

内禀无序蛋白构象与带电氨基酸残基排布关系——以精氨酸和天冬氨酸组成的随机多肽为例

康文斌 王骏 王炜

Conformation of disordered peptides modulated by distributions of charged residues: Case study of random peptides composed of arginines and aspartic acids

Kang Wen-Bin Wang Jun Wang Wei

引用信息 Citation: *Acta Physica Sinica*, 67, 058701 (2018) DOI: 10.7498/aps.20172246

在线阅读 View online: <http://dx.doi.org/10.7498/aps.20172246>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2018/V67/I5>

您可能感兴趣的其他文章

Articles you may be interested in

[基于氨基酸位置特异性的蛋白质 Loop 区结构预测改进方法](#)

An improved algorithm for prediction of protein loop structure based on position specificity of amino acids
物理学报.2016, 65(15): 158701 <http://dx.doi.org/10.7498/aps.65.158701>

[电离辐射致植物诱变效应的损伤-修复模型](#)

Damage-repair model for mutagenic effects of plant induced by ionizing radiation
物理学报.2015, 64(24): 248701 <http://dx.doi.org/10.7498/aps.64.248701>

[生物大分子多尺度理论和计算方法](#)

Multiscale theory and computational method for biomolecule simulations
物理学报.2015, 64(9): 098701 <http://dx.doi.org/10.7498/aps.64.098701>

[微环境促进的侵袭性肿瘤生长的元胞自动机模拟](#)

Microenvironment-enhanced invasive tumor growth via cellular automaton simulations
物理学报.2015, 64(5): 058706 <http://dx.doi.org/10.7498/aps.64.058706>

内禀无序蛋白构象与带电氨基酸残基排布关系 ——以精氨酸和天冬氨酸组成的随机多肽为例*

康文斌^{1)2)3)†} 王骏^{1)‡} 王伟^{1)††}

1)(南京大学物理学院, 南京 210093)

2)(湖北医药学院数理教研室, 十堰 442000)

3)(湖北医药学院 Bio-X 研究中心, 十堰 442000)

(2017年10月17日收到; 2017年12月11日收到修改稿)

内禀无序蛋白的结构特征与其氨基酸序列有着密切的联系. 其中一个核心问题是正负带电氨基酸残基的排列如何影响无序蛋白或者多肽的构象? 为了回答这一问题, 本研究以天冬氨酸和精氨酸两种带电残基组成的随机多肽为研究对象, 利用全原子蒙特卡罗模拟和温度副本交换采样方法, 研究了随机多肽的电荷排布与结构之间的定性关系. 结果表明: 正负带电残基在序列上混合均匀时, 由于肽链内部的静电吸引和排斥相互抵消, 肽链倾向于形成无规卷曲的构象; 正负带电残基分离时, 由于长程静电相互吸引, 多肽倾向于形成类 β -发卡的形状.

关键词: 内禀无序蛋白, 构象, 序列, 分子模拟

PACS: 87.10.Rt, 87.14.ef, 87.15.A-, 87.15.Qt

DOI: 10.7498/aps.67.20172246

1 引言

内禀无序蛋白(下文简称为无序蛋白), 在蛋白质的转录调控和信号转导方面扮演着非常重要的角色^[1,2]. 无序蛋白不能自发折叠, 原因可能在于缺乏驱动折叠的疏水驱动力. 研究表明, 至少75%的无序蛋白是聚两性电解质^[3-10]. 无序蛋白或者无序多肽的序列中缺乏疏水性的氨基酸而富含极性和带电氨基酸^[3], 其中, 赖氨酸(Lys, K)、谷氨酸(Glu, E)、精氨酸(Arg, R)和天冬氨酸(Asp, D)出现的频率都很高^[11,12]. 近年来, 研究人员用带电残基百分比(fraction of charged residue, FCR)、残基平均净电量(net charge per residue, NCPR)

等特征量粗略地描述了无序蛋白序列与构象之间的关系^[6,13,14]. 例如, 2010年, Mao等^[13]研究发现NCPR可以作为序参量来预测聚两性电解质倾向形成紧密的球形还是松散的无规卷曲; 2013年, Das和Pappu^[4]发现赖氨酸和谷氨酸组成的随机多肽的构象依赖于正负电荷的排布情况.

