

Fn 氨基酸的全部序列. 组成一级结构氨基酸残基的特点是: a. 高度保守性. 不同种属哺乳动物之间, 序列差异不到 10%, b. 含大量的重复序列. 重复序列分三型: I 型、II 型和 III 型. c. 从 cDNA 序列得知, 机体内存在几种不同类型的 Fn 多肽, 它们之间的差异集中在三个可变区域: EIIIA, EIIIB 和 V 区 (参见图 1)^[1]. 三个可变区的位置是: EIIIB 插入在 III₇ 和 III₈ 之间, EIIIA 在 III₁₁ 与 III₁₂ 之间, V 区在 III₁₄ 与 III₁₅ 之间.

每条单链中含 12 个 I 型, 2 个 II 型, 15 个 III 型重复序列和三个可变区. 每个 I 型结构长约 45 个氨基酸, 含 4 个 Cys 残基 (仅 I—12 例外, 它含 6 个 Cys), Cys1 与 Cys2, Cys3 与 Cys4 配对构成两对二硫键, 形成双环结构. 二级结构分析显示第一个环呈反向平行的 β 结构, 其保守的一个 Gly 残基使肽链在环的尾部折叠.

I 型结构长约 60 个残基, 四个保守的 Cys 构成两对二硫键 (Cys1-Cys3, Cys2-Cys4), 形成与 I 型结构不同的双环结构, II 型结构中有 7 个保守的芳香族氨基酸残基. III 型结构主要位于分子的中央区域, 每个 III 型结构约含 90 个氨基酸, 无 Cys, 含三个高度保守的芳香族氨基酸, 围绕这三个残基的氨基酸也相当保守. 在 III₇ 和 III₁₅ 中还分别有一个自由巯基.

上述三种重复序列上一定残基的高度保守性, 表明每个重复结构可能都是一个基本功能单位. Fn 的 I 型结构与组织型纤溶酶原激活因子 (tissue plasminogen activator, t-PA) N 端 43 个残基组成的指状区 (finger domain) 相似, 二者均构成纤维蛋白的结合点, Fn 可能与 t-PA 竞争结合纤维蛋白^[2]. 在 IV 型胶原酶的序列中也含 I 型结构. II 型结构在凝血因子 XII 和两种牛精液蛋白中也能找到, 均由两个外显子编码. 它们与 t-PA, 尿激酶, 纤溶酶原等分子中的环饼状结构 (kingles) 同源, 二者 β 片层及二硫键的模式也明显一致, 可能由一个共同前体所演化, 推测它们各自的功能相似, 即介导多功能区的蛋白质与其它分子的结合. 同样 III 型结构也在其它蛋白分子中出现, 在一种基

质蛋白 (cytotactin) 中含 11 个 III 型结构; 一种神经细胞表面蛋白分子接触素也含两个 III 型结构. Fusab^[3] 用一个分子量为 11 500 的 Fn 片段来研究 III 型结构的功能, 此片段为 III₁₀ 的序列, 含细胞结合点, 当在该片段 N 端不断延长, 加入与天然 Fn 序列相应的 III 型结构时, 重组的片段对细胞的亲和力不断增加, 影响程度同所加入的 III 型结构与 III₁₀ 的距离呈负相关, 改变各 III 型结构的相对位置大大降低了重组片段亲合细胞的活性. 如果在 III₁₀ 中去除细胞的结合位点 RGDS 肽 (Arg-Gly-Asp-Ser), 其亲和细胞的活性完全丧失. 这些结果表明 III 型结构的数量、型号及所处位置决定 Fn 粘附细胞的活性.

1.2 多功能区结构

每条 Fn 单链有 30 个链内二硫键, 分布在多肽链的两端, N 端 22 个, C 端 8 个. 它们使得整个肽链折叠成多个球形区域结构 (domains), 在单链上呈线状排列. Fn 在生理病理中有着广泛的作用, 其功能有区域性特点. 主要的几个功能区包括: N 端功能区、明胶结合区、肝素高亲和性结合区、细胞亲合区、C 端功能区及中部一长段 III 型结构组成的中央区域 (参见图 2)^[4].

