Currie S L, Rosen M K 2017 *J. Biol. Chem.* **292** 19110
- [21] Lin Y H, Brady J P, Forman-Kay J D, Chan H S 2017 *New J. Phys.* **19** 115003
- [22] Schuster B S, Dignon G, Jahnke C, Good M C, Hammer D A, Mittal J 2019 *Biophys. J.* **116** 453a
- [23] Smith J, Calidas D, Schmidt H, Lu T, Seydoux G 2016 *ELIFE* **5** e21337
- [24] Uebel C J, Anderson D C, Mandarino L M, Manage K I, Aynaszyan S, Phillips C M 2018 *Plos Genet.* **14** e1007542
- [25] Kang W B, Wang J, Wang W 2018 *Acta Phys. Sin.* **67** 058701 (in Chinese) [康文斌, 王骏, 王炜 2018 物理学报 **67** 058701]

- [26] Pak C W, Kosno M, Holehouse A S, Padrick S B, Mittal A, A R, Yunus A A, Liu D R, Pappu R V, Rosen M K 2016 *Mol. Cell* **63** 72
- [27] Gates Z P, Baxa M C, Yu W, Riback J A, Li H, Roux B, Kent S B, Sosnick T R 2017 *Proc. Natl. Acad. Sci. U.S.A.* **114** 2241
- [28] Riback J A, Katanski C D, Kear-Scott J L, Pilipenko E V, Rojek A E, Sosnick T R, Drummond D A 2017 *Cell* **168** 1028
- [29] Das R K, Ruff K M, Pappu R V 2015 *Curr. Opin. Struct. Biol.* **32** 102
- [30] Vitalis A, Wang X, Pappu R V 2007 *Biophys. J.* **93** 1923
- [31] Papoian G A 2008 *Proc. Natl. Acad. Sci. U. S. A.* **105** 14237
- [32] Levine Z A, Shea J E 2017 *Curr. Opin. Struct. Biol.* **43** 95
- [33] Dignon G L, Zheng W, Kim Y C, Mittal J 2019 *ACS Central Sci.* **5** 821
- [34] Murthy A C, Dignon G L, Kan Y, Zerze G H, Parekh S H, Mittal J, Fawzi N L 2019 *Nat. Struct. Mol. Biol.* **26** 637
- [35] Ruff K M, Pappu R V, Holehouse A S 2019 *Curr. Opin. Struct. Biol.* **56** 1
- [36] Monahan Z, Ryan V H, Janke A M, Burke K A, Rhoads S N, Zerze G H, O'Meally R, Dignon G L, Conicella A E, Zheng W W, Best R B, Cole R N, Mittal J, Shewmaker F, Fawzi N L 2017 *EMBO J.* **36** 2951
- [37] Best R B 2017 *Curr. Opin. Struct. Biol.* **42** 147
- [38] Robustelli P, Piana S, Shaw D E 2018 *Proc. Natl. Acad. Sci. U.S.A.* **115** E4758
- [39] van L R, Buljan M, Lang B, Weatheritt R J, Daughdrill G W, Dunker A K, Fuxreiter M, Gough J, Gsponer J, Jones D T, Kim P M, Kriwacki R W, Oldfield C J, Pappu R V, Tompa P, Uversky V N, Wright P E, Babu M M 2014 *Chem. Rev.* **114** 6589
- [40] Mittag T, Forman-Kay J D 2007 *Curr. Opin. Struct. Biol.* **17** 3
- [41] Flory P J 1942 *J. Chem. Phys.* **10** 51
- [42] Huggins M L 1942 *J. Chem. Phys.* **46** 151
- [43] Samanta H S, Zhuravlev P I, Hinczewski M, Hori N, Chakrabarti S, Thirumalai D 2017 *Soft Matter* **13** 3622
- [44] Zhou H X, Nguemaha V, Mazarakos K, Qin S B 2018 *Trends Biochem. Sci.* **43** 499
- [45] Lin Y H, Forman-Kay J D, Chan H S 2016 *Phys. Rev. Lett.* **117** 178101
- [46] Liu Y X, Zhang H D, Tong C H, Yang Y L 2011 *Macromolecules* **44** 8261
- [47] Speck T, Menzel A M, Bialke J, Lowen H 2015 *J. Chem. Phys.* **142** 224109
- [48] Burke M G, Woscholski R, Yaliraki S N 2003 *Proc. Natl. Acad. Sci. U.S.A.* **100** 13928
- [49] Ruff K M, Khan S J, Pappu R V 2014 *Biophys. J.* **107** 1226
- [50] Zuccato C, Valenza M, Cattaneo E 2010 *Physiol. Rev.* **90** 905
- [51] Condon J E, Martin T B, Jayaraman A 2017 *Soft Matter* **13** 2907
