



基于弹性网络模型的蛋白质变构路径与关键残基识别研究

李 娇¹⁾ 王韦卜¹⁾ 苏计国^{1,2)*}¹⁾ 燕山大学理学院, 河北省微结构材料物理重点实验室, 秦皇岛 066004;²⁾ 国药中生生物技术研究院有限公司第六研究室, 北京 101111)

摘要 **目的** 变构效应在蛋白质生物学功能执行过程中发挥着重要的调控作用, 如何基于蛋白质空间结构, 有效识别变构信号的传播路径和关键的残基位点是蛋白质结构-功能关系研究领域的热点科学问题。**方法** 本研究利用基于弹性网络模型 (elastic network model, ENM) 的力分布计算方法, 通过分析蛋白质对外力的响应过程, 来识别体系的变构路径以及变构过程中的关键残基。在该方法中, 对蛋白质的关键变构位点施加外力, 通过对体系形变以及内力分布情况的分析, 有效识别与外力承载区域形变相耦合的关键残基, 从而得到力信号在蛋白质结构内的传播路径。**结果** 利用该方法研究了人类磷酸甘油酸激酶 (human phosphoglycerate kinase, hPGK) 和蛋白质酪氨酸磷酸酶 (protein tyrosine phosphatase, PTP) PDZ2 结构域的变构调控路径和关键残基。对于 hPGK, 识别出从底物结合位点到铰链区的两条变构信号传导路径。对于 PTP PDZ2, 也成功识别出从配体结合位点传递到蛋白质远端的两条长程变构调控路径。计算结果与实验和分子动力学 (molecular dynamics, MD) 模拟得到的结果一致。**结论** 本研究为蛋白质体系关键残基识别及变构路径研究提供了有效的分析方法。

关键词 人类磷酸甘油酸激酶, 酪氨酸磷酸酶 PDZ2 结构域, 弹性网络模型, 力分布, 关键残基, 变构路径

中图分类号 Q7

DOI: 10.16476/j.pibb.2021.0227

变构调节是蛋白质功能调节的一种重要机制, 变构效应因子在变构位点的结合, 可以通过特定的变构信号传输路径传递到远处的活性位点, 使活性部位的局部结构或构象运动发生改变, 进而对蛋白质的功能进行调节^[1-2]。变构相互作用最早由 Umbarger 等^[3] 在 1956 年提出, 他们发现了左旋异亮氨酸对左旋苏氨酸脱氨酶的反馈抑制作用。之后, Changeux 等^[4-6] 在含有几个亚单位的蛋白质体系中, 观察到配体结合引起蛋白质结构发生一定的构象变化, 由此提出了变构效应。尽管蛋白质的变构调节在结构生物学领域的研究进行了大半个世纪, 但是蛋白质变构调节的物理机制以及相应变构路径的有效识别, 仍是一个有待深入研究的重要科学问题。变构调节在蛋白质功能反馈和调控机制中起着关键性作用, 许多蛋白质的活性及其功能的调节都是通过变构效应来实现的。因此, 蛋白质变构

路径的有效识别, 不仅有助于揭示蛋白质功能调控的分子机制, 还可以为变构药物的开发或纳米生物传感器件^[7-9] 的开发提供理论依据。

分子动力学模拟 (molecular dynamics, MD) 已广泛应用于蛋白质的变构效应研究中, 通过对构象运动和运动关联性的分析等来揭示蛋白质体系的变构路径和关键残基^[10-15]。Roy 等^[16] 运用 MD 模拟结合运动相关性分析, 对人类鼻病毒 (HRV) 的变构调控过程进行了研究, 他们通过对不同残基之间径向运动关联性的分析, 揭示了抗病毒化合物 WIN 52084 对 HRV 衣壳蛋白“呼吸”运动的长程变构调控机制。Bowerman 和 Wereszczynski^[17] 基于 MD 模拟轨迹, 利用残基运动相关性分析、互信

* 通讯联系人。

Tel: 13260014653, E-mail: jiguosu@ysu.edu.cn

收稿日期: 2021-08-06, 接受日期: 2021-10-10

息关联分析以及图论的方法来识别蛋白质体系的变构网络,利用该方法他们详细分析了凝血酶变构位点与催化中心之间的长程变构信息传播路径。MD模拟虽然在蛋白质变构效应研究中得到了成功应用,但是,很多情况下,蛋白质的变构调控并不一定伴随明显的构象运动,而是通过刚性区域来耗散变构所产生的应力,这些刚性区域不会发生大的构象变化,而是产生较大的力信号^[18],常规的平衡态MD模拟很难识别到这些力信号。因此,一些研究者通过对体系施加外界的扰动,利用MD模拟结合力分布分析(force distribution analysis, FDA)方法来揭示蛋白质内力信号的长程传播过程,进而获得体系的变构调控路径^[19-23]。Stacklies等^[24]对肌联蛋白I27结构域的长程力学变构调控进行了研究,对I27的N端、C端施加300 pN的恒力进行拉伸,运用MD结合FDA的方法,计算了不同残基间的力分布,得到了蛋白质力信号的传导路径。随后,该小组运用同样的方法,对MetJ蛋白进行了研究,识别了力信号从变构位点传递到活性位点的变构路径^[18]。2013年,Palmai等^[25]利用MD结合FDA的方法,研究了3-磷酸还原酶(hPGK)催化功能执行过程中的变构调控机制,揭示了变构信号从底物结合位点到功能性运动铰链区的传导路径。这些研究表明,基于MD模拟轨迹,通过对蛋白质体系内部力分布的分析来研究蛋白质变构调控路径是一种有效的分析方法。但是,利用MD模拟来计算蛋白质内的力分布存在一定的局限性。首先,由于蛋白质体系的热涨落,力学信号常常被热噪声所淹没,需要设计一定的去噪音手段才能获取有价值的力学信号。其次,为了去除热噪声,获得具有统计意义的力分布结果,需要进行大量轨迹的模拟,计算非常耗时。而弹性网络模型(elastic network model, ENM)不需要考虑热噪声的影响,并且对于体系的配分函数和热力学性质可以进行解析求解,大大的减少计算量,提高计算效率。

