胶原蛋白异源三聚体的理性设计

吕成, 强书敏, 孙添添, 许菲*

江南大学生物工程学院, 无锡 214122

* 联系人, E-mail: feixu@jiangnan.edu.cn

收稿日期: 2019-02-07; 接受日期: 2019-03-19; 网络版发表日期: 2019-05-16

摘要 胶原蛋白是高等动物中最丰富的蛋白质, 约占总蛋白质质量的1/3. 同时, 胶原又是重要的生物医用材料, 被广泛应用于伤口缝线、软骨缺损填充、骨骼金属植入物镀膜等. 目前, 在已知的人类28种胶原蛋白中, 其中有1/4的胶原构型(7/28)是以异源三聚体的形式存在. 由于缺乏对胶原蛋白链间特异性识别与折叠机制的了解, 如何通过理性设计来制备清洁来源、纳米尺度可调控的胶原蛋白材料一直是蛋白质工程与生物材料工程交叉领域中的热点和难点问题. 迄今为止, 在胶原三螺旋中鉴定的最重要的成对相互作用是轴向电荷对, 其可以在适当放置的赖氨酸与天冬氨酸或谷氨酸残基之间形成. 这些成对氨基酸相互作用将允许制备具有高特异性的异源三聚体螺旋, 以及对螺旋结构和稳定性的总体改进控制. 此外, 对这些相互作用的详细研究将帮助我们修改天然胶原蛋白序列, 制造清洁来源、高度可控和成分响应的胶原蛋白生物材料. 本综述将总结我们对这种相互作用和其他电荷对相互作用的理解, 以及它们如何成功地用于控制胶原三螺旋自组装并探索新的蛋白质设计策略.

关键词 胶原蛋白, 盐桥相互作用, 计算设计

胶原蛋白是组织工程中常用的支架材料, 它具有凝胶性、吸水性、高生物相容、可降解等优良特性^[1-4]. 胶原蛋白自组装的设计有可能影响从分子识别到细胞信号再到生物材料的许多生物学领域, 同时也是检验我们对蛋白质折叠和维持结构的基本理解的一项严格测试^[5-9]. 轴向电荷对(或称作肽链间盐桥相互作用)作为胶原蛋白疏水相互作用的可行替代品应用, 使我们能够以极大的特异性制备异源三聚体螺旋, 同时全面改善对螺旋结构和稳定性的控制^[10-13]. 了解它们的结构和能量学特征将使我们能够寻找或开发胶原蛋白模拟结构所需要的“设计规则”, 并辅助研究蛋白质结合和胶原相关的疾病. 本文总结了近期关于轴向

电荷对相互作用的理解以及它们是如何成功用于控制胶原三股螺旋自组装.

1 胶原蛋白简介

1.1 胶原蛋白的背景知识

胶原蛋白是高等动物中最丰富的蛋白质, 约占总蛋白质质量的1/3. 天然胶原蛋白由三条多肽链盘绕形成三股螺旋结构^[1], 这些肽链紧密堆积, 同时它们的排列会精确地错开一个氨基酸(图1A). 在胶原蛋白肽链中, 每三个位置就会被一个最小的氨基酸甘氨酸(Glycine, 简写Gly或G)所占据, 所以胶原蛋白的三股螺旋形成域

引用格式: 吕成, 强书敏, 孙添添, 等. 胶原蛋白异源三聚体的理性设计. 中国科学: 生命科学, 2019, 49: 615-624

Lu C, Qiang S M, Sun T T, et al. Computational design of heterotrimeric collagen helices (in Chinese). Sci Sin Vitae, 2019, 49: 615-624, doi: 10.1360/N052018-00260

是由一个典型的Gly-X-Y重复序列定义的, 其中X和Y位置分别是脯氨酸(Proline, 简写Pro或P)和(4R)-羟脯氨酸(Hydroxyproline, 简写为Hyp或O)^[2]. 这些重复序列可以延伸超过1000个氨基酸, 形成300 nm左右的三股螺旋结构. 具备三股螺旋结构的胶原蛋白分子进一步组装成更高级的长纤维网状结构, 这些三维空间网络为人体组织提供抗拉强度和柔韧性. 以寡聚化状态存在于细胞外基质中的胶原蛋白还可以帮助调节细胞行为、指导组织形态发生, 如通过纤维形貌、孔径尺寸、配体结合位点的空间分布等因素影响细胞黏附、增殖、分化. 此外, 胶原蛋白在调节细胞极性、启动血栓形成和调节肿瘤转移等方面也起着重要的作用^[3,4].