- [52] Dignon G L, Zheng W, Kim Y C, Best R B, Mittal J 2018 *PLOS Comput. Biol.* **14** e1005941
- [53] Das S, Eisen A, Lin Y H, Chan H S 2018 *J. Phys. Chem. B* **122** 5418
- [54] Kapcha L H, Rossky P J 2014 *J. Mol. Boil.* **426** 484
- [55] Ashbaugh H S, Hatch H W 2008 *J. Am. Chem. Soc.* **130** 9536
- [56] Ghavami A, Veenhoff L M, van der Giessen E, Onck P R 2014 *Biophys. J.* **107** 1393
- [57] Borgia A, Borgia M B, Bugge K, Kissling V M, Heidarsson P O, Fernandes C B, Sottini A, Soranno A, Buholzer K J, Nettek D, Kragelund B B, Best R B, Schuler B 2018 *Nature* **555** 61
- [58] Wang J, Choi J M, Holehouse A S, Lee H O, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovskiy A, Drechsel D, Poser I, Pappu R V, Alberti S, Hyman A A 2018 *Cell* **174** 688
- [59] Song J, Gomes G N, Shi T, Gradinaru C C, Chan H S 2017 *Biophys. J.* **113** 1012
- [60] Noid W G 2013 *J. Chem. Phys.* **139** 090901
- [61] Voegler Smith A, Hall C K 2001 *Proteins* **44** 344
- [62] Marchut A J, Hall C K 2006 *Biophys. J.* **90** 4574
- [63] Marchut A J, Hall C K 2007 *Proteins* **66** 96
- [64] Nguyen H D, Hall C K 2006 *J. Am. Chem. Soc.* **128** 1890
- [65] Cheon M, Chang I, Hall C K 2010 *Proteins* **78** 2950
- [66] Yeo J J, Huang W W, Tarakanova A, Zhang Y W, Kaplan D L, Buehler M J 2018 *J. Mater. Chem. B* **6** 3727
- [67] Bereau T, Deserno M 2009 *J. Chem. Phys.* **130** 235106
- [68] Rutter G O, Brown A H, Quigley D, Walsh T R, Allen M P 2015 *Phys. Chem. Chem. Phys.* **17** 31741
- [69] Chen M C, Wolynes P G 2017 *Proc. Natl. Acad. Sci. U.S.A.* **114** 4406
- [70] Seo M, Rauscher S, Pomes R, Tieleman D P 2012 *J. Chem. Theory Comput.* **8** 1774
- [71] Miao L, Schulten K 2009 *Structure* **17** 449
- [72] Monticelli L, Kandasamy S K, Periole X, Larson R G, Tieleman D P, Marrink S J 2008 *J. Chem. Theory Comput.* **4** 819
- [73] Marrink S J, Risselada H J, Yefimov S, Tieleman D P, Vries A H 2007 *J. Phys. Chem. B* **111** 7812
- [74] Lu L Y, Dama J F, Voth G A 2013 *J. Chem. Phys.* **139** 121906
- [75] Hills R D, Lu L, Voth G A 2010 *PLOS Comput. Biol.* **6** e1000827
- [76] Ando D, Zandi R, Kim Y W, Colvin M, Rexach M, Gopinathan A J 2014 *Biophys. J.* **106** 1997
- [77] Kim Y C, Hummer G 2008 *J. Mol. Biol.* **375** 1416
- [78] Ryan V H, Dignon G L, Zerze G H, Chabata C V, Silva R, Conicella A E, Amaya J, Burke K A, Mittal J, Fawzi N L 2018 *Mol. Cell* **69** 465
- [79] Choi J-M, Dar F, Pappu R V 2019 *PLOS Comput. Biol.* **15** e1007028
- [80] Tran H T, Mao A, Pappu R V 2008 *J. Am. Chem. Soc.* **130** 7380
- [81] Holehouse A S, Garai K, Lyle N, Vitalis A, Pappu R V 2015 *J. Am. Chem. Soc.* **137** 2984
- [82] Asthagiri D, Karandur D, Tomar D S, Pettitt B M 2017 *J. Phys. Chem. B* **121** 8078
- [83] Karandur D, Wong K Y, Pettitt B M 2014 *J. Phys. Chem. B* **118** 9565
- [84] Vitalis A, Wang X, Pappu R V 2008 *J. Mol. Biol.* **384** 279
- [85] Vitalis A, Lyle N, Pappu R V 2009 *Biophys. J.* **97** 303
- [86] Kang H, Vazquez F X, Zhang L, Das P, Toledo-Sherman L, Luan B, Levitt M, Zhou R 2017 *J. Am. Chem. Soc.* **139** 8820
- [87] Daggett V, Kollman P A, Kuntz I D 1991 *Biopolymers* **31** 1115

