

# 基于结构的丝氨酸蛋白酶超家族进化分析

季星来, 孙之荣

(清华大学生物信息研究所, 北京 100084)

**摘 要:** 丝氨酸蛋白酶是一类水解酶, 得名于 Ser195 对底物的亲核攻击。丝氨酸蛋白酶在长期进化过程中, 虽然蛋白质序列之间的差异性很大, 但仍然保持了一个稳定的核心催化结构; 同时, 丝氨酸蛋白酶也衍生出各种各样的生物学功能。但是, 序列的差异以及大量的复制现象使得丝氨酸蛋白酶超家族的进化研究进展缓慢。本文用 SSAP 算法从结构比较的层次上进行蛋白质超家族进化的研究, 分析了丝氨酸蛋白酶超家族结构与功能的演化, 研究了丝氨酸蛋白酶结构上的进化及其对功能的影响, 对于蛋白质结构与功能关系的研究以及蛋白质分子设计都有一定的指导意义。

**关键词:** 丝氨酸蛋白酶超家族; 分子进化分析; 结构比较; SSAP 算法

中图分类号: Q617 文献标识码: A 文章编号: 0372-2112 (2001) 12A-1756-03

## Structure Based Evolution Analysis of the Serine Protease Superfamily

Ji Xing-lai, SUN Zhi-rong

(Department of Biological Sciences & Biotechnology, Bioinformatics Institute, Tsinghua University, Beijing 100084, China)

**Abstract:** The serine protease superfamily is a subgroup of the hydrolases, whose name comes from the conserved Ser-195 catalyzing the substrates. Serine proteases maintain a common fold over a very long evolutionary span, in spite of divergent sequences. This class of enzymes has developed a wide variety of important biological functions. However, strongly divergent protein sequences and a mass of gene duplications make it very difficult to perform evolution analysis on the whole superfamily. In the present work, structure based evolution analysis was performed on the serine proteases superfamily using SSAP algorithm, and the structure and function evolution were analyzed, especially the effect on functions by structure evolution of the serine proteases. The research would have certain guiding significance on the relationships between protein structures and functions and the molecular design.

**Key words:** serine protease superfamily; molecular evolution analysis; structure alignment; SSAP

## 1 引言

丝氨酸蛋白酶属于蛋白水解酶, 有一个独特易反应的活性丝氨酸残基。丝氨酸蛋白酶超家族的成员多而广, 物种跨度大, 功能明确。丝氨酸蛋白酶在结构上为全  $\beta$  蛋白, 核心结构由两个非常相似的结构域构成, 每个结构域是一个由 6 个  $\beta$  折叠片构成的  $\beta$  折叠桶, 而丝氨酸蛋白酶的活性催化位点以及其它的重要功能位点就位于这两个结构域之间(图 1)。两个结构域由于结构上的差异, 它们在丝氨酸功能和进化上的作用也是不同的。

在长期的进化过程中, 丝氨酸蛋白酶的序列产生了很大的差异, 在超家族成员之中, 某些蛋白质序列的相似性甚至低于 20%, 为整个超家族的进化研究带来了一定的困难。但是从结构上看, 由于丝氨酸蛋白酶的结构与功能关系密切, 蛋白酶的结构保持了极大的稳定性<sup>[1,2]</sup>, 这使得从结构上进行超家族的进化分析成为可能。实际上, 正是由于丝氨酸蛋白酶结构上的微小变化, 引起了功能上的进化, 而丝氨酸蛋白酶也因

此具有诸多的生理功能, 几乎涉及到生物体发育和代谢过程的方方面面。

本文从结构的角度进行丝氨酸蛋白酶超家族的进化研究, 探讨了丝氨酸蛋白酶结构与功能的进化, 以及两个子结构域对于功能与进化的不同作用, 对于蛋白质功能的分化及其机制有重要的意义。

## 2 材料与方法

### 2.1 材料

选取丝氨酸蛋白酶超家族作为研究对象。从 SCOP 分类数据库中<sup>[3]</sup>选取 Trypsin-like 丝氨酸蛋白酶超家族和 Herpes virus 丝氨酸蛋白酶超家族成员。在每个谱系的各个记录中, 只选择一条催化结构域记录, 挑选的原则是尽量选择天然蛋白的晶体结构, 否则从突变结构或者复合结构中选取最接近天然结构的晶体记录。这样, 一共选择了 60 个丝氨酸蛋白酶催化域晶体结构; 其中有两条记录是同一个蛋白质的不同结构域, 具有不同的结构与功能。