ENM方法是研究蛋白质结构和动力学性质的有效方法^[26-27],在蛋白质结构-功能关系研究中得到了广泛应用,包括蛋白质结构稳定性的分析、关键残基的识别、蛋白质结构域的分解、大幅度功能性构象运动的分析等^[28-35]。Wang等^[36]基于ENM,通过力分布的计算来分析蛋白质体系对外力的响应过程,研究了GB1蛋白质力学性能的变构调控机

制。对GB1蛋白的N端、C端施加20 pN的外力进行拉伸,分别研究结合配体与未结合配体情况下蛋白质体系的形变情况和力分布情况。研究发现,配体的结合能够通过长程变构效应调控抗力部位的机械性能,显著增强GB1的力学稳定性。

本工作将基于ENM的力分布计算方法用于蛋白质变构路径和关键残基位点的识别研究,基于该方法,分析了人类磷酸甘油酸激酶(human phosphoglycerate kinase, hPGK)和蛋白质酪氨酸磷酸酶(protein tyrosine phosphatase, PTP) PDZ2结构域的变构路径和关键残基。对两种蛋白质的变构位点施加外力进行拉伸,通过计算外力作用下蛋白质内部的力分布,来识别与力承载区域形变相耦合的残基位点,并研究力信号在蛋白质结构内的传播路径。

1 理论与方法

ENM将蛋白质的空间结构简化为一个由大量节点和节点之间的弹簧所构成的弹性网络。将蛋白质体系中的每一个残基简化为网络中的一个节点,用C α 原子来代替,其他的原子忽略。选取一个几何距离作为截断半径 r_c ^[26]。如果两个残基间的距离小于该截断半径,认为它们之间存在相互作用,则用弹簧连接;若两残基间距离大于该截断半径,则认为不存在相互作用。所有连接的弹簧的弹性系数均相同。在本工作中,弹簧的弹性系数设为1.0 N/m,选取的截断半径为8.0 Å。由整个体系中弹簧的连接情况,可得到体系的势能为

$$V = \frac{1}{2} \gamma [\Delta R^T H \Delta R] \quad (1)$$

其中, γ 为弹簧的弹性系数; ΔR 为蛋白质内各个氨基酸偏离其平衡位置的位移,是 $3N$ 维列向量, N 为残基的个数;右上角的T表示向量的转置; H 为 $3N \times 3N$ 的Hessian矩阵,该矩阵可表示为

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{NN} \end{bmatrix} \quad (2)$$

上式中的 h_{ij} 为 3×3 的子矩阵, i, j 分别表示第 i, j 个残基。

当 $i \neq j$ 时,有

$$h_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_i \partial x_j} & \frac{\partial^2 V}{\partial x_i \partial y_j} & \frac{\partial^2 V}{\partial x_i \partial z_j} \\ \frac{\partial^2 V}{\partial y_i \partial x_j} & \frac{\partial^2 V}{\partial y_i \partial y_j} & \frac{\partial^2 V}{\partial y_i \partial z_j} \\ \frac{\partial^2 V}{\partial z_i \partial x_j} & \frac{\partial^2 V}{\partial z_i \partial y_j} & \frac{\partial^2 V}{\partial z_i \partial z_j} \end{bmatrix} \quad (3)$$

当 $i = j$ 时, 有

$$h_{ii} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_i^2} & \frac{\partial^2 V}{\partial x_i \partial y_i} & \frac{\partial^2 V}{\partial x_i \partial z_i} \\ \frac{\partial^2 V}{\partial y_i \partial x_i} & \frac{\partial^2 V}{\partial y_i^2} & \frac{\partial^2 V}{\partial y_i \partial z_i} \\ \frac{\partial^2 V}{\partial z_i \partial x_i} & \frac{\partial^2 V}{\partial z_i \partial y_i} & \frac{\partial^2 V}{\partial z_i^2} \end{bmatrix} \quad (4)$$

其中, 矩阵元为势能函数 V 对坐标 x 、 y 求二阶偏导, 表达式为

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{-\gamma(x_j - x_i)(y_j - y_i)}{r_{ij}^2} \bigg|_{r_{ij} = r_{ij}^0} \quad (5)$$

$$\frac{\partial^2 V}{\partial x_i \partial y_i} = \gamma \sum_{j \neq i} \frac{(x_j - x_i)(y_j - y_i)}{r_{ij}^2} \bigg|_{r_{ij} = r_{ij}^0} \quad (6)$$