1.2 同源三聚体和异源三聚体

虽然胶原蛋白的结构简单, 但是理解胶原蛋白系统的一个主要挑战是多种同源与异源胶原蛋白并存. 目前, 在已知的人类28种胶原蛋白中^[14], 大多数的胶原蛋白类型是均质三聚体, 也就是说, 三股螺旋的所有三条肽链都是相同的; 而其中只有1/4的胶原蛋白构型(7/28)是以异源三聚体的形式存在, 即构建胶原蛋白三股螺旋的三条链具有不同的序列(图1)^[14,15]. 虽然总数上不及同源三聚体, 但是在真核生物的结缔组织中异源三聚体胶原蛋白的含量更为丰富. 例如, 在骨和皮肤中发现的形成原纤维的I型胶原蛋白, 它是由两个 $\alpha 1$ 和一个 $\alpha 2$ 链的结合形成的AB₂型胶原蛋白, 它们在细胞迁移、生长和分化尤其是止血中起着至关重要的作用; 而另一种非纤维状IV型胶原蛋白由三种不同的链组成ABC型胶原蛋白^[4,16]. 随着更多组件合并的可能性, 使其构型的复杂性呈指数增长. 以两个易形成三股螺旋的肽链A和肽链B的混合物为例, 该体系可以形成所有的(或任意竞争性的)同源三聚体和异源三聚体会有四种不同的组合方式, 分别为A₃, B₃, A₂B和AB₂; 而且肽链的单一残基在三重螺旋中相互错开排列的特征又增加了区分唯一靶向胶原蛋白的另一层复杂性, 如唯一的肽链B在A₂B螺旋中可以有出现在首链(BAA)、中链(ABA)或者尾链(AAB)三种不同位置的区分. 对于ABC型异源三聚体而言, 三种不同的肽链组分理论上可以形成 $3^3=27$ 个不同的三聚体组合形式. 理解异种特异性装配的分子间力对于分子级工程很重要, 多组分组件的组成控制可用于调节从金属-有机配体络合^[5,6]到DNA折纸^[7,8]和蛋白质纳米笼^[9]等各种分

子系统的材料性质.

1.3 胶原蛋白的折叠

蛋白质折叠的主要驱动力包括: 疏水作用、氢键与静电作用力等. 球状蛋白主要是通过无数静电和非共价弱相互作用辅助下的疏水塌陷进行折叠, 相比之下, 胶原蛋白三股螺旋结构的疏水包埋的作用很小. 在拓扑结构上, α 螺旋结构域与胶原蛋白最接近, 但 α 螺旋主要通过疏水相互作用来稳定寡聚体的形成. 胶原蛋白的原纤维区域没有疏水核, 三条多肽链都含有相邻的GXY氨基酸三联体, 折叠为右手超螺旋; Gly残基被隔离到螺旋的核心, 在那里它们通过形成多个链间氢键网络来稳定三重螺旋结构. 由于胶原蛋白序列的重复性, 我们可以设想单个链的交错连接, 其中一条或多条链的三个氨基酸被移位. 这种错位在两个方面影响了三股螺旋的稳定性. 随着链的移动, 在N和C端上的空悬部分不能形成链间氢键, 从而降低胶原蛋白的热稳定性.

胶原蛋白中富含带电荷氨基酸, 它们之间的两两交互的静电相互作用也可以提供额外的稳定性. 利用互补静电相互作用的策略^[4,10~16], 还可以通过沿蛋白质表面引入侧链相互作用网络来设计特异性^[17~20]. 例如, 在一个1014残基长的人类I型胶原蛋白三股螺旋结构中, 大约1/3的G-X-Y三联体参与了某种形式的良好的静电相互作用.

1.4 胶原蛋白相关疾病

胶原蛋白的序列突变会导致遗传性疾病. 在许多遗传性结缔组织疾病中, 其致病原因是由于胶原蛋白固定重复氨基酸序列遭到破坏. I型胶原蛋白作为骨骼中的主要蛋白质, 其任何链中的突变都会引起成骨不全症^[21]. 如果将其中一个甘氨酸改变为侧链更大的丝氨酸, 该突变会延迟蛋白质的折叠并降低胶原蛋白分子的稳定性; 而折叠缺陷导致细胞内胶原蛋白的保留及其最终死亡, 结构改变还会导致骨形成所需的其它组分的缺陷结合^[22,23]. 胶原蛋白的关键作用已经被超过1000种突变所引起广泛的疾病清楚地阐明(<http://www.le.ac.uk/genetics/collagen>), 迄今为止已在22种基因中鉴定出这些突变, 包含超过20种胶原蛋白中的12种类型^[24]. 这些疾病包括成骨不全、许多软骨发育不良、先天性结缔组织发育不全综合征的几种亚型、埃

图 1 胶原蛋白异三聚体结构、序列及盐桥. A: 胶原蛋白异三聚体结构示意图及序列(PDB ID: 5YAN); B: DK盐桥示意图; C: KD盐桥示意图

Figure 1 Construction, sequence and salt bridge of heterotrimer collagen. A: Structure and sequence of heterotrimer collagen ABC(PDB ID: 5YAN); B: schematic of lateral DK and axial DK; C: schematic of lateral KD and axial KD

布尔氏症候群、Bethlem肌病、大疱性表皮松解症的某些亚型、Knobloch综合征以及一些骨质疏松症、动脉瘤、骨关节病和椎间盘疾病^[25].

目前, 具有基因工程胶原蛋白突变的小鼠是理解胶原蛋白突变最有价值的研究手段, 已被广泛应用于定义各种胶原蛋白的功能及研究许多相关的疾病^[26-28]; 而整个胶原蛋白分子的大小和复杂性使得人们很难在结构基础上解释突变对胶原蛋白稳定性和折叠变化的影响. 幸运的是, 异常胶原蛋白的多肽模型已经开始帮助阐明氨基酸突变如何导致临床问题^[2]. 最近, Gauba和Hartgerink^[11,29]报道了一种可用来研究成骨不全症的多肽模型. 研究人员将 I 型胶原蛋白序列片段合并到模型肽的中间, 在这些插入的序列中携带 0~3 个成骨不全甘氨酸-丝氨酸突变. 位于胶原蛋白序列区域两侧的正、负、中性序列仍然能够将多肽组装成胶原蛋白异三聚体. 该模型允许它们在组成胶原蛋白的三条肽链中的任何一条或全部中产生与疾病相关的突变. 通过对三聚体到单体的热跃迁的测量表明, 单

个突变极大地降低了这些肽的稳定性. 并且实验结果与计算研究一致, 计算了将这种突变引入胶原蛋白三股螺旋的效果. 当研究肽链三聚体的折叠时, 他们看到了同样的趋势: 第一个甘氨酸突变造成三股螺旋折叠的最大延迟. Gauba和Hartgerink^[29]提供的方法, 还允许我们改变这些肽链上突变位点周围的序列, 以及通过不同方式取代甘氨酸, 研究影响成骨不全症状严重程度的因素. 尽管临床症状取决于含有这种突变的胶原蛋白的功能和位置, 原则上这种研究方法还可以进一步扩展到由胶原蛋白的甘氨酸突变引起的其他疾病的致病机理, 例如, 阿尔波特综合征由于 IV 型胶原蛋白的突变导致肾病, Ehler-danlos 综合征由于 III 型胶原蛋白的突变影响血管功能^[25].

