



生物化学与生物物理进展
Progress in Biochemistry and Biophysics
ISSN 1000-3282, CN 11-2161/Q

《生物化学与生物物理进展》网络首发论文

题目: 基于 GNM 方法研究人 TDP-43-DNA 相互作用动力学及关键位点
作者: 邓雪晴, 王世豪, 巩卫康, 李春华
收稿日期: 2021-04-29
网络首发日期: 2021-06-09
引用格式: 邓雪晴, 王世豪, 巩卫康, 李春华. 基于 GNM 方法研究人 TDP-43-DNA 相互作用动力学及关键位点. 生物化学与生物物理进展.
<https://kns.cnki.net/kcms/detail/11.2161.Q.20210609.1302.004.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于 GNM 方法研究人 TDP-43-DNA 相互作用 动力学及关键位点*

邓雪晴、王世豪、巩卫康、李春华**

(北京工业大学环境与生命学部, 北京 100124)

摘 要 TAR DNA 结合蛋白-43 (Transactive response DNA binding protein 43, TDP-43), 一种可变剪切因子, 可以特异性地结合富含 TG 序列的 DNA, 涉及多种神经退行性疾病。分子动力学模拟方法虽然是研究分子间相互作用强有力的工具, 但它非常耗时, 且难以对大的构象变化的体系进行充分采样来研究其变构行为。在本工作中, 我们使用粗粒化的基于弹性势的高斯网络模型(Gaussian network model, GNM) 研究了人 TDP-43 与靶标 DNA 间相互作用的动力学, 结果发现 DNA 结合后, TDP-43 上富含正电残基的 loop1 和 loop3 片段有较大的柔性损失, 这反映了它们在识别和结合中的诱导契合作用; 进一步, 利用小组之前提出的基于 GNM 的热力学循环方法识别了 TDP-43 与 DNA 相互作用的关键残基, 其微扰引起了大的结合自由能的变化。我们发现, 该方法不仅识别到了一些与 DNA 特异性相互作用有关的重要残基, 而且识别到了一些远离结合界面但在结合引起的分子构象变化中发挥重要作用的残基。本研究有助于理解 TDP-43 与 DNA 的特异性相互作用, 可为药物设计提供重要信息, 另外该方法可以很方便地拓展到其他蛋白质-核酸相互作用动力学的研究。

关键词: TDP-43-DNA 相互作用; 高斯网络模型; 关键残基; 相互作用动力学

学科分类号 Q61

*基金项目: 国家自然科学基金(31971180, 11474013)资助项目

**联系人: 李春华, Tel.: +86-010-67392001, E-mail: chunhuali@bjut.edu.cn

收稿日期: 2021-04-29, 接受日期: 2021-06-07

TAR DNA 结合蛋白-43 (Transactive response DNA-binding protein 43, TDP-43) 是一种可变剪切因子, 在 DNA 转录和 RNA 翻译过程中发挥着重要作用。研究发现细胞内 TDP-43 的异常聚集会影响 DNA 的结合, 进而导致多种疾病^[1]包括额颞叶变性(frontotemporal lobar degeneration, FTL D)、肌萎缩性侧索硬化症(amyotrophic lateral sclerosis, ALS)和阿尔兹海默病(Alzheimer's disease, AD) 等的发生^[2-4]。从结构上看, TDP-43 含有两个 RNA 识别基序(RNA recognition motif, RRM), 它们特异性地识别并结合富含 UG 序列的 RNA 片段或富含 TG 序列的 DNA 片段^[5], 其中 RRM1 在核酸特异性识别中扮演着主要作用。Kuo 等人研究发现, 人 TDP-43 与靶标 DNA 的结合会通过结合界面对 TDP-43 的结构动力学产生扰动, 进而引起两个 RRM s 区域发生构象重排, 从而增强了分子间的结合亲和性^[6-8]。另外, TDP-43 的某些位点突变会导致其与 DNA 结合亲和性和特异性发生显著变化。因此, 探索 TDP-43 与 DNA 之间的特异性相互作用, 并识别在结合过程中发挥关键作用的残基, 对于深入理解其相互作用机制, 以及开发以 TDP-43 为靶点的药物具有重要意义。