在本工作中, 由于对蛋白质关键位点施加外力进行拉伸, 体系的总势能 V_{total} 由原来的弹性势能 V 变为弹性势能与外力做功之和^[33, 37], 即:

$$V_{\text{total}} = \frac{1}{2} \Delta R^T H \Delta R - F^T \Delta R \quad (7)$$

其中, F 表示施加在蛋白质上的外力, 为 $3N$ 维列向量。对 i 、 j 残基施加外力进行拉伸, 则外力 F 矩阵的构建为: $F = (000 \cdots F_x^i F_y^i F_z^i \cdots F_x^j F_y^j F_z^j \cdots 000)^T_{1 \times 3N}$ 。将体系的总势能在平衡位置求梯度, 可得到体系达到平衡后, 各个残基偏离平衡位置的位移为:

$$\Delta R = H^{-1} F \quad (8)$$

上式中, H^{-1} 矩阵的矩阵元为 $[H^{-1}]_{ij} = \sum_{k=7}^{3N} \frac{[U_k]_i [U_k]_j}{\lambda_{kk}}$, $[H^{-1}]_{ij}$ 表示 H^{-1} 矩阵的 i 行 j 列的矩阵元, λ_{kk} 为 Hessian 矩阵的第 k 个本征值, U 为 Hessian 矩阵的本征向量, $[U_k]_i$ 表示第 k 个本征向量中的第 i 个元素。残基位移所导致的体系内弹簧形变可以表示为^[38-39]:

$$\Delta r_{M \times 1} = A_{M \times 3N}^T \Delta R_{3N \times 1} \quad (9)$$

$\Delta r_{M \times 1}$ 为残基间弹簧的形变, M 表示组成弹性网络的弹簧总数, A^T 是残基偏移平衡位置位移在弹簧形变矢量上的投影。残基间相应的内力分布可以表示为:

$$f_{M \times 1} = K_{M \times M} \Delta r_{M \times 1} \quad (10)$$

K 是弹簧弹性系数所构成的 M 维的对角矩阵, f 是体系中所有弹簧承受的力, 为 $M \times 1$ 维的向量。

2 结果与讨论

2.1 hPGK 结构域中的关键残基及变构路径

hPGK 在生物体内参与糖酵解过程, 是催化 1,3 二磷酸甘油酸转化为 3-磷酸甘油酸的关键酶, 可产生 1 分子的 ATP, 在糖酵解的过程中发挥着不可或缺的作用^[25, 40]。hPGK 的三维结构 (PDB 代码 2XE7) 由两个大小相近的结构域组成 (图 1a)。研究证实, hPGK 两个结构域的开合运动在催化功能的发挥中起着重要作用, 而 Arg38、Glu192、Asn336、Gly394 残基是两个结构域打开与闭合过程中的关键残基^[25]。其中, 残基 Arg38 和 Asn336 直接参与了底物的结合。残基 Glu192 和 Gly394 位于体系两个结构域之间的铰链区, 直接调控着两个结构域的相对运动。大量的实验结果表明, 底物的结合会引起铰链区的构象调整, 从而导致 hPGK 的两个结构域发生大幅度的构象变化, 从开放构象转变到闭合构象, 进而催化底物的化学反应, 完成其生物学功能^[41-44]。底物的结合如何调控体系铰链区的构象运动, 变构信号如何从底物结合位点有效传递到体系的铰链区是值得研究的重要问题, 变构路径的识别研究有助于更好地揭示 hPGK 发挥功能的分子机制。

在本工作中, 首先基于 hPGK 的三维结构构建弹性网络, 然后对残基 Arg38 和 Asn336 施加 1 pN 的外力进行拉伸, 拉伸方向为两残基位点连线的方向。进而, 通过公式 (9) 和 (10) 计算蛋白质在外力作用下的形变情况和相应的内力分布情况。计算得到体系内力的分布情况如图 1b 所示。图中蓝色到红色表示力的分布从小到大, 受力越大表明相应残基对于力信号的传导越重要, 相应位点为力信号变构传导的关键位点。

将图 1b 中不同残基之间弹簧的受力情况映射到蛋白质的三级结构上, 并根据不同残基间力的分布情况来识别力信号的传播路径 (图 2)。将外力所施加的两个残基位点 Arg38 和 Asn336 在图中用紫色球标出, 为了更清晰的显示力信号的传播路径, 将受力大于 0.05 pN 的弹簧显示在图 2 中。图中残基间连线的粗细表示其承受的内力大小, 连线越粗, 表示所承受的力越强, 对力信号传导也越重要; 反之, 连线越细表示受力越弱。由残基间内力