1.5 胶原蛋白设计存在的挑战

长期以来, 多肽模型一直是胶原蛋白结构研究中不可或缺的组成部分, 作为探索胶原蛋白组装稳定性、特异性和化学计量学分子基础的重要工具, 多肽

模型提供了改变氨基酸序列的机会, 允许定义突变位点变化的分子细节及其相互作用, 是理解胶原蛋白突变病理学以及正常胶原蛋白结构特征的有价值的模型. 迄今为止, 大多数模型多肽的工作都集中在同源三聚体上. 如何通过理性或计算设计肽链之间的静电相互作用力, 形成非共价键诱导的胶原蛋白异源三聚体是目前的难点问题^[30-32].

从20世纪90年代以来, 胶原蛋白模式多肽研究均为同三聚体. 虽然在目前发现的28种天然胶原类型中有7种为异三聚体, 其中包括最主要的I型胶原蛋白, 但是由于胶原蛋白折叠机制复杂, 迄今尚不能体外重构天然序列来源的胶原异三聚体. 如果能人工设计胶原蛋白异三聚体, 则不仅可作为更接近天然序列的模式多肽来解析胶原相关疾病的致病机理, 还可进一步构建可特异性修饰单链与精确调控的胶原蛋白纳米材料.

胶原蛋白模拟系统的设计由于缺少疏水核心而变得更具挑战性. 从序列结构的角度揭示胶原蛋白三股螺旋稳定性的分子机制由于蛋白质折叠问题会变得十分复杂. 尽管天然胶原蛋白三股螺旋的整体蛋白质折叠相对简单, 但设计特定异源胶原蛋白非常困难. 首先, 高解析度的胶原蛋白分子结构数量较少, 蛋白质数据库(Protein Data Bank)^[33]中目前只有大约40个胶原蛋白结构, 并且以高脯氨酸/羟脯氨酸含量的胶原蛋白序列为主^[34], 从而缺乏对驱动三股螺旋形成的作用力的了解. 在这些结构中, 除了主链间氢键之外, 侧链间的强相互作用, 如侧链盐桥等信息稀少. 其次, 位于胶原蛋白表面上的盐桥高度暴露于溶剂分子表面, 这些基团对折叠和稳定性的能量贡献高度依赖于与水的相互作用. 在计算方法中, 溶剂对蛋白质折叠和结构的贡献的建模一直是一个主要的挑战, 试图在连续溶剂化的效率和显式水模型的准确性之间取得平衡. 此外, 胶原蛋白的折叠是缓慢的, 而不是双态的, 使氨基酸序列对自然状态和展开状态能量贡献的热力学特性复杂化.

由于胶原蛋白具有无毒、高生物相容、可降解等优良特性, 所以胶原蛋白作为重要的生物医用材料被广泛应用在伤口缝线、软骨缺损填充、骨骼金属植入物镀膜等. 长期以来, 胶原蛋白主要从动物皮肤中提取, 所以其来源比较混杂, 甚至会利用旧皮革制品中提取的有毒胶原蛋白生产药物胶囊, 而且动物来源的

胶原蛋白可控性较差、易被动物源带有的疾病(如朊蛋白)污染等问题. 目前, 制备清洁来源胶原蛋白主要可分为两种途径: 化学合成胶原样多肽和生物异源表达, 如转基因植物培育、动物细胞培养、宿主微生物表达等. 化学合成的胶原样多肽具有可控性好、易于修饰等优点, 但是制备成本高、无法批量生产. 高等真核表达体系的普遍问题是培养成本高、周期长、表达量低; 通过基因重组技术由微生物表达体系制备胶原蛋白的制备方法虽然成本低、表达量高, 但是胶原蛋白分子难以在纳米尺度上形成有序的空间排列. 所以, 如何制备清洁来源、纳米尺度可调控的胶原蛋白材料一直是蛋白质工程与生物材料工程交叉领域中的热点问题.