目前, 实验人员已经通过 X-ray 方法获得了人 TDP-43 的 RRM1 与 DNA 结合的复合物结构(PDB code: 4IUF), 如图 4(B)所示。TDP-43 的 RRM1 是由 N 端、RRM 区域以及 linker 区域构成的。通常, RRM 含有 4 个反向平行排列的 β 片层, 比如 hnRNP-A1^[9]和 Hud^[10]中的 RRM, 但 TDP-43 的 RRM 却是具有 $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4-\beta 5$ 的二级结构排列, 包含一个额外的 $\beta 4$ 片层。实验表明该 β 片层扩展了 TDP-43 与 DNA 结合的界面, 有助于加强分子间的结合亲和性。TDP-43 具有一个带正电的结合凹槽, 其对 DNA 的结合很重要, 主要由 β 片层、loop 和 linker 构成; RRM1 的 loop1 和 loop3 伸向 DNA, 并与 DNA 形成静电和氢键相互作用; β 片层构成了主要的 DNA 结合界面, 其上有两个高度保守的片段 RNPs (Ribonucleoprotein domains): RNP1 (Lys145-Phe152) 主要位于 $\beta 3$, RNP2 (Leu106-Leu111) 主要位于 $\beta 1$ 。

在实验方面, 研究主要集中在探索 TDP-43 与 DNA 相互作用的关键残基上。Baralle 等人利用电泳迁移率 (electrophoretic mobility shift assay, EMSA) 进行研究, 发现 RRM1 和 RRM2 在与 DNA 的识别中分别起主要和辅助结合的作用^[11]。之后, 研究者通过交联和免疫沉淀分析(cross-linking and immunoprecipitation assay, CLIP) 研究发现, 除了 β 片层以外, 连接 β 片层和 α 螺旋的 loop1 和 loop3 对 DNA 的识别也非常关键, 它们主要发挥了特异性识别的作用^[12]。在理论研究方面, 目前对 TDP-43 与 DNA 之间相互作用的研究还不多, 只有几个关于 TDP-43 自身结构动力学研究的工作。Chiang 通过分子动力学 (molecular dynamics, MD) 模拟发现, 位点突变 Asp169Lys 引起了 RRM1 的连接 $\beta 4$ 和 $\beta 5$ 的 β 转角附近构象的变化, 这增加了 RRM1 核心的疏水相互作用^[13]。Prakash 等人通过 MD 模拟研究了 TDP-43 的去折叠过程, 观察到 RRM1 和 RRM2 具有完全不同的去折叠路径^[14]。

MD 模拟, 尽管能够在原子水平洞察分子动力学行为与功能的关系, 但其耗时耗力, 且难以对有大构象变化的体系进行充分采样以研究变构行为。针对此问题, 人们提出了很多

粗粒化模型的方法, 其中弹性网络模型(Elastic Network Model, ENM) 是一种被广泛用于研究生物大分子功能性变构的有效方法^[15, 16]。高斯网络模型 (Gaussian network model, GNM) 和各向异性网络模型 (anisotropic network model, ANM) 是两种常用的弹性网络模型。ENM 得到的低频运动模式常常代表了与分子功能相关的大尺度集合运动。最初, Tirion 等人提出用弹性势代替分子内部的相互作用来构建 ENM 模型, 其中, 一定截断半径内的所有残基对被认为有相互作用, 并用具有相同力常数的弹簧连接^[17]。Bahar 等人利用 GNM 研究了 tRNA 的动力学, 捕获到了不同结构域间功能性的协同运动^[18]。另外, 基于 ENM 模型, 一些微扰方法被提出用于识别生物大分子与功能有关的重要残基。Atilgan 等人提出线性微扰扫描方法, 来预测受体与配体结合中的关键残基, 其中外力微扰导致整个结构涨落变化较大的残基被认为是关键残基^[19]。Brooks 等人提出了改变弹簧系数的微扰方法, 来识别蛋白质变构中的铰链残基^[20]。我们组 Su 等曾提出了基于 GNM 的热力学循环方法, 来识别在蛋白质-配体结合中发挥关键作用的残基, 其中位点微扰导致其结合自由能较大变化的残基被认为是关键残基^[21]。前两种方法是通过施加外力微扰, 根据其对整个体系涨落的影响来识别关键残基, 而后一种方法是通过评价位点微扰 (模拟突变) 导致的受体-配体结合自由能的变化来识别关键残基, 该方法不仅能够识别与配体结合有关的重要残基, 还能识别与配体结合偶联的变构中的关键残基。

在本研究中, 我们基于 GNM 探索了 DNA 结合对 TDP-43 动力学的影响。此外, 利用基于 GNM 的热力学循环方法识别了 TDP-43 中与 DNA 特异性识别以及 DNA 结合引起的构象变化有关的重要残基。

1 方法和体系

1.1 高斯网络模型(GNM)