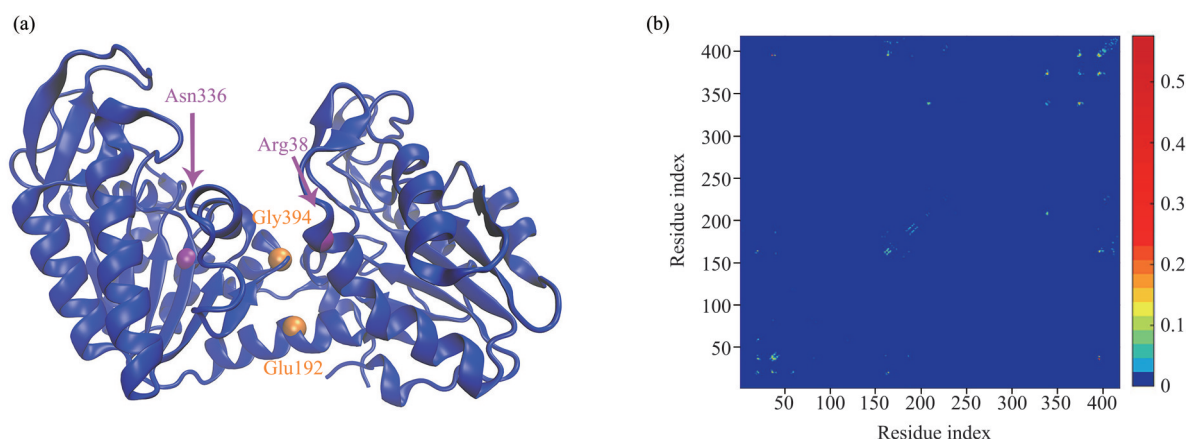


Fig. 1 Three-dimensional structure of hPGK (a) and internal force distribution of hPGK in response to the exertion of external forces (b)

In Figure 1a, the purple spheres are the force-loading sites, and the orange spheres are the hinge points between the two domains of hPGK.

分布情况, 将底物结合位点与铰链区之间受力最大的弹簧相连 (图2红色连线), 得到力信号的传导路径。根据该方法, 从Arg38出发, 可以得到两条到达铰链区的变构路径: Arg38→Thr393→Gly394 (铰链区) 和 Arg38→Thr393→Ala41→Leu188→Glu192 (铰链区)。这两条变构路径用红色箭头显示在了图2中, 其中所涉及到的关键残基用黄色球

标出。采用类似方法, 从Asn336施力位点出发, 也可以得到两条传播到铰链区的力信号传导路径: Asn336→Gly371→Thr393→Gly394 (铰链区) 和 Asn336→Gly371→Thr393→Ser392→Glu192 (铰链), 传导路径在图2中用蓝色箭头标出。变构路径中残基间的内力值见表1。

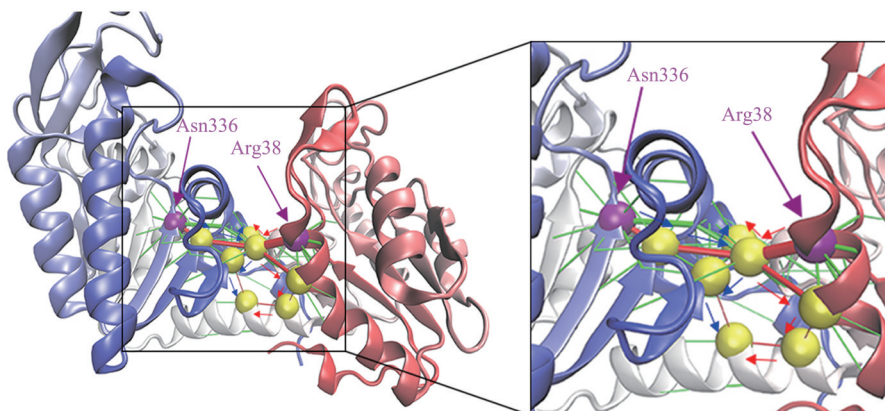


Fig. 2 Allosteric pathways in hPGK identified by our method

The force-exerting sites are displayed by purple balls, and the springs with force larger than 0.05 pN are shown in green lines. The identified allosteric pathways are represented by red or blue arrows. The inter-residue springs in the allosteric pathways are highlighted in red lines and the associated residues are displayed in yellow balls.

Table 1 Internal forces between residues in the allosteric pathway of hPGK in response to the exertion of external forces

Allosteric pathway		Internal force values/pN			
Arg38→Thr393→Gly394	0.576	0.123			
Arg38→Thr393→Ala41→Leu188→Glu192	0.576	0.251	0.050	0.070	
Asn336→Gly371→Thr393→Gly394	0.266	0.330	0.123		
Asn336→Gly371→Thr393→Ser392→Glu192	0.266	0.330	0.241	0.050	

Szabo 等^[45]通过残基突变、酶反应动力学实验以及结构分析,研究了底物结合对hPGK功能性开合构象运动的变构调控过程,并识别了变构信号从底物结合位点到构象运动铰链区传播过程中的关键残基,发现Arg38、Lsy219、Asn336、Glu343等残基在结构域开合过程中发挥着重要作用。另外,Palmai 等^[25]利用MD模拟与FDA相结合的方法,研究了底物结合对铰链区构象转变的变构调控作用,得到了两条长程变构路径,分别是结合位点残基Arg38传递到Thr393再到铰链区残基Glu192、Gly394,以及从结合位点残基Asn336到Gly371、Ser398、Ser392再传递到铰链位区Glu192和Gly394。本文的计算结果与上述实验和MD模拟结果能够很好地吻合。