正是由于电荷数量和排布的重要性, 自然界中无序蛋白常通过甲基化、磷酸化等化学修饰过程来改变自身的电荷分布特性^[15-21](如甲基化会通过精氨酸或赖氨酸的带电侧链的化学修饰, 使得相应残基变为电中性), 从而影响其结构系综, 进而激活或者抑制相应的生物功能. 这种现象在近来的研究中被广泛观察到^[15,17-22]. 例如, Pappu等^[21]发现, 细胞周期抑制蛋白p27的磷酸化导致电荷分

* 国家重点基础研究发展计划(批准号: 2013CB834100)、国家自然科学基金(批准号: 11334004, 11174133, 81421091, 11774157)、2017年国家级大学生创新创业训练计划(批准号: 201710929002)和湖北医药学院研究生启动金(批准号: 2011QDZR-11)资助的课题.

† 通信作者. E-mail: wbkang@hbm.edu.cn

‡ 通信作者. E-mail: wangj@nju.edu.cn

†† 通信作者. E-mail: wangwei@nju.edu.cn

布的改变, 进而引起p27蛋白结构的改变, 最终引起细胞周期的变化等. 这种通过影响多肽或蛋白质中的电荷分布来调控无序多肽或者无序蛋白构象和功能的策略也已经在多个真实体系中得到了应用^[15,17-22].

蛋白质结构与序列关系的定量研究对认识无序蛋白的功能具有非常重要的意义^[23-31]. Das等^[4]针对由谷氨酸和赖氨酸组成的随机多肽开展了相关研究. 众所周知, 无序蛋白中还存在很多不同种类的带电氨基酸残基. 例如, 精氨酸是无序蛋白发挥功能的重要调控位点, 因此研究富含精氨酸的多肽有着非常重要的理论意义. 同时, 一些具有相同电荷的残基在局部柔性上表现出不同的特征. 谷氨酸在 α -螺旋结构中比较常见, 天冬氨酸在 β -转角中出现频率较高, 两者表现出不同的二级结构倾向性. 这种局部结构倾向性和静电相互作用孰强孰弱? 比如由天冬氨酸和精氨酸组成的无序多肽的构象与电荷排布之间是否依然服从同样的物理规律? 为了回答这一问题, 我们系统研究了由精氨酸和天冬氨酸组成的链长为50的随机多肽体系. 基于ABSINTH (self-Assembly of biomolecules studied by an implicit, novel, and tunable hamiltonian) 模型^[32], 利用全原子蒙特卡罗模拟方法研究了随机多肽的构象特征. 结果表明, 当两种正负带电氨基酸混合均匀时, 多肽倾向于形成扩展的无规卷曲形状; 当正负电荷显著分离时, 多肽形成的结构为紧密的类 β -发卡形状. 这反映了电中性的随机多肽中正负带电氨基酸的排布显著地影响着其结构特征, 进一步展示了静电相互作用是离子型多肽体系中最主要的相互作用.

2 模型与方法

2.1 模式参数 κ 的定义

本研究中, 以带正电的精氨酸 Arg 和带负电的天冬氨酸 Asp 为基本组成成分, 产生了一系列链长 $l = 50$ 的随机序列 (共 30 条). 这里带电性质的选择对应生理条件的情形. 对每一条多肽的 N 端和 C 端, 分别利用乙酰基 (ACE) 和 N 甲基 (NME) 进行保护. 这里仅对 NCPR = 0 的序列进行研究, 具体的序列清单如表 1 所列. 这些序列涵盖了模式参数 κ 整个范围 (0, 1). 序列的模式参数的定义如下^[6]:

$$\sigma = \frac{(f_+ - f_-)^2}{(f_+ + f_-)^2}, \quad (1)$$

$$\sigma_i = \frac{(f_+ - f_-)_i^2}{(f_+ + f_-)_i^2}, \quad (2)$$

$$\Delta = \frac{1}{N_{\text{blob}}} \sum_{i=1}^{N_{\text{blob}}} (\sigma_i - \sigma)^2, \quad (3)$$

$$\kappa = \frac{\Delta}{\Delta_{\text{max}}}, \quad (4)$$

这里 f_+ 和 f_- 分别表示带正、负电氨基酸残基的百分比. 对每一条序列, 可以把序列划分为 N_{blob} 个不重复的片段来计算模式参数, 每个片段可以包含 $g = 5$ 或者 $g = 6$ 个氨基酸残基^[33]. 对第 i 个片段电荷不对称性 σ_i 定义为 (2) 式, 序列的整体电荷不对称性 σ 的方差 Δ 定义为 (3) 式. 不同序列的电荷不对称性的方差 Δ 是不同的, 其中正负电荷完全分离的状态方差最大 (记为 Δ_{max}). 其余序列的模式参数 κ 定义为该序列的方差 Δ 除以 Δ_{max} . 当 Arg 和 Asp 完全分离时, κ 为 1; 当 Arg 和 Asp 混合均匀时, κ 接近于零.