GNM 是一种粗粒化的方法, 能够获得生物分子的低频运动, 可以用于研究与功能有关的分子动力学行为。GNM 将生物分子结构抽象为一个粗粒化的弹性网络, 其中残基简化为节点 (这里氨基酸用 C_α 原子, 核苷酸用 P 原子代替), 一定距离 (R_c) 范围内的节点间用弹性系数 γ 相同的弹簧相连。简化后, 整个体系的势能为:

$$V = \frac{1}{2} \gamma [\Delta \mathbf{R}^T (\mathbf{\Gamma} \otimes \mathbf{E}) \Delta \mathbf{R}] \quad (1)$$

其中, $\Delta \mathbf{R}$ 是一个 N 维的列向量, 向量的每个元素表示节点偏离平衡位置的位移, T 代表转置, \mathbf{E} 是单位矩阵, \otimes 代表直积。 $\mathbf{\Gamma}$ 是 GNM 的 Kirchhoff 矩阵

每个节点的均方涨落以及不同节点间涨落的交叉相关性表示为:

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \frac{3k_B T}{\gamma} [\mathbf{\Gamma}^{-1}]_{ii}, \quad \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} [\mathbf{\Gamma}^{-1}]_{ij} \quad (2)$$

根据 Debye-Waller 理论, 分子内原子涨落的 B-factor 为该节点涨落的 $8\pi^2/3$ 。

1.2 模型评价指标

对于模型中的参数, 我们通过最大化理论与实验 B-factor 的皮尔逊相关系数(Pearson correlation coefficient, PCC) 来确定。两组数据完美正 (负) 相关 PCC 为 1 (-1), 不相关为 0。

1.3 基于 GNM 的热力学循环方法

基于 GNM 的热力学循环方法可以识别与结合和变构有关的关键残基, 其中位点突变导致受体-配体结合自由能变化较大的残基为关键残基。如图 1 所示, ΔG_1 是 TDP-43 与 DNA 的结合自由能, ΔG_2 是 TDP-43 中某一个残基突变后, TDP-43' 与 DNA 的结合自由能。通过计算残基突变前后结合自由能的变化 $\Delta\Delta G$, 来识别对结合有较大影响的关键残基。

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 \quad (3)$$

由于 TDP-43 与 DNA 的相互作用非常复杂, 直接计算 ΔG_1 和 ΔG_2 是非常困难的。因此, 构建了热力学循环法来计算结合自由能的变化。如图 1 所示, 我们假设了两个非物理的过程, 分别是对 TDP-43 和复合物中的同一残基引入微扰 (模拟突变), 其中涉及到的自由能变化分别为 ΔG_3 和 ΔG_4 。

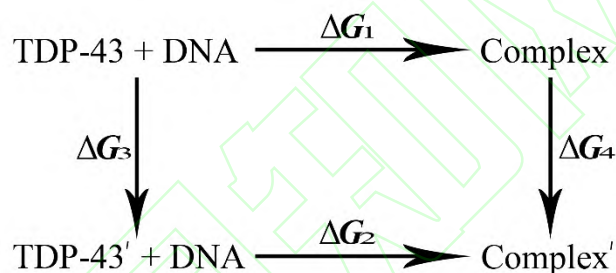


图 1. 热力学循环图。 ΔG_1 和 ΔG_2 分别是野生型和突变型 TDP-43 与 DNA 的结合自由能, ΔG_3 和 ΔG_4 分别是微扰 TDP-43 和复合物中的某一残基引起的自由能的变化。

Fig 1. Thermodynamic cycle diagram. ΔG_1 and ΔG_2 are the binding free energies between wild-type and mutant TDP-43 proteins and DNA respectively, and ΔG_3 and ΔG_4 are the free energy changes caused by a residue perturbation to TDP-43 and complex, respectively.

由于自由能是一个态函数, 只取决于初始和终态, 因此突变一个残基引起的结合自由能的变化可以表示为分别微扰复合物和 TDP-43 中的该残基引起的自由能变化的差值, 即 $\Delta\Delta G = \Delta G_4 - \Delta G_3$ 。在两个假想的微扰过程中, 与力常数微扰相关的势能变化是相同的, 所以 $\Delta\Delta G$ 只和熵变有关, 即

$$\Delta\Delta G = -T(\Delta S_4 - \Delta S_3) \quad (4)$$

当突变一个残基时, 所有与该残基连接的弹簧都会被微扰, 突变所导致的自由能的变化 $\Delta\Delta G$ 为这些弹簧被微扰所产生的自由能变化的总和。这样, 分别计算蛋白质单体和复合物中与突变残基相连的弹簧被微扰所产生熵变的总和, 即 ΔS_3 和 ΔS_4 , 然后得到 $\Delta\Delta G$ 。其中微扰某一对残基 i 和 j 之间的弹簧所引起的熵变为^[21, 22]:

$$\Delta S_{ij} = \frac{1}{2T} \left(\langle (\Delta R_i)^2 \rangle + \langle (\Delta R_j)^2 \rangle - 2 \langle \Delta R_i \cdot \Delta R_j \rangle \right) \Delta \gamma_{ij} \quad (5)$$