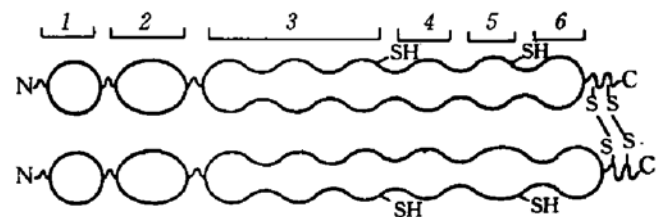


图 2 纤维粘连蛋白各球型功能区

1 N 端功能区, 2 明胶结合区, 3 中央区域, 4 细胞结合区, 5 肝素高亲和性结合区, 6 C 端功能区.

用蛋白酶降解 Fn 分离到一个分子量为 27 000 的 N 端片段即 N 端功能区. 其第一个残基为焦谷氨酸, 因而其 N 端封闭, 由连续五个 I 型结构 (I₁—I₅) 组成. 通过因子 XIII_a 与其 N 端第三个保守的残基 Glu 连结, 再交联纤维蛋白. 它也含有一个肝素低亲和性结合点.

Lambert 等在纯化该片段过程中发现,在 Ca 离子存在条件下,27kD 片段中残基 Lys(85)-Asp(86) 之间的肽键裂解,显示有丝氨酸蛋白酶活性,具有钙依赖性^[5]. 紧随 N 端功能区的是一个 46kD 的明胶结合区 (gelatin binding site, GBD), 含大量二硫键,由四个 I 型结构和两个 II 型结构组成: I₆-II₁-II₂-I₇-I₈-I₉. 该片段通过亲水键如氢键,范德华力与明胶结合. Baralle 在 *E. coli* 中表达特定的融合蛋白来进一步定位明胶的结合位点,其实验证明明胶结合活性是处在连接结构 II₂ 与邻近 I₇ 之间的 14 个氨基酸残基上^[6]. Litvinovich 研究 GBD 区发现,此区中六个结构域都是独立折叠,其中 I₁, I₂ 和 I₈ 对温度相对敏感, I₆, I₇, I₉ 对温度不敏感,反映出 GBD 区的变构是分步完成^[7]. 其它许多分子的结合位点也在 GBD 区,如组织转氨酶也能与 GBD 区结合^[8]. 最近, Chernousov 用单抗证明此区中的 I₉ 与其后面的 II₁ 构成一功能单位,能介导前体 Fn 分子间相互作用,参与 Fn 聚合过程^[9]. GBD 区含多个糖化位点,能起到稳定 GBD 区抵抗蛋白酶解的作用. 中央区域由 GBD 后面六个 II 型结构组成,前三个 II 型结构对蛋白酶极敏感,难以获得其稳定片段,故功能不详. 后三个 II 型结构约 30kD,是 DNA 和另一个低亲合性肝素结合区,因亲和力低,其生理状态下的意义不明. 中央区域在不同种属动物间同源性高达 90%,此区可能还有某些重要的结合点,有待深入研究. 另一个重要的功能区是细胞结合区 (cell binding site, CBD), 它能促进细胞粘附与聚集,由 II₈-II₉-II₁₀ 组成. Pierschlacher 从该区上切下一个 12kD 小肽,它保留了亲和细胞的能力,用抗 II₁₀ 单抗能有效阻断其活性,证明 12kD 片段是包含 II₁₀ 序列的区域^[10]. 此后根据 II₁₀ 的 C 端残基序列合成的一四肽 Arg-Gly-Asp-Ser (简称 RGD 肽) 成功地模拟了 12kD 片段结合细胞的能力. 单独的 RGD 肽可为多种细胞所识别,但其邻近序列的参入可增强结合活力. 除去 RGD 肽 C 端的序列明显地降低了它的活力^[11]. RGD 肽是存在于多种蛋白分子中共同

的细胞亲合点. 除 Fn 外,体外粘连蛋白 (vitronectin), vWF, 人纤维蛋白原的 α 链,血小板糖蛋白 GP II b/III a 等均含 RGD 肽. 从结构上看,每个 II 型结构在 C 端第三个残基上都形成一个亲水环,暴露在 β 结构延伸区的表面,在 II₁₀ 中即为 RGD 肽. 由于 II 型结构高度保守性,这些序列中可能还有某些特定功能. 一个肝素高亲合性结合区位于 CBD 的后面,由 II₁₂-II₁₃-II₁₄-V 组成. 对肝素亲合力远高于 N 端功能区 and 中央区域. Barkalow 研究证明,肝素活性区定位在 II₁₃ 和 II₁₄ 上,且发现主要的活性由 II₁₃ N 端的一半残基决定^[12]. 该区也能粘附细胞,且不为 RGD 肽抑制. Konoriya 最近的研究发现,在 V 区内含有另一个细胞结合点,活性存在于三个残基上: Leu-Asp-Val, 称 LDV 肽^[13], 人、大鼠、牛的 V 区上的此三肽是保守的. Fn 的 C 端功能区是一纤维蛋白结合区,由 I₁₀-I₁₁-I₁₂ 构成,不包括 C 端链间二硫键,此区结合纤维蛋白不需凝血 XIII_a 因子介导,也不能结合肝素,说明它与 N 端纤维蛋白的结合点不同.