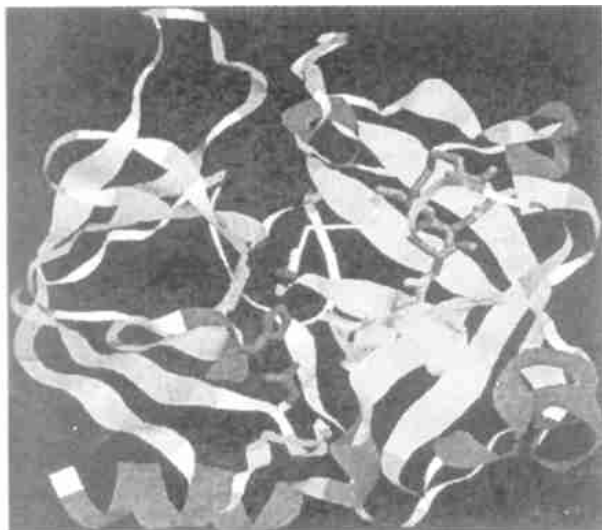


图1 牛糜蛋白酶结构(PDB: 2ga).  $\alpha$ 螺旋标记为品红色;  $\beta$ 片层标记为黄色; 拐角标记为淡蓝色; 其它标记为白色. 催化中心标记为红色; 氧离子孔标记为白色; 底物口袋标记为绿色; 主链结合位点标记为黄色; 其它底物相关位点标记为黄绿色.

## 2.2 方法

使用 SSAP(Sequential Structure Alignment Program) 程序进行丝氨酸蛋白酶结构比较. SSAP 算法<sup>[4]</sup>是一种新的蛋白质结构比较算法; 该算法执行两级动态寻优过程. 由于 SSAP 算法采用在蛋白质序列比较中广泛采用的动态规划技术, 因此它使蛋白质结构与序列比较统一化. 系统进化树由 Philip 软件包<sup>[5]</sup>中的 NEIGHBOR 程序构建, 使用 Neighbor-joining 算法<sup>[6]</sup>.

## 3 结果与讨论

### 3.1 系统发育分析

系统发育进化树(图2)表明丝氨酸蛋白酶超家族在整体上是趋异进化. 首先, 系统发育进化树在总体上根据物种类型而聚类, 明显地划分为病毒、原核生物和真核生物三大类, 这说明丝氨酸蛋白酶的超源非常早, 应该在原核生物和真核生物物种分歧之前. 另外, 丝氨酸蛋白酶基本上按照功能而聚类, 这一方面体现了丝氨酸蛋白酶在物种或者功能进化史上的重要地位, 也体现了结构与功能的紧密联系. *Streptomyces griseus* trypsin (1sgt00, SGT) 是来源于原核生物的蛋白酶, 但是同真核生物的蛋白酶首先聚类, 这说明胰蛋白酶是很早就起源的丝氨酸蛋白酶, 而它所具有的消化功能应该就是丝氨酸蛋白酶最早的功能之一, 这同丝氨酸蛋白酶最基本的降解功能是一致的. 值得注意的是, 丝氨酸蛋白酶许多特异性的功能都只在真核生物内才具有, 这一方面说明真核生物的功能复杂度, 也表明这些功能是在基本功能的基础上较晚进化的; 功能的特异化是蛋白质功能进化的一个大的趋势.

必须指出, 由于丝氨酸蛋白酶超家族在进化过程中呈现出大量的复制现象, 横向进化占有较大的比重, 因此不太适合用来进行纵向进化的研究, 对于物种进化和谱系进化的研究有很大的误导性. 当然, 也可以选择超家族中部分成员进行物种进化的研究, 比如选择哺乳动物的胰蛋白酶、凝血因子和弹

性蛋白酶等进行纵向进化的研究<sup>[7]</sup>.

### 3.2 结构域分析

鉴于丝氨酸蛋白酶两个结构域的相似性, 我们对结构域进行了结构比较并对对比结果进行了统计分析, 表1列出了这些统计结果. 从统计结果上看, 同  $N$  端结构域相比,  $C$  端结构域有很好的相似性, 差异性也小, 这说明  $C$  端结构域在进化上承受较大的环境压力, 因此必须保持一定的三维结构, 以执行相应的催化功能.  $C$  端结构域在结构与功能上的重要性在系统发育进化树上也得到了很好的体现: IfxyA0 是一种人工嵌合物,  $N$  端结构域来自于凝血因子 X,  $C$  端结构域来自于糜蛋白酶, 而在系统发育树中 IfxyA0 还是同糜蛋白酶聚类在一起, 这说明  $C$  端结构域在整个丝氨酸蛋白酶催化结构域中占据核心地位.