2 盐桥的相互作用力

在多肽模型中, 每一对相互作用的热力学和三维分子结构可以从自然态下的能量和结构因素的复杂相互作用中解构出来. 由于在长序列上重复的GXY三联体基序以及短肽形成天然胶原蛋白的能力, 胶原蛋白特别适合这种解剖. Brodsky课题组^[35]全面研究了各种氨基酸在同源三聚体 $(\text{GPO})_3\text{GXY}(\text{GPO})_4$ (P=脯氨酸和O=羟脯氨酸)宿主序列中X或Y位置的单取代作用, 并将取代后获得的熔融温度值(T_m)用于预测另一胶原蛋白多肽 $(\text{GPO})_3\text{GXYGX}'\text{Y}'(\text{GPO})_3$ 的热稳定性^[36]. 然后将预测的 T_m 与实验值进行比较, 得出了取代残基之间的两两相互作用. 其中令人感兴趣的结果是, 在进行Gly-Lys-Asp和Gly-Lys-Glu序列替换时, 实验测得的 T_m 值与没有任何取代的 $(\text{GPO})_8$ 的 T_m 值比较, 没有发生很大的变化^[37]. 这表明该序列替换因其之间成对的高度稳定性从而补偿因原序列中单个氨基酸替换引起的稳定性破坏, 并首次提出了三个肽链间Lys-Asp/Glu存在盐桥相互作用. 胶原蛋白的三股螺旋折叠构象将相邻链之间的相互作用限制为两个几何类别: 平行(Y-X')和垂直(Y-X)于螺旋轴, 分别称之为轴向盐桥和侧向盐桥. 在序列位置上, Y位置的首链或中链的赖氨酸(K) n 与X位置的中链或尾链的 $n+2$ 位天冬氨酸(D)/谷氨酸(Q), 分别形成轴向盐桥和侧向盐桥(图1B). 在后随链的赖氨酸 n 和首链的 $n+5$ 位酸性残基之间也可能存在对称的轴向盐桥. 由于Y-X和Y-X'的不同排列, 对Lys和Asp形成的盐桥将存在的四种组合形式, 导致

八种类型的电荷对相互作用。

为了进一步研究轴向和平行两种类型的盐桥相互作用, Hartgerink课题组^[38]利用X射线晶体学和溶液核磁共振解析了含有KGD和KGE序列的胶原蛋白样肽的分子结构。两个不同序列多肽的晶体结构中都能找到一个明确定义的轴向盐桥, 而利用核磁共振(nuclear magnetic resonance, NMR), 在每个三股螺旋中均显示存在两个轴向盐桥, 这与Persikov等人^[35]提出的盐桥结构一致。尽管在这两种多肽的晶体结构中都没有发现侧向盐桥, 但NMR实验中通过对带电侧链取样异质构象群, 其中有些构象能够观察到带电侧链的侧向接触。在KGE多肽链中采集了具有和没有侧向盐桥的构象, 带电侧链同时与主链和其他侧链极性原子都会相互作用; 相反, 在KGD多肽链采集的少量具有侧向盐桥的构象中, 侧链主要与主链发生作用。由于带电侧链构象的高度灵活性以及与主链之间发生相互作用的倾向性, 使得侧向盐桥的弱相互作用“受挫”, 其表现出的盐桥构象稳定性也不如轴向盐桥。同样在ABC型异源三聚体三股螺旋的NMR结构中也得到了类似的结论, 即使两种类型的盐桥在空间上都可以被Lys-Asp对呈现, 结构中也只能观察到侧链在轴向方向的接触^[39]。在GEK序列的晶体结构中, 虽然两个侧向相互作用在几何上也是可行的, 但是也没有观察到侧向盐桥^[40]。这些结果也同时解释了为什么GEK和GDK序列比未取代(GPO)₈螺旋少12和16℃的热稳定性^[36]。

通过这些短肽的研究发现, 单个氨基酸的位置偏好和多个氨基酸取代对稳定性和螺旋结构的累积效应与天然胶原蛋白中的优先序列基序相关。在XYG重复序列中, 真核胶原蛋白中的酸性残基显示出对X位置的偏好, 而赖氨酸(K)更喜欢Y位置。目前还没有报道这种位置偏好反向的DGK或EGK序列的热稳定性或分子结构, 从而无法与KGD或KGE基序进行直接比较。分子建模表明, 在DKG/EGK序列中, 任何可能存在的假定的轴向或侧向盐桥的几何形状将与KGD/KGE序列中观察到的情况相反。Hartgerink课题组^[41]将这些所谓的反向轴盐桥并入ABC异源三聚体中, 并与轴向盐桥稳定的Asp-Lys和Glu-Lys对相比, 这些具有反向盐桥的异源三聚体的稳定性明显较差, 并且在组成上更具有异质性。基于相互作用的带电基团之间的分离, 在四个反向的电荷对类型中, 轴向DK(7.05±1.43 Å)和侧向KD(6.3±0.62 Å), 在结构中表现出的相互作用始

终非常弱。而对于轴向KD(2.9±0.19 Å)和侧向DK(4.6±0.76 Å), 则始终表现出较强的耦合^[42]。这表明一种弱相互作用类似于在侧向盐桥观察到的, 并且与真核胶原蛋白中异常低频率的DGK或EGK密切相关^[43]。虽然RGE和RGD序列的热稳定性分别比(POG)₈三股螺旋低8和4.5℃^[36], 而其稳定性的减弱与现有的结构数据无关。例如, III型胶原蛋白的异源三聚体Gly991-Gly1032序列的晶体结构包含一个RGE基序, 该基序形成两个明确定义的轴向Arg-Glu盐桥^[44]。在另一种情况下, 胍基(精氨酸)在酸性残基存在的情况下会优先与主链和侧链极性原子相互作用会形成轴向作用^[45]。这些结果表明, 精氨酸残基的盐桥相互作用对局部环境有很强的依赖性。精氨酸作为未配对的残基, 与中性肽原子形成多极接触, 是Hyp后Y位置最稳定的氨基酸。为了稳定负电荷组分而形成的盐桥可能会受到精氨酸的其他相互作用的充分阻碍, 从而导致这种配对不够有效^[46,47]。