上述公式的物理意义为: (1) 式表示整条序列的电荷不对称性; (2) 式表示第 i 个以长度为 5 或者 6 个残基为单位时片段的电荷不对称性; (3) 式表示电荷不对称性的方差; (4) 式反映了一条序列中正负电荷混合均匀程度.

2.2 ABSINTH 模型

ABSINTH 模型是一种成功用于刻画内禀无序多肽体系的隐式溶剂模型^[32], 形式上可以分解为

$$U_{\text{total}} = U_{\text{tor}} + U_{\text{Disp}} + U_{\text{EV}} + W_{\text{Solv}} + W_{\text{el}}, \quad (5)$$

其中 U_{tor} 表示二面角扭转势能; $U_{\text{Disp}} + U_{\text{EV}}$ 对应于修正的范德瓦尔斯相互作用势, U_{Disp} 和 U_{EV} 分别表示短程吸引项和排斥项; W_{Solv} 为 ABSINTH 模型中的平均场相互作用; W_{el} 表示静电相互作用的贡献.

本研究中, 为了比较电荷排布情况对多肽结构的调控, 考虑了一些极限情况: 体积排斥极限 (excluded volume limit, EV)、球形极限 (Lennard-Jones, LJ)、随机线团极限 (flory random coils, FRC) 和完整势能 (ABSINTH) 模型. 其中 FRC 极限去除了 W_{el} ; EV 极限只打开体积排斥项 U_{EV} ; LJ 极限则仅保留 Lennard-Jones 势函数 $U_{\text{Disp}} + U_{\text{EV}}$ ^[6].

基对之间距离的平均结果:

$$\langle R_{ij} \rangle = \frac{1}{Z_{|i-j|}} \sum_{i,j,|i-j|=\text{const}} \left\langle \frac{1}{Z_{ij}} \sum_{m \in i} \sum_{n \in j} |r_m - r_n| \right\rangle, \quad (7)$$

这里, r_m 和 r_n 分别代表第 m 个原子和第 n 个原子的位置矢量, 分别属于第 i 号和第 j 号残基; Z_{ij} 表示两个残基之间的所有的独立的配对数目; 序列距离 $|i-j|$ 的范围是 0 到 $l+2$; $Z_{|i-j|}$ 表示求和平均归一化常数.

2.4 模拟方法

模拟采用软件包 CAMPARI2.0^[31] 实现, 应用 ABSINTH 模型^[32] 和 OPLS-AA/L 力场. 为了提高采样的效率, 利用温度副本交换采样方法 (replica-exchange Monte Carlo, REMC)^[33-41] 和马尔可夫链蒙特卡罗方法 (Markov chain metropolis Monte Carlo, MCMC). 模拟的盒子为球形, 半径为 75 Å. 采用基于原子的软球势边界条件. 为了考虑溶剂离子的影响, 本研究模拟了浓度为 15 mM 的氯化钠溶液. 这一数值和生理浓度 (150 mM) 有差别. 不过前人研究指出氯化钠的浓度对于谷氨酸和赖氨酸组成的随机多肽体系的结构特征没有显著影响^[6]. 因此, 本文选择 15 mM 的氯化钠浓度是合理的. 氨基酸残基原子间 Lennard-Jones 势和静电相互作用的截断距离分别设置为 10 Å 和 14 Å. 溶液中盐离子和侧链部分净电荷之间的静电相互作用则没有采取截断设置^[6].

模拟过程的温度副本交换方法 (REMC) 选择了从 280 K 到 430 K 的 12 个温度副本. 模拟的总步数为 4.65×10^7 步, 相邻温度副本间每隔 5×10^4 步交换一次构象. 在统计分析时, 将模拟初始 1.5×10^6 步舍掉. 在下文中, 仅仅对常温 (298 K) 这一温度副本的情况进行了系统讨论.