2 纤维粘连蛋白基因的结构特点

Fn 基因片段的重组克隆,所测得序列能很好的交叉重叠,说明体内仅有单一的 Fn 基因,所有来源于 cDNA 的不同 mRNA 均起源于同一 Fn 基因. cDNA 探针定位人 Fn 基因位于二号染色体上.

2.1 内含子-外显子结构

Fn 基因有明显的内含子-外显子结构. Patel 证实约有 50 个外显子^[14]. 除少数例外,基因上的外显子与蛋白肽链中的重复序列对应. 研究大鼠 Fn 基因发现,12 个 I 型结构中的 8 个; 2 个 II 型结构中的 1 个 (即 II₂) 仅由一个外显子编码. 大部分 II 型结构由两个外显子编码. 三个额外的可变区分别由一个外显子编码,人 Fn 基因结构与大鼠相似. Fn 多肽中三种类型的序列重复出现的特点表明其编码基因起源于三个原始的遗传单位经内部复制而成. 每个重复序列都是由具体的遗传单位 (一个或两个

外显子) 编码. 这些重复序列在其它许多蛋白质分子中也出现, 这可能是生物进化过程中由外显子改组 (exon shuffle) 所造成.

2.2 选择剪切

F_n 基因是一较有特殊性的基因, 通过改变 RNA 转录过程, 单一的一条基因能产生序列稍异的几条多肽, 人体约有 20 种形式. F_n 多肽的多型性体现在三个可变区氨基酸残基的差异上, 这种差异来自前体 mRNA 水平上的选择剪切 (alternative splicing), 不同的剪切方式产生不同类型的 F_n, 后者分别与特定功能相关. Ffrench 等人用剪切区特异探针作原位杂交, 测定了鸡胚生长过程中不同 mRNA 剪切的分布及浓度, 结果发现在胚胎不同位置及不同生长时间里 mRNA 的剪切方式也不同^[15]. Norton 等也发现早期胚胎中 mRNA 大部分是编码 EIIIA⁺ 和 EIIIB⁺. 抗肽链 EIIIA 抗体与胚胎组织的亲合力比成熟组织高, 说明 EIIIA, EIIIB 在早期胚胎中参与细胞增殖与移动. Pagani 等发现, 所有幼鼠 V 区的表达均较成年组高, 成年鼠脑组织中 EIIIA 和 EIIIB 的表达大大高于老年鼠, 证明 EIIIA, EIIIB 在脑组织分化中可能有重要作用^[16]. 此外, 在大鼠损伤的皮肤局部也发现 EIIIA, EIIIB 的重新表达. 肿瘤也可能与 F_n 选择剪切有关, Oyama 等发现人肝恶性肿瘤组织中 mRNA 上 EIIIA 阳性比正常人明显升高, 而在非肿瘤组织中, 包括肝炎等, EIIIA 的选择剪切无明显差异, EIIIA 区剪切的增加与部分静脉血栓和肿瘤的肝内转移也存在密切相关^[17]. 这些结果均表明, F_n 前体 mRNA 组织特异性的选择剪切在人肝肿瘤细胞中被修饰, 这可能是控制 RNA 选择剪切的分子机制. 许多生长因子如 β 转染生长因子, 维甲酸, 1-25 二羟基 D3 都可能改变前体 mRNA 在三个可变区内的选择剪切.

人 F_n 基因 5' 端含一转录调节序列, 包括引物 TATAA 和 CCAAT 及 SP1 结合点 (GGGCGGG). 另外还有一些调节片段类似于

其它几种蛋白的调节序列, 如含有热休克蛋白, cAMP 反应蛋白的调节片段, 反应出 F_n 生物合成调节上的复杂性. 3' 非翻译区有 700 碱基长, 在 Poly-A 上游 150 个碱基处有一八核苷酸序列: TTATTTAT, 由于许多参与炎症反应蛋白的基因中在其相应位置上也有这一序列^[18], 推测这一序列可能与 F_n 在炎症反应中的作用有关.