3 设计方法

天然胶原蛋白的残基组分、序列组成、纤维形貌等具有高度多样性, 由于缺乏对胶原蛋白链间特异性识别与折叠机制的了解, 胶原蛋白材料的设计与胶原蛋白相关遗传病机理等的研究长期受限于同三聚体序列。轴向电荷对相互作用的应用, 尤其是Lys-Asp和Lys-Glu之间的相互作用, 允许折叠更多高度受控的异源三聚体胶原蛋白螺旋。通过第二部分对盐桥相互作用的讨论, 异源三聚体胶原蛋白类似物设计的有四个关键规则^[48]: (i) 轴向盐桥构象强健, 因此非常适合胶原蛋白螺旋设计; (ii) 侧向盐桥, 虽然在几何上是可能的, 但提供很少的稳定性^[42], 是螺旋设计的不利选择; (iii) 轴向盐桥在X位置的酸性残基和Y位置的赖氨酸(KGD/KGE)比位置相反时(DKG/EGK)提供了更大的热稳定性和成分均匀性; (iv) 与赖氨酸不同的是, 精氨酸很大程度上取决于其在三重螺旋中的局部环境, 而形成的任何盐桥似乎都不具有显著的稳定性。这些设计规则可以帮助我们进行异源三聚体胶原蛋白类似物的非共价设计。

3.1 理性设计胶原蛋白异三聚体

通常三股螺旋是由脯氨酸和羟脯氨酸的构象约束

来稳定的, 轴向电荷对的相互作用提供了另一种稳定机制。Hartgerink课题组^[39]通过设计这两种因素在胶原蛋白结构稳定性上的权衡, 引入了一种由(PKG)₁₀、(EPG)₁₀和(DKG)₁₀链组成的结构稳定的ABC异三聚体。虽然带电荷的残基降低了单个肽链的三重螺旋倾向, 但在异三聚体中形成了盐桥, 为折叠提供了稳定性和很大的驱动力。其中含Lys-Asp盐桥(KD)的ABC异三聚体最为稳定; 其次是Lys-Glu(KE)和Arg-Glu(RE)的盐桥, 两者 T_m 值相近。含有Arg-Asp盐桥(RD)的ABC异三聚体表现出非常差的热稳定性; 并且在热熔融曲线上出现了两个展开跃迁, 表明其成分纯度不高。

这些分析形成了之后大部分理性设计胶原蛋白异源三聚体的方法基础。Hartgerink课题组^[17]使用Lys-Glu和Lys-Asp盐桥设计了一个稳定的、电中性的AAB杂化三聚体, 它由两个(PKGPOG)₅和一个(DOG)₁₀或(EOG)₁₀链组成。研究的一个重要观察结果是, 使用Lys-Glu盐桥(KE)稳定的异三聚体比使用Lys-Asp(KD)稳定的异三聚体有更好的排列控制。虽然这种现象学效应的分子基础尚不清楚, Jalan等人^[49]还是利用它成功地演示了同时控制AAB-异三聚体。Jalan等人^[49]在AAB设计中做了两个关键的修改: 首先, 将Lys-Asp和Lys-Glu盐桥组合到目标状态中, 在保持足够稳定性的同时, 提高特异性; 其次, 设计目标异源三聚体, 使其仅在目标蛋白中形成最大的盐桥数量。而其他7个竞争肽含有较少的盐桥和较多未配对的残基。以(PKGEOG)₅的同源三聚体为模板, 分别用羟基脯氨酸(O)和天冬氨酸(D)取代后随链中未配对的赖氨酸和谷氨酸残基。这些修饰增加了异三聚体相对于(PKGEOG)₅同源三聚体的热稳定性, 使其相差19℃, 并使其能对组合和排列进行同时控制。这里设计的异源三聚体不是电中性的, 含有周期性排列的谷氨酸残基。从折叠到更高阶结构的角度来看, 静电电荷稳定三重螺旋结构的设计是有用的, 但对于蛋白质结合研究和疾病建模而言, 电中性螺旋结构可能更好。这些更真实的折叠的异三聚体多肽为直接观察胶原蛋白疾病的结构扰动、合理设计治疗药物等方面提供结构模型, 进一步揭示突变如何影响高阶结构的形成以及与骨骼其他成分的相互作用。

3.2 计算设计胶原蛋白异三聚体

对于序列空间的有效探索是计算蛋白质设计领域

的重要概念, 并显著拓宽了人们对蛋白质结构与功能的理解^[50]。计算设计折叠异三聚体胶原蛋白多肽的目标是开发一种计算方法, 优化胶原蛋白ABC异源三聚体的稳定性, 同时应避免竞争性聚合体的形成。这两种目的通常分别被称为“正向”和“负向”设计(图2A), 典型的设计方案是从一个随机或非最优的序列映射到表示目标结构的三维框架上开始。突变可以单独引入, 也可以成组引入, 计算稳定性的变化取决于有利和不利交互的扰动。在模拟过程中, 有害的突变被舍弃, 稳定突变被保持。在“正向”设计中, 首要目标是获得稳定的状态, 所有的计算都集中在最大化目标拓扑中有利相互反应的数量上(图2B)。在“负向”设计中, 当评估突变的影响时, 可以显式或隐式地考虑交替构象状态。

我们寻求应用计算方法来优化螺旋结构域之间的界面静电相互作用, 以设计折叠异源三聚体胶原蛋白多肽^[51]。由于缺乏关于胶原蛋白两两相互作用的数据, 在计算方法中需要几个近似来定义能量函数。在这个计分函数中, 三股螺旋态的能量近似关联于标准的(POG)₁₀同源三聚体, 由于每一个POG三联体的存在使得体系的单体能量下降1.5 kcal/mol, 为胶原蛋白提供良好的骨架稳定性, 这种中性基序我们赋值为零。对不稳定螺旋的赖氨酸和天冬氨酸残基点突变赋负值, 而对满足轴向Lys-Asp盐桥序列要求的双突变赋正值。因为Arg的长而灵活的侧链更容易避免不利的相互作用, Arg-Arg斥力的能量权重小于Glu-Glu。然后, 利用从螺旋结构域设计中改编的计分函数^[48], 计算了目标ABC异三聚体和26种非理想状态(AAA, AAB, AAC, ..., CCA, CCB, CCC)的相互作用能量:

$$E_{i,j} = \begin{bmatrix} +2, R/R \\ +3, E/E \\ -1, R/E, E/R \end{bmatrix} \quad (1)$$