在识别关键残基的过程中,对每个残基都以上述相同的方式突变,没有进行特殊处理。另外,对于蛋白质,我们考虑了对残基涨落贡献超过 50% 的前 60 个慢运动模式,复合物的运动模式被投影到这 60 个模式上。

1.4 研究体系

人的 TDP-43 与 DNA 的复合物结构是由 X-ray 实验解析出来的, PDB code 为 4IUF。蛋白质受体的结构是从 TDP-43-DNA 的复合物结构中拆分出来的。

2 结果与讨论

2.1 基于 GNM 的理论与实验 B-factor 的比较

B-factor 的准确预测为理解生物分子动力学提供了一个良好的起点。对于构建的 GNM, 我们通过最大化理论与 X-ray 实验得到的 B-factor 间的 PCC 值来确定最优参数。传统 GNM 只有一个截断半径参数。对 TDP-43-DNA 复合物, 我们采用三个截断半径, 分别针对蛋白质内部、RNA 内部和界面处, 其寻优范围为: 1-25 Å, 步长为 1 Å。图 2 显示了在最优截断半径 (蛋白质内部、RNA 内部和界面处残基间的截断半径分别为 11、14 和 12 Å) 下, 由 GNM 得到的理论 B-factor 和 X-ray 实验的 B-factor。从图 2 可以看到, GNM 能较好地再现复合物 TDP-43-DNA 的柔性信息, 实验与理论 B-factor 的 PCC 值为 0.67。

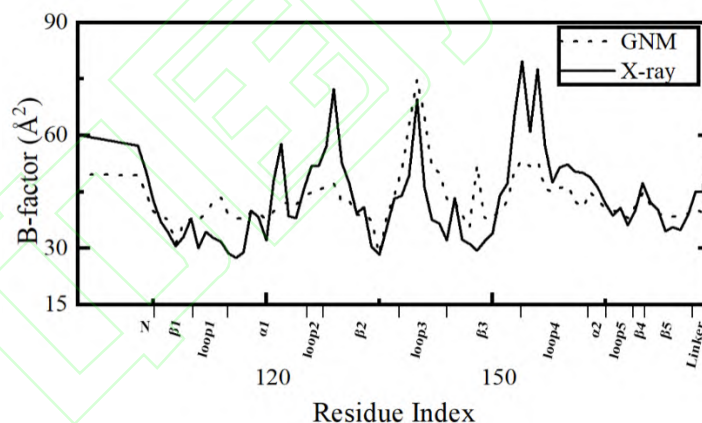


图 2. TDP-43-DNA 复合物的理论 (虚线) 和实验 X-ray (实线) B-factor 的比较。

Fig 2. Comparison between the theoretical (dotted line) and experimental X-ray (solid line) B-factors of TDP-43-DNA complex.

2.2 DNA 结合对蛋白质动力学的影响

为了探究 DNA 结合对蛋白 TDP-43 动力学的影响, 我们基于公式(2) 计算并比较了 TDP-43 在 DNA 结合和未结合状态下的残基的均方涨落, 结果展示在图 3 中。从图 3 可以看出, DNA 结合导致蛋白一些片段的柔性损失, 如 loop1、loop3 和 5 个 β 片层 $\beta 1$ - $\beta 5$ 。

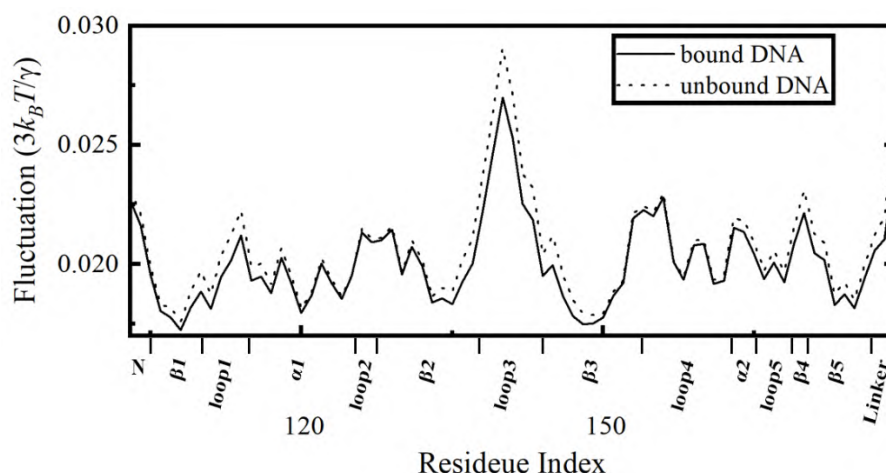


图 3. TDP-43 在未结合 DNA (虚线) 和结合 DNA (实线) 状态下的残基均方涨落。

Fig 3. Residue mean square fluctuation (MSF) of TDP-43 in DNA-unbound (dotted line) and DNA-bound (solid line) states.