$C$  端结构域在丝氨酸蛋白酶催化结构域中发挥重要的功

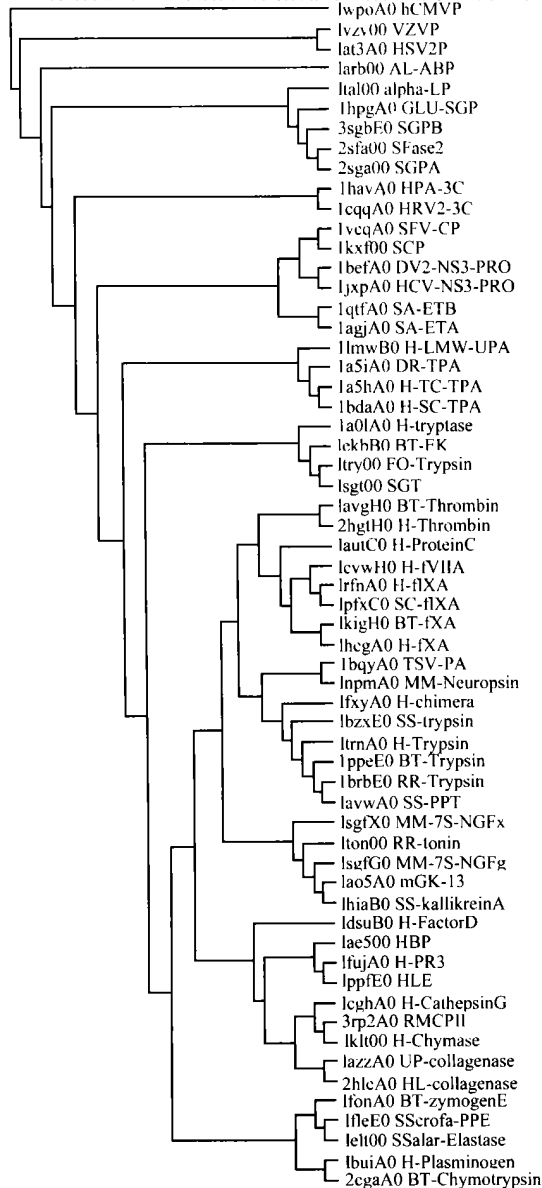


图2 丝氨酸超家族基于结构比较的系统发育进化树

能作用,这个重要性有其结构基础.从丝氨酸蛋白酶的三维结构上看(图1),虽然催化中心的三个主要残基中有两个(Asp102&His57)位于N端结构域,但是最重要的Ser195在C端结构域上;但是其它重要的功能位点都位于C端结构域,其中包括稳定催化中间产物的氧离子孔(Gly193&Ser195),底物特异性口袋(Ser189, Gly216&Gly226),主链底物结合位点(Ser214, Trp215&Gly216)以及其它一些底物相关位点.另外,丝氨酸蛋白酶催化结构域含有三个十分重要的保守二硫键,其中有两个位于C端结构域.正是由于C端结构域上包含绝大部分的功能位点,C端结构域因此而执行丝氨酸蛋白酶的主要催化功能,也因此承受较大的进化压力;N端结构域的主要功能可能就是同C端结构域配合保证催化中心三联体的结构稳定和功能活性.

N端结构域和C端结构域由于执行不同的功能,而且承受不同的进化压力,它们的进化模式也并不相同.C端必须保持一定的三维结构以执行相应功能,而N端所承受的进化压力要小得多,表现在序列和结构上,N端结构域呈现出相当大的差异性,很可能是中性进化的结果.另外,从统计结果上看,原核生物和真核生物在N端结构域上的进化模式相当稳定,而C端结构域进化模式变化相当大,而且真核生物蛋白酶的两个结构域的保守程度要比原核生物大得多.从结构上看,重要的功能位点都位于两个结构域之间,而且真核生物蛋白酶拥有另外三个非保守的二硫键.这些证据都表明,丝氨酸蛋白酶催化结构域很可能是由这两个结构域融合形成的.