同时优化稳定性和特异性, 将计算出的靶向异源三聚体ABC稳定性与其他竞争异源三聚体和同源三聚体分离开来。ABC物种的稳定性采用“正向”设计优化, 在离散的、基于序列的相互作用模型中最大化有利相互作用的数量。针对特异性采用“负向”设计, 将相同的离散交互模型应用于竞争物种, 并优化基于玻尔兹曼概率的特异性评分(图2B):

$$P_{ABC} = e^{-\beta E_{ABC}} / \sum_{i=1}^{27} e^{-\beta E_i} \quad (2)$$

图2 胶原蛋白异三聚体的计算序列优化策略. A: 多种组合可能及正负设计策略; B, C: 计算机辅助优化过程^[51]

Figure 2 Computational sequence optimization strategy of collagen heterotrimer. A: Multiple combination possibilities and positive and negative design strategies; B, C: optimization process by computer design^[51]

其中 $i \in (\text{AAA}, \text{AAB}, \text{AAC}, \dots, \text{CCC})$; β 为系统温度, 设置为1.0. 如果所有的肽在溶液中被预测以ABC的形式聚集, P_{ABC} 接近于1. 如果没有形成ABC, P_{ABC} 趋近于零. 为了评估正设计和负设计对能量和特异性分布的相对贡献, 我们对每个场景进行了1000次模拟: 单独进行正设计, 单独进行负设计, 或者同时优化稳定性和特异性(图2C). 使用天然目标与不希望的竞争态的最低能量之间的能量差:

$$E_{\text{gap}} = E_{\text{ABC}} - \min(E_i)_{i \neq \text{ABC}} \quad (3)$$

基于此简单能量函数的遗传搜索算法对肽序列进行了优化, 增加了目标序列与其26个竞争态之间的能量差距. 合成的ABC序列使用这种方法得到的分数为-12, 这表明竞争肽与目标序列在能量上最少相差了12个未配对残基. 即使在这样一个限制的搜索空间, 三元混合物的A, B和C序列自我组装成一个高度稳定的单一成分——单一的ABC异源三聚体(图1A). 这些设计的实验表征与计算预测不匹配, 经过仔细评估, 发现方法上的误差和能量函数的不足都是不准确的. 然而, 这第一轮的设计提出的表征与计算预测不匹配的问题将与未来成功构建异源聚体胶原蛋白息息相关. 该领域的未来工作应集中于扩大算法可用的序列空间, 这将需要对氨基酸成对相互作用进行详细研究.

该胶原蛋白ABC异源三聚体设计的一个特征是三种肽链的酸性和碱性三联体近乎平衡. 肽链A主要是酸性的, B主要是碱性的, 而C为一半酸性一半碱性.

在得分最高的结构中, 酸性和碱性三联体之间的差异从未超过4/30. 这在正设计方面是有意义的, 有利电荷对的数目越大, ABC三聚体的能量越低. 静电平衡对特异性的作用是电荷差异越大, 无论是酸性的还是碱性的残基, ABC和下一个最稳定的状态之间的能量差距越小. 因此, 静电平衡确保了稳定性和特异性. 在蛋白质浓度非常高的细胞中, 序列多样性可以防止非特异性聚集. 平衡电荷的另一个结果是三种肽的序列多样性增加:

$$S = \ln \left(\sum_i^{20} N_i / \prod_i^{20} N_i! \right), \quad (4)$$

其中 N_i 是特定氨基酸类型的实例数量.

通过以上计算优化所得到的胶原蛋白异源三聚体序列, 获得了1.77 Å分辨率的晶体结构(PDB ID: 5YAN)^[42]. 从原子水平上验证了全新计算设计的方法在胶原蛋白序列设计的准确性与实用性. 在此晶体结构基础之上, 借助分子动态模拟胶原蛋白结构, 采集了带电荷残基侧链相互作用中可能出现的各种构象, 以及统计强静电作用出现的概率; 同时通过设计丙氨酸定点突变, 破坏侧链间的静电引力, 以圆二色谱测量热融合温度的变化, 对侧链静电引力对胶原蛋白热稳定性的贡献进行排序: 轴向KD>侧向DK>轴向DK≈侧向KD, 与本节开始部分提到的设计规则相符合. 结合实验测定与动态模拟研究结果, 找到了特定相互作用对能量学的贡献并提出了胶原蛋白序列设计的经验公式:

$$T_m = \sum_{i=1}^4 n_i p_i \Delta T_{m,i} + \sum_{j=1}^4 n_j \Delta T_{m,j} \quad (5)$$

对于盐桥类型*i*, n_i 是观察数, $\Delta T_{m,i}$ 是基于诱变的稳定性的贡献, p_i 是从分子模拟获得的权重. 该预测模型可准确地预测胶原蛋白的热稳定性. 在该计算方法中, 避开对胶原蛋白折叠过程的模拟, 直接预测假定最终态的热稳定性; 避开溶剂分子与侧链间相互竞争的计算, 而是直接以侧链间的距离来标度静电作用的强度, 在连续溶剂化的效率和显式水模型的准确性之间取得平衡. 通过优化立场与计算方法, 从而降低了计算成本, 同时推动胶原蛋白设计向计算准确性、序列多样性、功能完善性等多方面发展, 促进用于工程新型胶原蛋白结构的开发.