从图 3 可以看出, DNA 结合导致 loop1 和 loop3 片段有较大的柔性损失, 这主要是由于 TDP-43 的 loop1 和 loop3 伸向 DNA, 与 DNA 形成了静电、堆叠和氢键相互作用^[6, 8, 23]。此外, loop1 和 loop3 中富含一些带正电荷的残基, 如 Lys114、Lys140 和 Lys145, 它们在吸引 DNA 中发挥了重要作用, 表明它们在 TDP-43-DNA 识别的诱导契合方面做出了重要贡献。

此外, DNA 结合导致 $\beta 1$ - $\beta 5$ 片段都有一定程度的柔性损失。 β 片层构成了与 DNA 结合的界面(图 4(B)), 其氨基酸残基与 DNA 核苷酸间的氢键和堆叠相互作用, 如 Asp105-G8、Ile107-C7、Met132-G8、Gln134-G8、Lys136-G8、Phe149-C7、Phe149-G8、Phe149-G4 和 Arg171-G4 在分子间特异性识别中发挥了关键作用^[8]。RNPs 是 β 片层上一段富含芳香族和疏水残基的高度保守的序列, 其中 RNP1 和 RNP2 (分别主要位于 $\beta 3$ 和 $\beta 1$ 上) 与 DNA 形成了特异性相互作用, 在 DNA 的特异性识别中发挥了重要作用^[8]。

总之, TDP-43 与 DNA 间相互作用的形成使得 loop1、loop3 和 $\beta 1$ - $\beta 5$ 片段发生了一定程度柔性的损失。loop1 和 loop3 对分子间静电吸引和诱导契合贡献了主要的作用, $\beta 1$ - $\beta 5$ (构成了与 DNA 结合的表面) 在 TDP-43 特异性识别 DNA 的过程中发挥了关键作用。

2.3 基于热力学循环识别 TDP-43 的关键位点

我们利用基于 GNM 的热力学循环方法识别了 TDP-43 上与 DNA 结合和构象变化有关的关键残基。对 TDP-43 上的残基进行微扰 (模拟突变), 导致结合自由能 $\Delta\Delta G$ 变化较大的残基被认为是关键残基簇的中心。

我们利用上述方法识别到了 8 个关键残基簇, 如图 4 (A) 所示, 分别以 Ile107、Lys114、Leu139、His143、Gly148、Asp169、Lys176 和 Asn179 为中心。为了方便展示, 我们将这些中心残基标注在结构上, 如图 4 (B) 所示。根据识别的中心残基在二级结构上的分布, 我们将其分为 β 、loop 和 linker 区域三个部分, 并将预测的关键残基与实验和理论数据进行了比

较, 下面将详细讨论它们的功能情况。

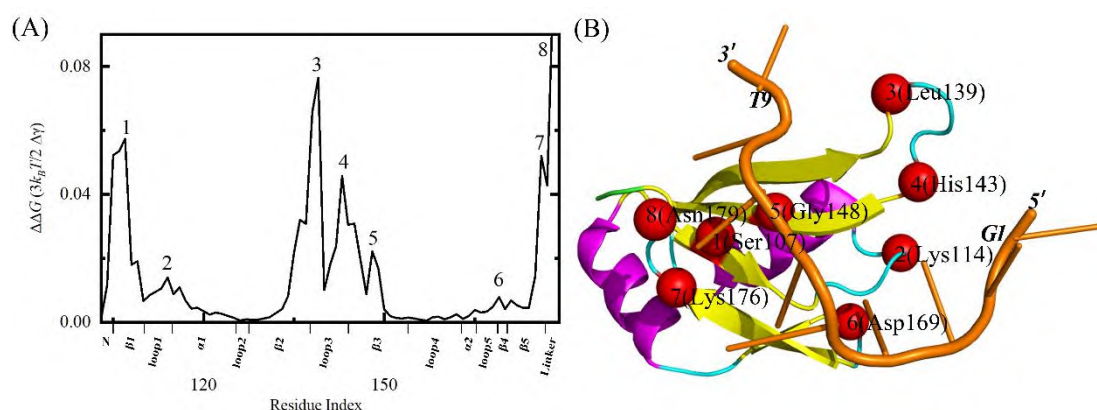


图 4. 基于 GNM 的热力学循环方法识别到的关键残基簇。(A) $\Delta\Delta G$ 是残基突变引起的结合自由能的变化, 具有相对高的 $\Delta\Delta G$ 值的关键残基被分别用数字 1-8 标注。(B) 8 个关键残基簇的中心残基在结构上的位置。