3 结果与讨论

3.1 回转半径的变化反映了多肽的构象转变

图 1 显示了 (Arg-Asp)₂₅ 不同序列的平均回转半径 $\langle R_g \rangle$ 随着模式参数 κ 的变化趋势, 图 2 给出了 (Arg-Asp)₂₅ 系统的代表构象. 总体来看, 平均回转半径 $\langle R_g \rangle$ 随着模式参数 κ 的增大而减小. 也就是

随着正负电荷的分离, 多肽的模式参数增大, 多肽由松散的结构转变到紧密堆积. 我们对回转半径与

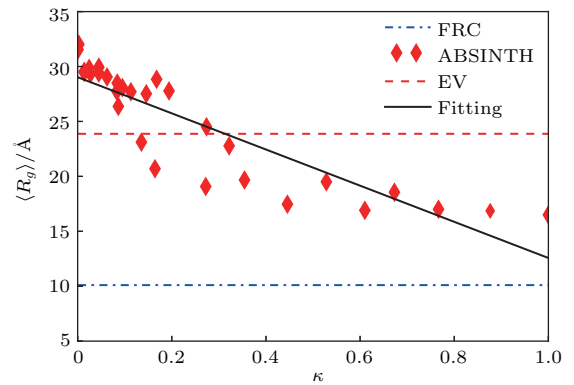


图 1 (Arg-Asp)₂₅ 不同序列回转半径 $\langle R_g \rangle$ 与模式参数 κ 的关系 (EV 和 FRC 分别表示体积排斥极限和 Flory 随机线团极限情况, ABSINTH 表示一般情况, 黑色的实线为回转半径与模式参数的线性拟合)

Fig. 1. Ensemble-averaged radii of gyration $\langle R_g \rangle$ for sequence variants of the (Arg-Asp)₂₅ system. Two dash lines intersect the ordinate at $\langle R_g \rangle$ values expected for all sequence variants modeled in the EV limit or as Flory random coils (FRC). The solid black line is the fitting line for radii of gyration and patterning parameter.

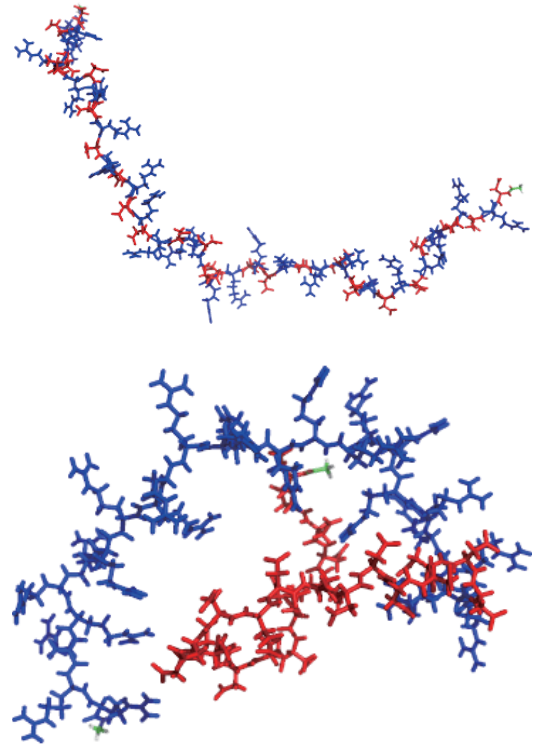


图 2 (Arg-Asp)₂₅ 系统, $\kappa = 0.0009$ 和 $\kappa = 1.0000$ 时的代表构象 (红色的表示带负电荷的天冬氨酸, 蓝色的表示带正电荷的精氨酸)

Fig. 2. Representative conformation for two sequence variants of the (Arg-Asp)₂₅ system: up ($\kappa = 0.0009$), down ($\kappa = 1.0000$). Side chains of Asp are shown in red, and side chains of Arg are shown in blue.

模式参数进行了线性拟合,发现皮尔森线性相关系数 $r = -0.89$, 显著性为 $P = 8.1 \times 10^{-11}$. sv30 多肽的回转半径的最小值约为 16 \AA , 这种情况对应的模式参数 κ 趋近于 1. 当正负电荷混合均匀 (κ 趋近于 0) 时, 序列 sv1 的回转半径的值约为 32 \AA . 这说明了模式参数越小, 多肽的结构越松散, 结构类似于自回避随机行走结构; 模式参数越大, 多肽的结构越紧密, 结构类似于密堆积球形. 图 1 还给出了两种极限情况 (EV 极限和 FRC 极限). 两者对应的多肽回转半径分别约为 24 \AA 和 10 \AA . 这两种极限情形中, 多肽的结构显然不依赖于序列中电荷的分布. FRC 极限显然给出了多肽回转半径的下限.

3.2 链末端距的变化反映多肽的结构转变

图 3 给出了链末端距 $\langle R_{ee} \rangle$ 与模式参数 κ 之间的关系. 可以看到, 随着模式参数增大, 多肽链末端距呈现出逐渐减小的趋势. 对正负电荷完全混合均匀的情况, 链末端距约为 84 \AA , 而正负电荷完全分离的情况对应的链末端距约为 30 \AA . 同样, 链末端距随着模式参数的变化趋势和图 1 显示的结论是一致的. 这一结果表明, 随着正负电荷的分离, 多肽结构由扩展的直杆状逐渐转变到紧密的球形结构.