2.2 PTP PDZ2结构域的关键残基及变构路径

PTP是生物体内的关键酶^[46-48]。它与蛋白酪氨酸激酶协同作用,共同维持着酪氨酸磷酸化的平衡,进而参与体内的细胞生长、新陈代谢等活动。PDZ结构域存在于许多蛋白质内,被认为具有蛋白质定位、信号传输等功能^[46, 49-50]。PTP包含有5个PDZ结构域,其中,第二个结构域由6条 β 链($\beta_1 \sim \beta_6$)和2条 α 螺旋(α_1 、 α_2)组成,含有96个氨基酸,属于典型的PDZ结构^[51-52]。图3a中不同的颜色表示不同的局部二级结构。 β_2 与 α_2 螺旋之间存在一个配体结合口袋,实验研究表明,配体的结合能够长程变构调控远端的 α_1 和 β_1 区域与其他蛋白质的相互作用,变构信号如何从配体结合口袋传递到远端区域是有待深入研究的重要问题,有助于更好揭示PDZ2功能的分子机制。

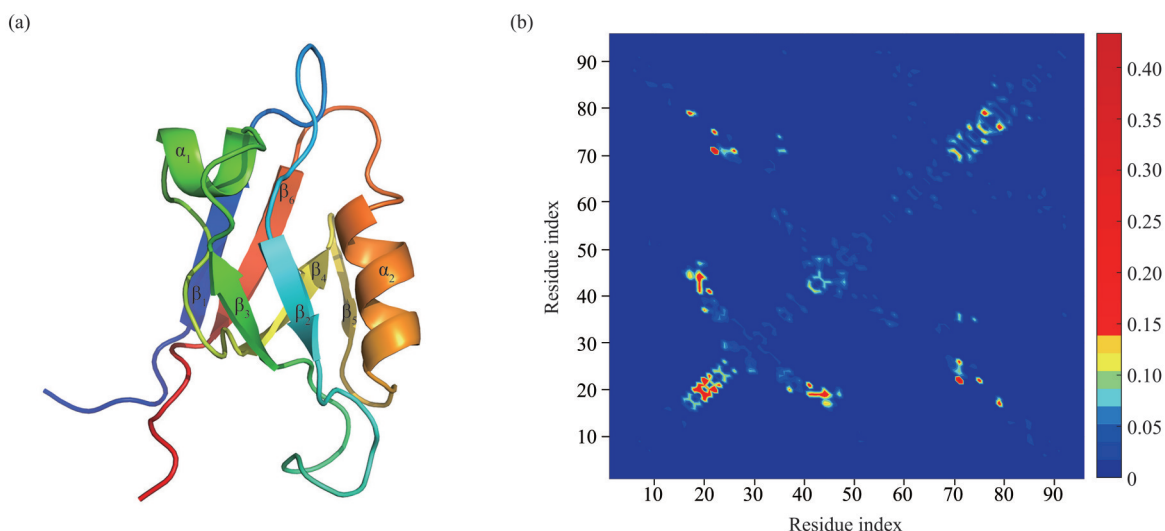


Fig. 3 The three-dimensional structure of PTP PDZ2 (a) and internal force distribution of PTP PDZ2 in response to the exertion of external forces (b)

本研究基于PTP PDZ2的三级结构(PDB代码3PDZ)构建相应的弹性网络,然后对配体结合口袋区域的变构位点残基Gly19和His71施加1 pN外力进行拉伸,拉伸方向为两残基的连线方向。运用公式(10)计算蛋白质体系在外力作用下的内力分布情况(图3b)。图3b中的红色区域对应于承受内力较大的残基,是力信号传导过程中的关键残基。

为了更清晰地分析力信号的传导路径,将计算得到的力分布显示到蛋白质的三级结构上(图4)。图中仅画出了受力大于0.01 pN的弹簧,显示为绿色连线。连线的粗细表示相应弹簧承受的内力情

况,连线越粗表示弹簧所承受的内力越强。由残基间内力分布情况,将施力位点与远端 β_1 区域间受力最大的弹簧相连,得到两条特异性变构路径(图4红色连线)。一条变构路径从 β_2 的施力位点开始,传递到 α_1 螺旋的N端,最终到达 β_1 - β_2 之间的loop区,即Gly19、Ser21(β_2) $\rightarrow$ Ile41、Gln43、Gly45(α_1) $\rightarrow$ Ser17(β_1 - β_2 之间的loop区),如图4右侧上图中红色箭头所示,变构路径中的关键残基用黄色球显示在图中。另一条力信号的传导路径是 α_2 螺旋的施力位点开始,穿过 β_2 、 β_3 、 β_4 、 β_6 ,最终到达 β_1 ,即His71(α_2) $\rightarrow$ Val22、Ile20(β_2) $\rightarrow$ Val37