4 总结与展望

蛋白质从头设计和计算蛋白质设计的目标是将关于分子间相互作用和蛋白质折叠的知识应用于构建适用于新型生物学或工业应用的新蛋白质. 迄今为止,

胶原蛋白体系的计算设计在很大程度上是由三股螺旋的分子结构和稳定性问题驱动的. 蛋白质从头设计是从目标蛋白出发来设计, 我们已成功设计胶原蛋白异三聚体; 计算蛋白质设计是从氨基酸序列出发设计, 基于已获得的分子序列, 进行进一步计算蛋白设计, 建立极性不带电荷、疏水残基侧链相互作用力与胶原蛋白热稳定性的相关性经验公式; 其次, 以计算机模拟退火算法设计残基侧链多样化、环境耐受性高的胶原蛋白异三聚体; 最后, 将计算设计的胶原蛋白序列整合到天然序列中作为胶原蛋白纤维间的非共价键作用位点, 调控胶原蛋白纤维的空间排列, 并以微生物异源表达制备与材料学实验验证. 随着我们从对序列-结构关系的理解转向结构-功能模式, 将会促进我们设计越来越复杂的胶原蛋白螺旋, 其序列选择更接近于自然界中存在的序列选择. 这不仅是对生物大分子折叠机理的科学探索, 胶原蛋白材料可作为组织再生、伤口愈合和细胞移植等生物医学基底材料, 胶原蛋白纳米材料的研发符合国家对于发展纳米新材料与高值医用耗材等高端制造业的战略需要.

参考文献

- Bella J, Brodsky B, Berman H M. Hydration structure of a collagen peptide. *Structure*, 1995, 3: 893–906
- Brodsky B, Baum J. Modelling collagen diseases. *Nature*, 2008, 453: 998–999
- Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol*, 2014, 15: 786–801
- Johansson C, Butkowski R, Wieslander J. The structural organization of type IV collagen. Identification of three NC1 populations in the glomerular basement membrane. *J Biol Chem*, 1992, 267: 24533–24537
- Sepehrpour H, Saha M L, Stang P J. Fe–Pt twisted heterometallic bicyclic supramolecules via multicomponent self-assembly. *J Am Chem Soc*, 2017, 139: 2553–2556
- Fujita D, Ueda Y, Sato S, et al. Self-assembly of tetravalent Goldberg polyhedra from 144 small components. *Nature*, 2016, 540: 563–566
- Han D, Pal S, Nangreave J, et al. DNA origami with complex curvatures in three-dimensional space. *Science*, 2011, 332: 342–346
- Mao C, LaBean T H, Relf J H, et al. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 2000, 407: 493–496
- Fletcher J M, Harniman R L, Barnes F R H, et al. Self-assembling cages from coiled-coil peptide modules. *Science*, 2013, 340: 595–599
- Fallas J A, Lee M A, Jalan A A, et al. Rational design of single-composition ABC collagen heterotrimers. *J Am Chem Soc*, 2012, 134: 1430–1433
- Gaub V, Hartgerink J D. Surprisingly high stability of collagen ABC heterotrimer: Evaluation of side chain charge pairs. *J Am Chem Soc*, 2007, 129: 15034–15041
- Gaub V, Hartgerink J D. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J Am Chem Soc*, 2007, 129: 2683–2690
- Jalan A A, Demeler B, Hartgerink J D. Hydroxyproline-free single composition ABC collagen heterotrimer. *J Am Chem Soc*, 2015, 135: 6014–6017
- Heino J. The collagen family members as cell adhesion proteins. *Bioessays*, 2007, 29: 1001–1010
- Ricard-Blum S. The collagen family. *Cold Spring Harbor Perspectives Biol*, 2011, 3: a004978