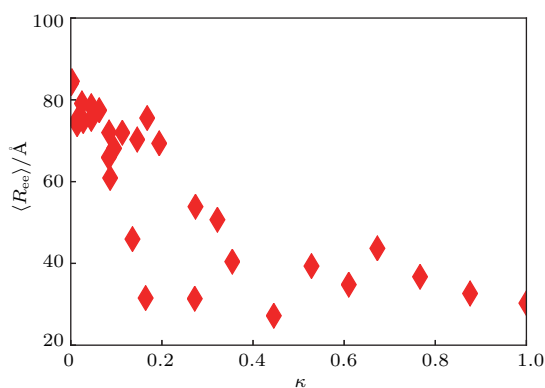


图 3 (Arg-Asp)₂₅ 系统各种序列对应的多肽的链末端距与模式参数 κ 的关系

Fig. 3. Ensemble-averaged end-to-end distance $\langle R_{ee} \rangle$ for sequence variants of the (Arg-Asp)₂₅ system.

3.3 内部平均距离 $\langle R_{ij} \rangle$ 的变化反映了多肽结构的转变

图 4 给出了内部平均距离 $\langle R_{ij} \rangle$ 与序列距离 $|i - j|$ 之间的关系曲线. 这一曲线可以很好地反映多肽结构的转变. 具体地, sv1 序列对应的曲线 (紫色菱形), 内部平均距离 $\langle R_{ij} \rangle$ 随着氨基酸序列

距离 $|i - j|$ 增大而逐渐增长, 反映多肽形成了扩展或者松散的结构 (类似于杆状). 这一情形对应 $\kappa = 0.0009$, 其中正负电荷混合均匀, 静电相互作用相互抵消. sv30 对应的曲线 (红色菱形), $\langle R_{ij} \rangle$ 表现出先增大再减小而后又增大的变化趋势, 表明多肽形成了具有弯折形状的紧密结构. 这一情形对应 $\kappa = 1.0000$, 其中正负电荷显著分离, 两者相互吸引而形成紧密结构. 这一趋势也反映了电荷分布变化导致多肽结构由松散结构向紧密结构转变的过程. 图 4 还给出了 EV 极限 (红色实线), FRC 极限 (黑色实线) 和 LJ 极限 (蓝色实线) 的情形. 很容易发现 $\kappa = 0.0009$ 时多肽结构比 EV 极限情况更加松散, 而 $\kappa = 1.0000$ 时, 多肽结构趋近于 LJ 极限.

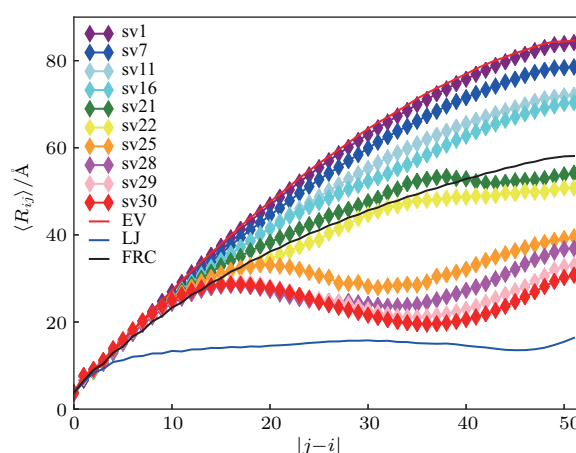


图 4 (Arg-Asp)₂₅ 系统各种序列对应的多肽的内部平均距离 $\langle R_{ij} \rangle$ 与模式参数 κ 的关系 (红色实线对应 EV 极限情形, 黑色实线表示 FRC 极限情形, 蓝色实线表示 LJ 极限情形)

Fig. 4. $\langle R_{ij} \rangle$ profiles for the (Arg-Asp)₂₅ sequence variants. The solid red, black and blue curves denote the profile expected for (Arg-Asp)₂₅ polymers in the EV limit, in FRC limit, and in LJ limits.

3.4 非球形指数的变化展示了从非球形结构向球形结构的转变

图 5 给出了非球形指数 δ^* 与模式参数 κ 的关系曲线, 可以看出随着模式参数 κ 的增加, 非球形指数从 0.6 下降到 0.2. 理论上, 非球形指数 $\delta^* = 1$ 表示理想的杆状结构, $\delta^* = 0$ 表示理想的球形结构. 非球形指数的变化反映了随着电荷的分离, 多肽紧密球形结构的形成过程. 伴随着正负电荷的分离, 多肽左右两端形成较大的电荷积聚区域, 在静电吸引的作用下, 类似球形的结构将逐渐形成. 非球形指数的变化趋势进一步反映了多肽从松散结构到紧密结构的转变.