参 考 文 献

- 1 Hynes R O. *Annu Rev Cell*, 1985; **1**: 67—90
- 2 Beckmann B, Geiger M, Carlie de Vries *et al.* *J Biol Chem*, 1990; **266** (4): 2227—2232
- 3 Kimizuka K, Ohdate Y, Kaqase Y *et al.* *J Biol Chem*, 1991; **266** (5): 3040—3051
- 4 Richard O H. In: *Fibronectins*. New York: Springer-Verlag Inc, 1990: 126
- 5 Lambert V S, Lottspeich F, Emid I *et al.* *Eur J Biochem*, 1991; **20** (1): 71—77
- 6 Owens F J, Barolle F E. *EMBO J*, 1986; **5**: 2825—2830
- 7 Litvinovich S V, Strickland D K, Medved L V *et al.* *J Mol Biol*, 1991; **217**: 563—575
- 8 Turner P M, Lorand L. *Biochem*, 1989; **28**: 628—635
- 9 Chernousov M A, Fogerty F J, Koteliansky V E *et al.* *J Biol Chem*, 1991; **266** (17): 10851—10858
- 10 Pierschlacher M D, Ruoslahti E, Sundelin J *et al.* *J Biol Chem*, 1982; **257**: 9593—9597
- 11 Aota S, Nagai F, Yamada M *et al.* *J Biol Chem*, 1991; **266** (24): 15938—15943
- 12 Barkalow F J, Schwarzbauer J E. *J Biol Chem*, 1991; **266** (12): 7812—7818
- 13 Komoriya A, Green K J, Mervic M *et al.* *J Biol Chem*, 1991; **266** (23): 15075—15079
- 14 Patel V P, Lodish H V. *J Cell Biol*, 1987; **105**: 3105—3118
- 15 Ffrench C C, Hynes F O. *Development*, 1989; **106** (2): 375—388
- 16 Pagani F, Zagato L, Vergani C *et al.* *J Cell Biol*, 1991; **113** (5): 1223—1229
- 17 Oyama F, Hirohashi S, Shimosato Y *et al.* *J Biol Chem*, 1989; **264** (18): 10331—10334
- 18 Capri D, Beutler B, Hartog K *et al.* *Proc Natl Acad Sci USA*, 1986; **83**: 1670—1674

Enzyme Catalysis in Low-Water Organic Media. Yang Zhen, D. A. ROBB, Ji Liangnian. (*Center for Biotechnology, Department of Chemical Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA*). *Prog. Biochem. Biophys.* (China), 1994; 21 (2): 98

Enzymes are catalytically active not only in aqueous solution but also in organic media (low-water solvent system, reversed micelles, monophasic cosolvent system, and biphasic organic/aqueous system). Of special interest is the low-water solvent system, because it is important to organic synthesis. In this review, attention is focused on the factors that influence enzyme catalysis in a low-water solvent including the role of water, selection of the solvent and support, some special properties acquired by enzymes in such a system are discussed. Examples of applications with the use of enzymes in organic synthesis, analysis, and polymer chemistry, are listed.

Key words enzymatic catalysis, enzyme activity, organic media, low-water solvent system

The Superfamily of Plant Lectins. Sun Jianzhong, Wang Keyi. (*Shanghai Institute of Biochemistry Academia Sinica, Shanghai 200031*). *Prog. Biochem. Biophys.* (China), 1994; 21 (2): 104

Lectins are a kind of carbohydrate-binding proteins. Though they differ in their carbohydrate specificities, they resemble each other in many physicochemical properties. By now, a lot of plant lectins have been sequenced, and some of their three-dimensional structures have been established. A few lectin genes have been cloned. Comparison of their primary sequences and tertiary structures, we can find that plant lectins are several large groups of homologous

proteins belonging to a superfamily.

Key words plant lectins, homology comparison, protein superfamily

Progress in the Molecular Biological Research on Fibronectin. Wang Ling, Wang Zhenyi, Qi Zhengwu. (*Shanghai Institute of Hematology, Shanghai Second Medical University, Shanghai 200025*). *Prog. Biochem. Biophys.* (China), 1994; 21 (2): 109

Fibronectin (Fn) is a high molecular dimer of glycoprotein with a series of discrete structural domains. It is composed of three type repeats, type I, II and III. A serial repeats form a functional domain. There is only one Fn gene in body, and by alternative splicing it produces many kinds of Fn polypeptides, which have different sequences in three variable regions. Fn is involved in a variety of biological functions. It is very important to elucidate the relationship between the function and structure relationship by deeply analysing the structures of its domains and gene.

Key words fibronectin, repeats, alternative splicing

Progress in Topography of Ribosomal RNA and RNA N-Glycosidase Research (I). Zhang Jinsong, Liu Wangyi. (*Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031*). *Prog. Biochem. Biophys.* (China), 1994; 21 (2): 113

The α -sarcin domain of 28S rRNA is involved in protein synthesis reaction catalyzed by ribosomes. It was demonstrated that trichosanthin is an RNA N-glycosidase, and a new method for RNA N-glycosidase assay was preliminarily established. Trichosanthin cleaves the supercoiled double-stranded DNA to produce nicked