- 16 Tolstoshev P, Haber R, Crystal R G. Procollagen $\alpha 2$ mRNA is significantly different from procollagen $\alpha 1(I)$ mRNA in size or secondary structure. *Biochem Biophys Res Commun*, 1979, 87: 818–826
- 17 O’Leary L E R, Fallas J A, Hartgerink J D. Positive and negative design leads to compositional control in AAB collagen heterotrimers. *J Am Chem Soc*, 2011, 133: 5432–5443
- 18 Traub W, Fietzek P P. Contribution of the $\alpha 2$ chain to the molecular stability of collagen. *FEBS Lett*, 1976, 68: 245–249
- 19 Hulmes D J S, Miller A, Parry D A D, et al. Analysis of the primary structure of collagen for the origins of molecular packing. *J Mol Biol*, 1973, 79: 137–148
- 20 Salem G, Traub W. Conformational implications of amino acid sequence regularities in collagen. *FEBS Lett*, 1975, 51: 94–99
- 21 Marini J C, Forlino A, Cabral W A, et al. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: Regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat*, 2010, 28: 209–221
- 22 Byers P H, Wallis G A, Willing M C. Osteogenesis imperfecta: translation of mutation to phenotype. *J Med Genet*, 1991, 28: 433–442
- 23 Beck K, Chan V C, Shenoy N, et al. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc Natl Acad Sci USA*, 2000, 97: 4273–4278
- 24 Kuivaniemi H, Tromp G, Prockop D J. Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum Mutat*, 1997, 9: 300–315
- 25 Myllyharju J, Kivirikko K I. Collagens and collagen-related diseases. *Ann Med*, 2001, 33: 7–21
- 26 Kivirikko K I. Collagens and their abnormalities in a Wide Spectrum of Diseases. *Ann Med*, 1993, 25: 113–126
- 27 Olsen B. New insights into the function of collagens from genetic analysis. *Curr Opin Cell Biol*, 1995, 7: 720–727
- 28 Prockop D J, Kivirikko K I. Collagens: Molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem*, 1995, 64: 403–434
- 29 Gauba V, Hartgerink J D. Synthetic collagen heterotrimers: structural mimics of wild-type and mutant collagen type I. *J Am Chem Soc*, 2008, 130: 7509–7515
- 30 Thomson A R, Wood C W, Burton A J, et al. Computational design of water-soluble α -helical barrels. *Science*, 2014, 346: 485–488
- 31 Grigoryan G, Reinke A W, Keating A E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature*, 2009, 458: 859–864
- 32 Xu F, Zahid S, Silva T, et al. Computational design of a collagen A:B:C-type heterotrimer. *J Am Chem Soc*, 2011, 133: 15260–15263
- 33 Berman H M, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*, 2000, 28: 235–242
- 34 Kramer G R, Berman H M. An overview of structural studies of the collagen triple helix. In: Bansal M, Srinivasan N, eds. *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*. Singapore: World Scientific, 2013. 17–37
- 35 Persikov A V, Ramshaw J A M, Brodsky B. Prediction of collagen stability from amino acid sequence. *J Biol Chem*, 2005, 280: 19343–19349
- 36 Persikov A V, Ramshaw J A M, Kirkpatrick A, et al. Peptide investigations of pairwise interactions in the collagen triple-helix. *J Mol Biol*, 2002, 316: 385–394
- 37 Persikov A V, Ramshaw J A M, Kirkpatrick A, et al. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry*, 2005, 44: 1414–1422
- 38 Fallas J A, Dong J, Tao Y J, et al. Structural insights into charge pair interactions in triple helical collagen-like proteins. *J Biol Chem*, 2012, 287: 8039–8047
- 39 Fallas J A, Gauba V, Hartgerink J D. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J Biol Chem*, 2009, 284: 26851–26859
- 40 Kramer R Z, Venugopal M G, Bella J, et al. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J Mol Biol*, 2000, 301: 1191–1205
- 41 Wei F, Fallas J A, Hartgerink J D. Sequence position and side chain length dependence of charge pair interactions in collagen triple helices. *Macromol Rapid Commun*, 2012, 33: 1445–1452
- 42 Zheng H, Lu C, Lan J, et al. How electrostatic networks modulate specificity and stability of collagen. *Proc Natl Acad Sci USA*, 2018, 115: 6207–6212
- 43 Ramshaw J A M, Shah N K, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J Struct Biol*, 1998, 122: 86–91
- 44 Boudko S P, Engel J, Okuyama K, et al. Crystal structure of human type III collagen Gly⁹⁹¹–Gly¹⁰³² cystine knot-containing peptide shows both 7/

- 2 and 10/3 triple helical symmetries. *J Biol Chem*, 2008, 283: 32580–32589
- 45 Russell L E, Fallas J A, Hartgerink J D. Selective assembly of a high stability AAB collagen heterotrimer. *J Am Chem Soc*, 2010, 132: 3242–3243
- 46 Emsley J, Knight C G, Farndale R W, et al. Structure of the integrin $\alpha\beta 1$ -binding collagen peptide. *J Mol Biol*, 2004, 335: 1019–1028
- 47 Yang W, Chan V C, Kirkpatrick A, et al. Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. *J Biol Chem*, 1997, 272: 28837–28840
- 48 Nanda V, Zahid S, Xu F, et al. Computational design of intermolecular stability and specificity in protein self-assembly. *Methods Enzymol*, 2011, 487: 575–593
- 49 Jalan A A, Hartgerink J D. Pairwise interactions in collagen and the design of heterotrimeric helices. *Curr Opin Chem Biol*, 2013, 17: 960–967
- 50 Stiebritz M T. MetREx: A protein design approach for the exploration of sequence-reactivity relationships in metalloenzymes. *J Comput Chem*, 2015, 36: 553–563
- 51 Xu F, Zhang L, Koder R L, et al. *De novo* self-assembling collagen heterotrimers using explicit positive and negative design. *Biochemistry*, 2010, 49: 2307–2316

Computational design of heterotrimeric collagen helices

LU Cheng, QIANG ShuMin, SUN TianTian & XU Fei

School of Biotechnology, Jiangnan University, Wuxi 214122, China

As an important biomedical material, collagen has been widely used in wound sutures, cartilage defect filling, and bone metal implant coating. Seven out of the twenty-eight types of natural collagen are heterotrimers comprising three chains with different primary sequences. It is important to understand the intermolecular forces underlying the hetero-specific assembly of natural collagen to modulate material properties in protein engineering and biomaterials engineering. Till date, the most important pairwise interaction identified in the collagen triple helix is the axial charge pair, which can be formed between appropriately placed lysine and aspartate or glutamate residues. These pairwise amino acid interactions enable the preparation of heterotrimeric helices with high specificity, as well as excellent control over helix structure and stability. Further, detailed studies on these interactions will allow the modification of natural collagen sequences to develop clean-sourced, highly controllable, and compositionally responsive collagen biomaterials. This review will summarize our understanding of this interaction and other charged-pair interactions and describe how they have been successfully used to control collagen triple helix self-assembly and explore new protein design strategies.

collagen heterotrimer, salt bridge interaction, computational design

doi: [10.1360/N052018-00260](https://doi.org/10.1360/N052018-00260)