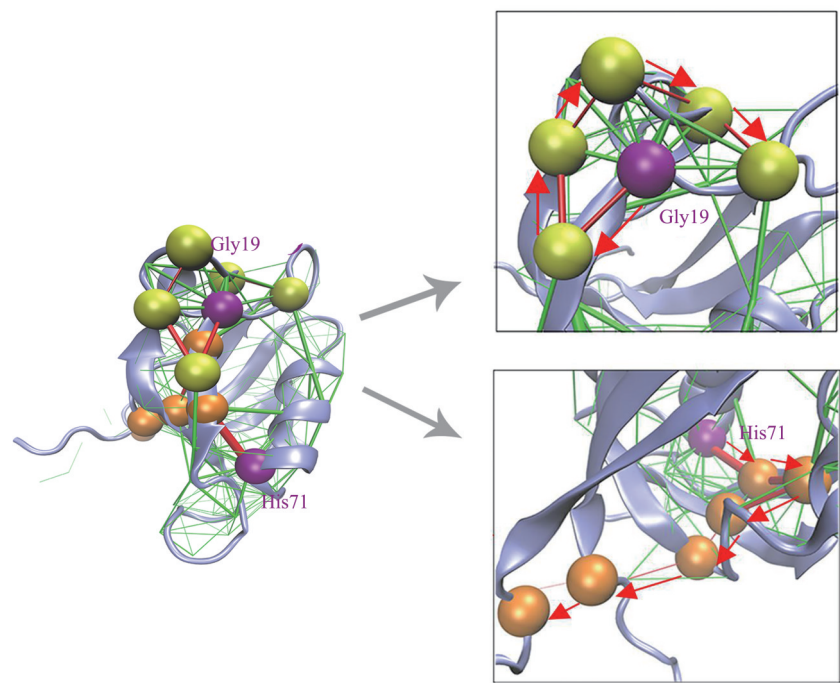


Fig. 4 Allosteric pathways in PTP PDZ2 identified by our method

The force-exerting sites are displayed by purple balls, and the springs with force larger than 0.01 pN are shown in green lines. The identified allosteric pathways are represented by red arrows. The inter-residue springs in the allosteric pathways are highlighted in red lines and the associated residues are displayed in yellow or orange balls.

(β_3) $\rightarrow$ Gly55 (β_4) $\rightarrow$ Gln93、Lys91 (β_6) $\rightarrow$ Gly4 (β_1), 如图4右侧下图中红色箭头所示, 变构路径中的关键残基用橙色球显示在图中。变构路径中残基间的内力值见表2。

Table 2 Internal forces between residues in the allosteric pathway of PTP PDZ2 in response to the exertion of external forces

Allosteric pathway	Internal force values/pN							
Gly19, Ser21 $\rightarrow$ Ile41, Gln43, Gly45 $\rightarrow$ Ser17	0.179	0.220	0.099	0.120	0.139			
His71 $\rightarrow$ Val22, Ile20 $\rightarrow$ Val37 $\rightarrow$ Gly55 $\rightarrow$ Gln93, Lys91 $\rightarrow$ Gly4	0.396	0.314	0.180	0.041	0.037	0.016	0.010	

Kong等^[48]基于MD模拟, 采用残基间相互作用的关联性分析, 研究了PTP PDZ2体系的长程变构调控机制, 识别出两条变构路径, 一条路径始于 β_2 上的变构位点长程传递到 α_1 螺旋, 另一条路径始于 α_2 上的配体结合位点, 垂直穿过 β_2 、 β_3 、 β_4 、 β_6 最终到达 β_1 链的N端。本研究利用基于ENM的力分布分析方法得到的变构路径与Kong等利用MD模拟得到的结果能够很好地吻合。相对于MD模拟, 基于ENM的力分布分析方法计算更为简单、快速。

3 结 论

本文利用基于ENM的力分布计算方法, 通过分析蛋白质对外力的响应, 研究了hPGK、PTP PDZ2两个蛋白质体系力信号的变构调控机制, 成功识别出体系中特异的变构路径和关键残基, 与实验和其他计算模拟结果能够很好地吻合。该方法将蛋白质的三级结构简化为弹性网络, 通过对变构位点施加外力, 计算蛋白质体系在外力作用下的形变情况和体系内部的力分布情况, 来识别力信号传播过程中起重要作用的关键残基和相应的变构传播路

径。相对于MD模拟方法, 基于ENM的力分布分析方法具有明显的优势, MD模拟需要耗费大量的模拟时间, 并且需要利用有效的统计分析方法, 去除噪音的干扰, 来识别变构信号, 而ENM方法可以通过解析求解来获得体系的力分布情况, 不需要进行数值模拟和轨迹采样, 也不包含噪声, 计算更加简洁快速, 能够用于超大蛋白质体系的研究。因此, 本文为蛋白质的变构效应研究提供了一种简单、有效的方法。但是, 需要指出的是, 该方法也存在一定的不足, 由于ENM是一种对蛋白质结构进行高度简化的粗粒化模型, 每个氨基酸简化为一个点, 很难考虑不同配体结合、残基的不同突变、翻译后修饰等相互作用的精细差别, 要更为精确的研究这些不同类型相互作用的影响, 还需要结合更精细的全原子模型。