REVIEW

Theoretical and computational methods of protein liquid-liquid phase separation*

Zhang Peng-Cheng¹⁾²⁾ Fang Wen-Yu¹⁾²⁾ Bao Lei¹⁾²⁾ Kang Wen-Bin^{1)2)†}¹⁾ (*School of Public Health and Management, Hubei University of Medicine, Shiyan 442000, China*)²⁾ (*Hubei Biomedical Detection Sharing Platform in Water Source Area of South to North Water Diversion Project, Shiyan 442000, China*)

(Received 24 March 2020; revised manuscript received 10 April 2020)

Abstract

Liquid-liquid phase separation (LLPS) of proteins is an emerging field in the research of biophysics. Many intrinsically disordered proteins (IDPs) are known to have the ability to assemble via LLPS and to organize into protein-rich and dilute phases both in vivo and in vitro. Such a kind of phase separation of proteins plays an important role in a wide range of cellular processes, such as the formation of membraneless organelles (MLOs), signaling transduction, intracellular organization, chromatin organization, etc. In recent years, there appeared a great number of theoretical analysis, computational simulation and experimental research focusing on the physical principles of LLPS. In this article, the theoretical and computational simulation methods for the LLPS are briefly reviewed. To elucidate the physical principle of LLPS and to understand the phase behaviors of the proteins, biophysicists have introduced the concepts and theories from statistical mechanics and polymer sciences. Flory-Huggins theory and its extensions, such as mean-field model, random phase approximation (RPA) and field theory simulations, can conduce to understanding the phase diagram of the LLPS. To reveal the hidden principles in the sequence-dependent phase behaviors of different biomolecular condensates, different simulation methods including lattice models, off-lattice coarse-grained models, and all-atom simulations are introduced to perform computer simulations. By reducing the conformational space of the proteins, lattice models can capture the key points in LLPS and simplify the computations. In the off-lattice models, a polypeptide can be coarse-grained as connected particles representing repeated short peptide fragments. All-atom simulations can describe the structure of proteins at a higher resolution but consume higher computation-power. Multi-scale simulation may provide the key to understanding LLPS at both high computational efficiency and high accuracy. With these methods, we can elucidate the sequence-dependent phase behaviors of proteins at different resolutions. To sum up, it is necessary to choose the appropriate method to model LLPS processes according to the interactions within the molecules and the specific phase behaviors of the system. The simulations of LLPS can facilitate the comprehensive understanding of the key features which regulate the membraneless compartmentalization in cell biology and shed light on the design of artificial cells and the control of neurodegeneration.

Keywords: liquid-liquid phase separation, intrinsically disordered proteins, multivalent interaction, multi-scale simulation

PACS: 87.15. A-, 87.15.Cc, 87.18.-h, 47.11.St

DOI: 10.7498/aps.69.20200438

* Project supported by the National Natural Science Foundation of China (Grant No. 11947006), the Cultivating Project for Young Scholar at Hubei University of Medicine, China (Grant Nos. 2019QDJZR12, 2018QDJZR22), and the Natural Science Foundation of Hubei Provincial Department of Education, China (Grant No. B2018434).

† Corresponding author. E-mail: wbkang@hbmh.edu.cn