Fig 4. Key residue clusters identified by GNM-based thermodynamics cycle method. (A) $\Delta\Delta G$ values in response to residue mutations with key residue clusters with relatively high $\Delta\Delta G$ values marked by the numbers 1-8, respectively. (B) Locations of the central residues for the 8 key residue clusters.

2.3.1 位于 loop 的残基

如图 4 所示, loop 上识别到了 3 个关键残基簇, 分别是第 2、3 和 6 簇。对于 loop1 片段, 识别到了关键残基簇 2。Kuo 等人指出, DNA 的 5' 端向上伸展与 loop1 上的残基发生了特异性相互作用^[8]。簇 2 上的 Trp113 与 G4 形成堆叠和氢键相互作用, Leu111 与 G4 形成氢键作用, 这增强了 TDP-43 和 DNA 之间的结合亲和性。此外, 突变实验表明, 簇 2 的残基 Trp113 突变为丙氨酸会导致 TDP-43 与 DNA 的结合亲和性有 6 倍的降低^[11, 24], 表明 Trp113 对 DNA 的结合非常关键。Pro112 与 G4 形成非极性相互作用, Pro112His 突变在一定程度上降低了 TDP-43 结合 DNA 的能力^[1]。另外, FoldX 方法的理论预测表明, 这一突变也会降低 TDP-43 蛋白质的稳定性^[24]。此外, DNA 结合导致残基簇 2 上的 Thr115 与残基簇 6 上的 Asp169 形成了氢键, 它们以分子内部相互作用的形式稳定了复合物的结构。对于位于 loop3 片段的关键残基簇 3, 其上的 Leu139 位于 TDP-43 与 DNA 的结合缝隙中。定点突变实验表明, Leu139Ala 突变会导致结合亲和性有 3 倍的降低^[7]。关键残基簇 6 位于 loop5 片段, MD 模拟实验表明, Asp169Gly 突变导致了 TDP-43 中连接 $\beta 4$ 和 $\beta 5$ 的 β 转角发生局部构象变化, 增加了结构核心的疏水相互作用, 从而提高了 TDP-43 的热稳定性, 增强了 TDP-43 与 DNA 的结合亲和性^[13]。

综合以上, 及 DNA 结合导致 loop1 和 loop3 发生柔性明显损失的结果, 说明 loop1 和 loop3 对 TDP-43 与 DNA 间的相互作用非常重要: 它们贡献了分子间的诱导契合作用, 主要与 DNA

形成了静电和特异性氢键及堆叠作用，其关键残基的突变导致了较大的结合自由能的变化。

2.3.2 位于 β 片层的残基

如图 4 所示，在 β 片层上识别到 4 个关键残基簇，分别是第 1、4、5 和 7 簇。关键残基簇 1 中的 Ile107、Asp105 和 Leu109 分别与 C7、G8 和 G6 形成了特异性的堆叠和氢键，及非极性相互作用，将这些核苷酸准确地定位在 β 片层表面^[8]。对于 $\beta 3$ 片段，两个关键残基簇在其上，分别是簇 4 和簇 5。对于簇 4，其上的 Lys145 与 G3 形成堆叠和氢键相互作用，Gly146 与 G4 形成氢键相互作用，稳定了复合物的结构。此外，实验发现 Lys145 的乙酰化大大地抑制了蛋白与 DNA 的结合^[25]。对于簇 5，其上的 Phe147 和 Phe149，位于 RNP1 上，是 TDP-43 与 DNA 特异性相互作用中非常重要的两个残基。DNA 主要结合在 β 片层上，与 RNP1 和 RNP2 片段发生了特异性相互作用^[8]。Phe147 插入到 C7 和 G8 的两个糖环之间，Phe149 与 G8 形成堆叠作用。突变实验表明，Phe147Leu 或 Phe149Leu 突变导致了 TDP-43 与 DNA 间结合亲和性有大约 20 倍的降低，表明它们对 DNA 的结合非常关键^[8]。对于残基簇 7，Lys176 与 A5 和 G6 形成了氢键相互作用，特异性地将它们精确定位在 β 片层表面，稳定了复合物的结构。残基簇 7 中的 Arg171 与 G4 形成氢键和堆叠相互作用，特异性地识别核苷酸 G4。突变实验表明，Arg171Ala 破坏了 G4 的特异性识别，导致 TDP-43 与 DNA 之间的结合亲和性有 6 倍的降低，这表明 Arg171 残基在 DNA 的特异性识别和结合方面扮演着重要角色^[8]。