参 考 文 献

- [1] Li G, Magana D, Dyer R B. Anisotropic energy flow and allosteric ligand binding in albumin. *Nat Commun*, 2014, **5**: 3100
- [2] Liang Z, Verkhivker G M, Hu G. Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *BriefBioinform*, 2020, **21**(3): 815-835
- [3] Umbarger H E, Brown B. Threonine deamination in *Escherichia coli*. I: d- and l-threonine deaminase activities of cell-free extracts. *J Bacteriol*, 1956, **71**(4): 443-449
- [4] Monod J, Changeux J P, Jacob F. Allosteric proteins and cellular control systems. *J Mol Biol*, 1963, **6**(4): 306-329
- [5] Monod J, Wyman J, Changeux J P. On the nature of allosteric transitions: a plausible model. *J Mol Biol*, 1965, **12**(1): 88-118
- [6] Changeux J P. 50th anniversary of the word "allosteric". *Protein Sci*, 2011, **20**(7): 1119-1124
- [7] Cao Y, Li H. Polypeptide of GB1 is an ideal artificial elastomeric protein. *Nat Mater*, 2007, **6**(2): 109-114
- [8] Cao Y, Yoo T, Li H. Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. *Proc Natl Acad Sci USA*, 2008, **105**(32): 11152-11157
- [9] Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci USA*, 2004, **101**(46): 16192-16197
- [10] De Groot B L, Van Aalten D M F, Scheek R M, *et al.* Prediction of protein conformational freedom from distance constraints. *Proteins*, 1997, **29**(2): 240-251
- [11] Amir M, Mohammad T, Kumar V, *et al.* Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. *Front Mol Biosci*, 2019, **6**: 41
- [12] Lange O F, Grubmüller H. Full correlation analysis of conformational protein dynamics. *Proteins*, 2008, **70**(4): 1294-1312
- [13] Harada R, Kitao A. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J Chem Phys*, 2013, **139**(3): 035103
- [14] Li Y, Deng L, Yang L Q, *et al.* Effects of CD4 binding on conformational dynamics, molecular motions, and thermodynamics of HIV-1 gp120. *Int J Mol Sci*, 2019, **20**(2): 260
- [15] Fenwick R B, Orellana L, Esteban-Martín S, *et al.* Correlated motions are a fundamental property of β -sheets. *Nat Commun*, 2014, **5**: 4070
- [16] Roy A, Post C B. Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. *Proc Natl Acad Sci USA*, 2012, **109**(14): 5271-5276
- [17] Bowerman S, Wereszczynski J. Detecting allosteric networks using molecular dynamics simulation. *Methods Enzymol*, 2016, **578**: 429-447
- [18] Stacklies W, Xia F, Gräter F. Dynamic allostery in the methionine repressor revealed by force distribution analysis. *PLoS Comput Biol*, 2009, **5**(11): e1000574
- [19] Seifert C, Gräter F. Force distribution reveals signal transduction in *E. coli* Hsp90. *Biophys J*, 2012, **103**(10): 2195-2202
- [20] Seifert C, Gräter F. Protein mechanics: how force regulates molecular function. *Biochim Biophys Acta*, 2013, **1830**(10): 4762-4768
- [21] Stacklies W, Seifert C, Gräter F. Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinformatics*, 2011, **12**: 101
- [22] Xiao S, Stacklies W, Debes C, *et al.* Force distribution determines optimal length of β -sheet crystals for mechanical robustness. *Soft Matter*, 2011, **7**(4): 1308-1311
- [23] Feher V A, Durrant J D, Van Wart A T, *et al.* Computational approaches to mapping allosteric pathways. *Curr Opin Struct Biol*, 2014, **25**: 98-103
- [24] Stacklies W, Vega M C, Wilmanns M, *et al.* Mechanical network in titin immunoglobulin from force distribution analysis. *PLoS Comput Biol*, 2009, **5**(3): e1000306
- [25] Palmaí Z, Seifert C, Gräter F, *et al.* An allosteric signaling pathway of human 3-phosphoglycerate kinase from force distribution analysis. *PLoS Comput Biol*, 2014, **10**(1): e1003444
- [26] Atilgan A R, Durell S R, Jernigan R L, *et al.* Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*, 2001, **80**(1): 505-515
- [27] Bastolla U. Computing protein dynamics from protein structure with elastic network models. *Wiley Interdiscip Rev Computat Mol Sci*, 2014, **4**(5): 488-503
- [28] Li H, Chang Y Y, Yang L W, *et al.* IGNM 2.0: the gaussian network model database for biomolecular structural dynamics. *Nucleic Acids Res*, 2016, **44**(D1): D415-D422
- [29] Yang L W, Liu X, Jursa C J, *et al.* iGNM: a database of protein functional motions based on gaussian network model. *Bioinformatics*, 2005, **21**(13): 2978-2987
- [30] Su J G, Jiao X, Sun T G, *et al.* Analysis of domain movements in