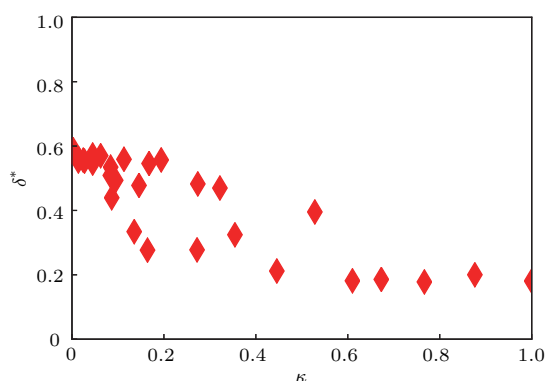


图5 (Arg-Asp)₂₅ 系统各种序列对应的多肽的非球形指数 δ^* 与模式参数 κ 的关系

Fig. 5. Plot of ensemble-averaged asphericity values δ^* against κ for sequence variants of (Arg-Asp)₂₅.

3.5 多肽结构转变规律的保守性

选择天冬氨酸和精氨酸组成的随机多肽为研究对象, 在组分上与 Pappu 等研究的谷氨酸和赖氨酸组成的随机多肽不同. 在蛋白质中, 天冬氨酸与谷氨酸形成二级结构的倾向性有着差别. 谷氨酸容易形成 α -螺旋结构, 但是天冬氨酸形成 β -转角的倾向性比较高^[42]. 但是, 两类多肽的电荷分布性质和结构特征之间的关系却是完全相似的. 这说明对这种富电荷体系, 在多肽结构转变过程中静电相互作用占主导地位. 因多肽中全部是带电的氨基酸, 没有疏水氨基酸, 多肽系统缺乏折叠为特定二级结构的相互作用. 相反, 正负电荷分离引起明显的静电吸引, 导致多肽形成包含弯折结构的紧密结构. 不过这种结构依然处于无序状态^[8,10,11]. 因此, 由于正负电荷排布属性引起的结构变化规律是相似的.

4 结 论

从两种代表性正负带电氨基酸组分出发, 构建随机多肽序列, 利用全原子模型、温度副本交换及蒙特卡罗模拟方法, 获得了电荷排布情况如何调控随机多肽结构的规律. 着重考察了强两性电解质系统, 发现电荷的排布情况会调控多肽的构象. 结果证实了随着正负电荷的分离, 系统中出现了较强的静电相互吸引, 导致了紧密球形结构的形成. 我们的结论与 Pappu 等^[6]的结果符合得很好. 值得指出, 目前的研究仅仅是对强两性电解质系统的研究, 对含有其他疏水性氨基酸情况的研究需要进一步探索. 搞清多肽序列与结构的一般性关系, 对揭示多肽的功能有很大的帮助, 同时厘清多肽序列结

构形成过程中的物理规律, 对设计或者蛋白质工程有着重要的指导意义^[43–52].

感谢南京大学高性能计算中心的支持, 感谢唐乾元博士的启发性讨论.