综上所述， β 片层提供了 TDP-43 与 DNA 结合的界面，其上的关键残基主要通过与 DNA 形成特异性的堆叠和氢键作用将核苷酸精确定位在 β 片层界面上，它们对特异性识别 DNA 和增强结合亲和性非常重要。

2.3.3 位于 linker 的残基

对于 linker 区域，识别到了关键残基簇 8。毫无疑问，linker 的柔性比较大，它在 DNA 结合导致 TDP-43 构象变化中起到了重要作用，同时也构成了结合界面的一部分，对 DNA 的结合有一定程度的贡献^[25]。对于残基簇 8，其上的 Asn179 与 C7 形成了氢键相互作用，实验发现将 Asn179 突变为丙氨酸会导致结合亲和性有一定程度的改变^[8]。

总之，linker 参与了与 DNA 的直接相互作用，加强了结合亲和性，在结合识别动力学中扮演着重要作用。

3 总结

TDP-43 是一种可变剪切因子，能够特异性地识别 DNA 片段，在 DNA 转录和 RNA 翻译过程中发挥重要作用。首先，我们对 TDP-43-DNA 复合物构建了最优参数下的 GNM 模型，实验与理论 B-factor 的 PCC 值达到 0.67，表明 GNM 可以较好地再现 TDP-43-DNA 复合物的柔性。接下来，通过比较分析 DNA 结合前后对蛋白质动力学的影响，发现 DNA 结合导致了 TDP-43 的 loop1、loop3 和 $\beta 1$ - $\beta 5$ 片段发生了一定程度的柔性损失，其中 loop1 和

loop3 在 TDP-43-DNA 识别的诱导契合方面发挥了重要作用, β 片层构成了主要的 DNA 结合界面, 它们对特异性的相互作用非常关键。

最后, 利用基于 GNM 的热力学循环方法识别了 TDP-43 上对 DNA 结合起关键作用的残基, 它们的突变导致了结合自由能较大的变化。根据识别到的关键残基在蛋白质结构中的位置, 将其分为三部分来分析。loop1 和 loop3 片段的柔性较高, 其上的关键残基对 TDP-43-DNA 的诱导契合和静电吸引比较重要; 位于 β 片层的关键残基提供了与 DNA 之间的特异性相互作用, 将其上的核苷酸精确定位于 β 片层构成的结合界面上; 位于 linker 片段的关键残基对于构象调整非常重要, 并参与了与 DNA 的直接相互作用, 在结合识别动力学中扮演着重要角色。该方法不仅识别到了一些与 DNA 特异性相互作用有关的重要残基, 而且识别到了一些远离结合界面但在结合引起的分子构象变化中发挥重要作用的残基。我们识别到的关键残基与实验信息吻合程度较高, 因此本研究有助于理解 TDP-43 与 DNA 的特异性相互作用, 可为药物设计提供重要信息。

虽然 GNM 是一种粗粒化的弹性网络模型, 但已被成功地用于研究生物大分子的功能性变构中。考虑到该方法的可解析性, 以及与实验结果较好的一致性, 我们相信 GNM 及所发展的基于 GNM 的热力学循环方法是有前途的方法, 可以很容易地推广到其他蛋白质-蛋白质/核酸相互作用的结合和变构动力学研究中。