表1 子结构域结构比较结果统计分析

谱系	平均相似值 (N-terminal)	平均相似值 (G-terminal)	方差 (N-terminal)	方差 (G-terminal)
原核生物	79.76417	85.34444	74.00044	32.27627
真核生物	89.16942	89.83897	66.43925	4.904512
所有	81.25177	82.70374	143.3776	157.9728

### 3.3 功能进化

丝氨酸蛋白酶超家族成员不仅物种跨度大,而且执行的生理功能也是多种多样的,包括细菌降解、食物消化、血凝过程、发育过程、神经系统、免疫过程和细胞衰老等.从功能进化的角度看,丝氨酸蛋白酶的功能越来越特异性,而这个功能上的特异性来自于催化底物的特异性.原核生物的蛋白酶在执行降解功能时对底物特异性要求不高,而真核生物的蛋白酶对底物要求比较高,不仅针对某一类型的肽键,而且还对底物本身有特异性的要求.对底物特异性的要求在结构上表现为底物特异性口袋的结构变化(图3),正是由于底物特异性口袋结构和其它底物结合位点的结构变化使得丝氨酸蛋白酶具有不同的底物特异性,从而造成了蛋白酶的功能差异,这也是

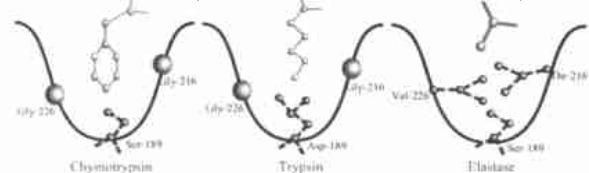


图3 糜蛋白酶、胰蛋白酶和弹性蛋白酶底物口袋结构示意图

功能进化的一种模式.

丝氨酸蛋白酶的功能进化除了表现为底物特异性的变化外,还表现为功能执行的复杂性.原核生物的蛋白酶主要执行降解功能或者食物消化功能,这些功能大都是直接完成的;但是真核生物的蛋白酶在执行各种生理功能时,每种功能的执行都具有一定的网络调控性质,生理功能并不是直接完成的,而是通过一系列的酶或蛋白质激活完成,在这个激活过程中,有些丝氨酸蛋白酶成为上游激活因子,而另一些则成为下游执行因子,使得丝氨酸蛋白酶在执行生理功能时的角色也出现了分化,凝血过程中的丝氨酸蛋白酶就是一个典型的例子.另外,多细胞生物的进化导致了细胞的分化和各种器官组织的分化,在这些分化过程中,产生了一了各种各样的生理过程,而丝氨酸蛋白酶也为适应这些新的功能而产生分化,这说明蛋白质功能的进化除了要适应物种进化的外在环境外,还必须适应物理内在分化所产生的体内环境.丝氨酸蛋白酶在适应这些分化过程中,主要依靠同其它已有的功能结构域融合形成新的蛋白,从而发展出新的蛋白功能,如神经生长因子就是由丝氨酸蛋白酶催化结构域同其它结构域融合而成.结构域融合也是蛋白质功能进化的一种模式.

致谢 感谢 Prof. W. R. Taylor & Prof. C. A. Orengo 提供 SS-AP 软件.

### 参考文献:

- [1] Blow DM. Structure and Mechanism of Chymotrypsin [J]. Acc. Chem. Res., 1976, 9: 145- 152.
- [2] Barker WC, Dayhoff MO. Evolutionary and functional relationships of homologous physiological mechanisms [J]. BioScience, 1980, 30: 593 - 600.
- [3] Murzin AG, Bernner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures [DB]. J. Mol. Biol., 1995, 247: 536- 540.
- [4] Taylor WR, Orengo CA. Protein Structure Alignment [J]. J. Mol. Biol., 1989, 208: 1- 22.
- [5] Felsenstein J. PHYLIP (phylogeny inference package) version 3. 5c. Distributed by the author, Department of Genetics, University of Washington, Seattle [CP]. 1993.
- [6] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. Mol. Biol. Evol., 1987, 4(4): 406-425.
- [7] Sun ZR, Ji XL, Hu SM, Wang Y, He FC. A new method of research on molecular evolution of proteinase superfamily [J]. Chinese Science Bulletin, 2001, 46(7): 578.

### 作者简介:



季星来 男, 1979 年 4 月出生于江苏南京. 1999 年清华大学生物科学与技术专业毕业, 获理学学士学位. 现攻读生物信息学博士学位, 发表论文 3 篇.