参考文献

- [1] Tantos A, Han K H, Tompa P 2012 *Mol. Cell. Endocrinol.* **348** 457
- [2] Dyson H J, Wright P E 2005 *Nat. Rev. Mol. Cell Biol.* **6** 197
- [3] Uversky V N 2002 *Eur. J. Biochem.* **69** 2
- [4] Das R K, Pappu R V 2013 *Proc. Natl. Acad. Sci. USA* **110** 13392
- [5] Yu J F, Dou X H, Sha Y J, Wang C L, Wang H B, Chen Y T, Zhang F, Zhou Y, Wang J H 2017 *BMC Bioinform.* **18** 206
- [6] Piovesan D, Tabaro F, Micetic I, et al. 2017 *Nucl. Acids Res.* **45** D1123
- [7] Potenza E, Di Domenico T, Walsh I, Tosatto S C 2015 *Nucl. Acids Res.* **43** D315
- [8] Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker A K, Felli I C, Forman-Kay J D, Kriwacki R W, Pierattelli R, Sussman J, Svergun D I, Uversky V N, Vendruscolo M, Wishart D, Wright P E, Tompa P 2014 *Nucl. Acids Res.* **42** D326
- [9] Sickmeier M, Hamilton J A, LeGall T, Vacic V, Cortese M S, Tantos A, Szabo B, Tompa P, Chen J, Uversky V N, Obradovic Z, Dunker A K 2007 *Nucl. Acids Res.* **35** D786
- [10] Sim K L, Uchida T, Miyano S 2001 *Bioinformatics* **17** 379
- [11] Forbes J G, Jin A J, Ma K, Gutierrez-Cruz G, Tsai W L, Wang K 2005 *J. Muscle Res. Cell Motil.* **26** 291
- [12] Uversky V N 2002 *Protein Sci.* **11** 739
- [13] Mao A H, Crick S L, Vitalis A, Chicoine C L, Pappu R V 2010 *Proc. Natl. Acad. Sci. USA* **107** 8183
- [14] Kumar S, Hoh J H 2004 *Biochem. Biophys. Res. Commun.* **324** 489
- [15] Hendus-Altenburger R, Lambrugh M, Terkelsen T, Pedersen S F, Papaleo E, Lindorff-Larsen K, Kragelund B 2017 *Cell. Signal* **37** 40
- [16] Malka-Gibor E, Kornreich M, Laser-Azogui A, Doron O, Zingerman-Koladko I, Harapin J, Medalia O, Beck R 2017 *Biophys. J.* **112** 892
- [17] Khan S H, McLaughlin W A, Kumar R 2017 *Sci. Rep.* **7** 15440
- [18] Lousa P, Nedozralova H, Zupa E, Novacek J, Hritz J 2017 *Biophys. Chem.* **223** 25
- [19] Stakkestad O, Lyngstadaas S P, Thiede B, Vondrasek J, Skallehegg B S, Reseland J E 2017 *Front. Physiol.* **8** 531
- [20] Liu J J, Dai J, He J F, Niemi A J, Ilieva N 2017 *Phys. Rev. E* **95** 032406
- [21] Das R K, Huang Y, Phillips A H, Kriwacki R W, Pappu R V 2016 *Proc. Natl. Acad. Sci. USA* **113** 5616

- [22] Lange J, Wyrwicz L S, Vriend G 2016 *Bioinformatics* **32** 932
- [23] Arya S, Mukhopadhyay S 2014 *J. Phys. Chem. B* **118** 9191
- [24] Kister A E, Potapov V 2013 *Biochem. Soc. Trans.* **41** 616
- [25] Hoang T X, Trovato A, Seno F, Banavar J R, Maritan A 2012 *Phys. Rev. E* **86** 050901
- [26] Huang Y Q, Liu Z R 2010 *Acta Phys. Chim. Sin.* **26** 2061 (in Chinese) [黄永棋, 刘志荣 2010 物理化学学报 **26** 2061]
- [27] Dunker A K, Lawson J D, Brown C J, Williams R M, Romero P, Oh J S, Oldfield C J, Campen A M, Ratliff C M, Hipps K W, Ausio J, Nissen M S, Reeves R, Kang C, Kissinger C R, Bailey R W, Griswold M D, Chiu W, Garner E C, Obradovic Z 2001 *J. Mol. Graph. Model.* **19** 26
- [28] Wang J, Wang W 1999 *Nat. Struct. Biol.* **6** 1033
- [29] Li W F, Qin M, Tie Z X, Wang W 2011 *Phys. Rev. E* **84** 041933
- [30] Wang J, Wang W 2016 *Adv. Phys. X* **1** 444
- [31] Vitalis A, Pappu R V 2009 *Annu. Rep. Comput. Chem.* **5** 49
- [32] Vitalis A, Pappu R V 2009 *J. Comput. Chem.* **30** 673
- [33] Pappu R V, Wang X, Vitalis A, Crick S L 2008 *Arch. Biochem. Biophys.* **469** 132
- [34] Cragnell C, Durand D, Cabane B, Skepo M 2016 *Proteins* **84** 777
- [35] Venditto J G, Wolf S, Curotto E, Mella M 2015 *Chem. Phys. Lett.* **635** 127
- [36] Wu H H, Chen C C, Chen C M 2012 *J. Comput. Aided Mol. Des.* **26** 363
- [37] Liu Y, Kellogg E, Liang H J 2012 *J. Chem. Phys.* **137** 045103
- [38] Odriozola G, Berthier L 2011 *J. Chem. Phys.* **134** 054504
- [39] Turner C H, Brennan J K, Lital M 2007 *J. Phys. Chem. C* **111** 15706
- [40] Kokubo H, Okamoto Y 2004 *J. Chem. Phys.* **120** 10837
- [41] Nakazawa T, Ban S, Okuda Y, Masuya M, Mitsutake A, Okamoto Y 2002 *Biopolymers* **63** 273
- [42] Uversky V N 2013 *Intrinsically Disordered Proteins* **1** e24684
- [43] Childers M C, Towse C L, Daggett V 2016 *Protein Eng. Des. Sel.* **29** 271
- [44] Yu C, Niu X, Jin F, Liu Z, Jin C, Lai L 2016 *Sci. Rep.* **6** 22298
- [45] Guharoy M, Bhowmick P, Tompa P 2016 *J. Biol. Chem.* **291** 6723
- [46] Nagibina G S, Tin U F, Glukhov A S, Melnik T N, Melnik B S 2016 *Protein Pept. Lett.* **23** 176
- [47] Noivirt-Brik O, Horovitz A, Unger R 2009 *PLoS Comput. Biol.* **5** e1000592
- [48] Cheng Y, LeGall T, Oldfield C J, Mueller J P, Van Y Y, Romero P, Cortese M S, Uversky V N, Dunker A K 2006 *Trends Biotechnol.* **24** 435
- [49] Ambroggio X I, Kuhlman B 2006 *Curr. Opin. Struct. Biol.* **16** 525
- [50] Meng G Z 1986 *Prog. Biochem. Biophys.* **13** 3 (in Chinese) [孟广震 1986 生物化学与生物物理进展 **13** 3]
- [51] Wang D C 2008 *Protein Engineering* (Vol. 1) (Beijing: Chemical Industry Press) p65 (in Chinese) [王大成 2008 蛋白质工程 (北京: 化学工业出版社) 第65页]
- [52] Deng H Y, Jia Y, Zhang Y 2016 *Acta Phys. Sin.* **65** 178701 (in Chinese) [邓海游, 贾亚, 张阳 2016 物理学报 **65** 178701]