参考文献

- [1] Furukawa Y, Suzuki Y, Fukuoka M, *et al.* A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Scientific reports*, 2016, **6**(20576)
- [2] Neumann M, Sampathu D M, Kwong L K, *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, 2006, **314**(5796): 130-133
- [3] Arai T, Hasegawa M, Akiyama H, *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and biophysical research communications*, 2006, **351**(3): 602-611
- [4] Gendron T F, Josephs K A, Petrucelli L. Review: transactive response DNA-binding protein 43 (TDP-43): mechanisms of neurodegeneration. *Neuropathol Appl Neurobiol*, 2010, **36**(2): 97-112
- [5] Josephs K A, Whitwell J L, Weigand S D, *et al.* TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta neuropathologica*, 2014, **127**(6): 811-824
- [6] Kuo P-H, Doudeva L G, Wang Y-T, *et al.* Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic acids research*, 2009, **37**(6): 1799-1808
- [7] Lukavsky P J, Daujotyte D, Tollervey J R, *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nature structural & molecular biology*, 2013, **20**(12): 1443-1449
- [8] Kuo P-H, Chiang C-H, Wang Y-T, *et al.* The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic acids research*, 2014, **42**(7): 4712-4722
- [9] Oubridge C, Ito N, Evans P R, *et al.* Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 1994, **372**(6505): 432-438
- [10] Wang X, Tanaka Hall T M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nature structural biology*, 2001, **8**(2): 141-145
- [11] Buratti E, Baralle F E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *The Journal of biological chemistry*, 2001, **276**(39): 36337-36343
- [12] Bhardwaj A, Myers M P, Buratti E, *et al.* Characterizing TDP-43 interaction with its RNA targets. *Nucleic acids research*, 2013, **41**(9): 5062-5074
- [13] Chiang C-H, Grauffel C, Wu L-S, *et al.* Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Scientific reports*, 2016, **6**(21581)
- [14] Prakash A, Kumar V, Meena N K, *et al.* Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *Journal of Biomolecular Structure & Dynamics*, 2019, **37**(1): 178-194
- [15] Han Z, Shao Q, Gong W, *et al.* Interpreting the dynamics of binding interactions of snRNA and U1A using a coarse-grained model. *Biophysical journal*, 2019, **116**(9): 1625-1636
- [16] Zhang P F, Su J G. Identification of key sites controlling protein functional motions by using elastic network model combined with internal coordinates. *The Journal of chemical physics*, 2019, **151**(4): 045101
- [17] Tirion M M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters*, 1996, **77**(9): 1905-1908
- [18] Bahar I, Jernigan R L. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. *Journal of molecular biology*, 1998, **281**(5): 871-884

- [19] Abdizadeh H, Guven G, Atilgan A R, *et al.* Perturbation response scanning specifies key regions in subtilisin serine protease for both function and stability. *Journal of enzyme inhibition and medicinal chemistry*, 2015, **30**(6): 867-873
- [20] Zheng W, Brooks B. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *Journal of molecular biology*, 2005, **346**(3): 745-759
- [21] Su J G, Du H J, Hao R, *et al.* Identification of functionally key residues in AMPA receptor with a thermodynamic method. *Journal of Physics Chemical B*, 2013, **117**(29): 8689-8696
- [22] Su J G, Xu X J, Li C H, *et al.* Identification of key residues for protein conformational transition using elastic network model. *The Journal of chemical physics*, 2011, **135**(17): 174101
- [23] Hudson W H, Ortlund E A. The structure, function and evolution of proteins that bind DNA and RNA. *Nature reviews Molecular cell biology*, 2014, **15**(11): 749-760
- [24] Zacco E, Graña-Montes R, Martin S R, *et al.* RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *Journal of molecular biology*, 2019, **431**(8): 1671-1688
- [25] François-Moutal L, Perez-Miller S, Scott D D, *et al.* Structural insights into TDP-43 and effects of post-translational modifications. *Frontiers in molecular neuroscience*, 2019, **12**(301)

Study on the interacting dynamics and key residues between human TDP-43 and DNA based on GNM model*

DENG Xueqing, WANG Shihao, GONG Weikang, LI Chunhua**

Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Abstract

Transactive response DNA binding protein 43 (TDP-43), an alternative-splicing regulator, can specifically bind the TG-rich DNAs, which is associated with a range of neurodegenerative diseases. Molecular dynamics simulation, although powerful in exploring inter-molecular interactions, is time-consuming, and moreover it is difficult to sample sufficiently the conformations for the system with large conformational changes to study the allosteric behavior. Here, we utilize a coarse-grained, elastic potential-based Gaussian network model (GNM) to characterize the interacting dynamics between human TDP-43 and DNA. The results reveal that upon DNA binding, an evident loss of flexibility occurs to TDP-43's loop1 and loop3 segments rich in positively charged residues, which indicates their induced fit role in TDP-43-DNA recognition and interactions. Furtherly, using our group's previously proposed thermodynamic cycle method based on GNM, we identify the key residues for DNA binding whose perturbations induce a large change in their binding free energy. We found that the approach identifies not only the residues important for DNA specific binding, but also the ones far away from the binding

interface but critical for the conformational changes of TDP-43 caused by the DNA binding. This study is helpful for the understanding of the specific interaction between TDP-43 and DNA, and can provide important information for the related drug design. In addition, this method can be easily extended to other protein-nucleic acid interacting dynamics studies.

Keywords: TDP-43-DNA interactions; GNM; key residues; interacting dynamics

*This work was supported by grants from the National Natural Science Foundation of China (31971180, 11474013).

** Corresponding author: LI Chunhua. Tel.: +86-010-67392001, E-mail: chunhuali@bjut.edu.cn

Received: April 29, 2021 Accepted: June 7, 2021