- glutamine-binding protein with simple models. *Biophys J*, 2007, **92**(4): 1326-1335
- [31] Su J G, Jin Xu X, Hua Li C, *et al.* Identification of key residues for protein conformational transition using elastic network model. *J Chem Phys*, 2011, **135**(17): 174101
- [32] Su J G, Han X M, Zhang X, *et al.* Analysis of conformational motions and related key residue interactions responsible for a specific function of proteins with elastic network model. *J Biomol Struct Dyn*, 2016, **34**(3): 560-571
- [33] Su J G, Qi L S, Li C H, *et al.* Prediction of allosteric sites on protein surfaces with an elastic-network-model-based thermodynamic method. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2014, **90**(2): 022719
- [34] Polles G, Indelicato G, Potestio R, *et al.* Mechanical and assembly units of viral capsids identified *via* quasi-rigid domain decomposition. *PLoS Comput Biol*, 2013, **9**(11): e1003331
- [35] Rader A J, Chennubhotla C, Yang L W, *et al.* The gaussian network model: theory and applications//Cui Q, Bahar I. *Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems*. New York: Taylor & Francis Group, 2006: 41-64
- [36] Wang W B, Zhu J Z, Li X Y, *et al.* Enhancement of protein mechanical stability: correlated deformations are handcuffed by ligand binding. *J Chem Phys*, 2019, **150**(15): 155102
- [37] Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc Natl Acad Sci USA*, 2003, **100**(23): 13253-13258
- [38] Atilgan C, Gerek Z N, Ozkan S B, *et al.* Manipulation of conformational change in proteins by single-residue perturbations. *Biophys J*, 2010, **99**(3): 933-943
- [39] Atilgan C, Atilgan A R. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput Biol*, 2009, **5**(10): e1000544
- [40] Vas M, Varga A, Gráczér E. Insight into the mechanism of domain movements and their role in enzyme function: example of 3-phosphoglycerate kinase. *Curr Protein Peptide Sci*, 2010, **11**(2): 118-147
- [41] Banks R D, Blake C C F, Evans P R, *et al.* Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature*, 1979, **279**(5716): 773-777
- [42] Watson H C, Walker N P, Shaw P J, *et al.* Sequence and structure of yeast phosphoglycerate kinase. *EMBO J*, 1982, **1**(12): 1635-1640
- [43] Bernstein B E, Michels P A M, Hol W G J. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature*, 1997, **385**(6613): 275-278
- [44] Cliff M J, Bowler M W, Varga A, *et al.* Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *J Am Chem Soc*, 2010, **132**(18): 6507-6516
- [45] Szabó J, Varga A, Flachner B, *et al.* Communication between the nucleotide site and the main molecular hinge of 3-phosphoglycerate kinase. *Biochemistry*, 2008, **47**(26): 6735-6744
- [46] 李华,林葵,吴更. PDZ 结构域的结构特点及其识别特异性配体的机制. *中国生物化学与分子生物学报*, 2011, **27**(12): 1107-1112
- Li H, Lin K, Wu G. *Chinese Journal of Biochemistry and Molecular Biology*, 2011, **27**(12): 1107-1112
- [47] Walma T, Spronk C A E M, Tessari M, *et al.* Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. *J Mol Biol*, 2002, **316**(5): 1101-1110
- [48] Kong Y, Karplus M. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins*, 2009, **74**(1): 145-154
- [49] Van Ham M, Hendriks W. PDZ domains-glue and guide. *Mol Biol Rep*, 2003, **30**(2): 69-82
- [50] Ye F, Zhang M. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem J*, 2013, **455**(1): 1-14
- [51] Gianni S, Calosci N, Aelen J M A, *et al.* Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Engin Des Sel*, 2005, **18**(8): 389-395
- [52] Van den Berk L C J, Landi E, Walma T, *et al.* An allosteric intramolecular PDZ – PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry*, 2007, **46**(47): 13629-13637

Identification of Protein Allosteric Pathway and Related Key Residues by Using Elastic Network Model

LI Jiao¹⁾, WANG Wei-Bu¹⁾, SU Ji-Guo^{1,2)*}

¹⁾Key Laboratory for Microstructural Material Physics of Hebei Province, College of Science, Yanshan University, Qinhuangdao 066004, China;

²⁾The Sixth Laboratory, National Vaccine and Serum Institute, Beijing 101111, China)

Abstract Objective Allostery plays important roles in regulating protein biological functions. How to effectively identify the allosteric signal transduction pathway and the related key residues from the tertiary structure of proteins is a hot scientific problem in the research field of protein structure-function relationship. **Methods** In the present work, a method based on elastic network model (ENM) combined with force distribution calculation was used to investigate the response of proteins to the external loading forces, and then the allosteric pathway and associated key residues in proteins were identified based on the analysis of internal force distribution. In this method, external forces were exerted on the allosteric site of the protein, and then the deformations and internal force distributions within the protein in response to the external loading forces were analyzed. Based on these analyses, the key sites that are coupled with the deformation of the force-loading region were identified, and the transduction pathway of the force signals in the protein was obtained. **Results** By using the proposed method, two proteins, the human phosphoglycerate kinase (hPGK) and protein tyrosine phosphatase (PTP) PDZ2 domain, were investigated to identify the allosteric pathway and the related key residues in the systems. For hPGK, two allosteric pathways were identified, which mediate the transduction of force signals from the substrate binding site to the hinge region of the protein. For PTP PDZ2, two long-range allosteric pathways, from the ligand binding site to the distantly opposite side of the protein, were also successfully revealed. The calculation results are consistent with the experimental observations and the results obtained with molecular dynamics simulations. **Conclusion** This study provides an effective method for the identification of key residues and allosteric pathway in proteins.

Key words human phosphoglycerate kinase, protein tyrosine phosphatase PDZ2, elastic network model, force distribution, key residues, allosteric pathway

DOI: 10.16476/j.pibb.2021.0227

* Corresponding author.

Tel: 86-13260014653, E-mail: jiguosu@ysu.edu.cn

Received: August 6, 2021 Accepted: October 10, 2021