Conformation of disordered peptides modulated by distributions of charged residues: Case study of random peptides composed of arginines and aspartic acids*

Kang Wen-Bin^{1)2)3)†} Wang Jun^{1)‡} Wang Wei^{1)††}

1) (School of Physics, Nanjing University, Nanjing 210093, China)

2) (Department of Mathematics and Physics, Hubei University of Medicine, Shiyan 442000, China)

3) (Bio-X Research Center, Hubei University of Medicine, Shiyan 442000, China)

(Received 17 October 2017; revised manuscript received 11 December 2017)

Abstract

The relationship between the sequential and structural features of intrinsically disordered peptides (IDPs) has attracted much attention during the recent decade. One essential problem relating to sequence-structure relationship is how the distribution of charged residues affects the structure of IDP. In this work, we address this problem with simulations on a series of random peptides composed of arginine and aspartic acids. With the ABSINTH implicit solvation model, the structural ensembles are generated with Markov Chain Monte Carlo method and replica-exchange sampling. The relations between various structural features (including the gyration radius, the tail distance, the distance between residues, and asphericity) and the distribution of charged residues are analyzed. Several limit cases (with parts of interactions switched off) are also calculated for comparison. The conversion from extended conformations to compact structures is observed, following the demixing of negatively and positively charged residues along the sequence. For the cases with well-mixed charges, the intra-chain electrostatic repulsions and attractions are balanced, which results in a generic Flory random coil-like conformation. Differently, for the case with well-separated charged residues, the electrostatic attraction between residues distant along the sequence induces a semi-compact hairpin-like conformation. This is consistent with the observations of Pappu group. Our results suggest that the structural dependence on charge distribution would not be sensitive to the selection of amino acid, and is determined by the patterns of charges, which demonstrates the robustness of the mechanism that the charge distribution modulates the structural features in the IDP system. Our results may broaden our understanding of the sequence-structure relation of IDP system.

Keywords: intrinsically disordered protein, structural ensemble, sequence, molecular simulation

PACS: 87.10.Rt, 87.14.ef, 87.15.A-, 87.15.Qt

DOI: 10.7498/aps.67.20172246

* Project supported by the National Basic Research Program of China (Grant No. 2013CB834100), the National Natural Science Foundation of China (Grant Nos. 11334004, 11174133, 81421091, 11774157), the 2017 National Training Program of Innovation and Entrepreneurship for Undergraduates, China (Grant No. 201710929002), and the Initial Project for Post-Graduates of Hubei University of Medicine, China (Grant No. 2011QDZR-11).

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

‡ Corresponding author. E-mail: wangj@nju.edu.cn

†† Corresponding author. E-mail: wangwei@nju.edu.cn