

人类抗体重排机制的起源

严信宇¹, 黄盛丰¹, 徐安龙^{1,2}

1. 中山大学生命科学学院, 广州 510275

2. 北京中医药大学生命科学院, 北京 100029

摘要 重组激活基因(RAG)蛋白介导的抗体重排是脊椎动物适应性免疫系统的核心, 抗体重排机制的起源一直是免疫学研究的热点。本研究以有活化石之称的文昌鱼为对象, 对多个文昌鱼基因组草图进行深度信息学分析, 发现了一个全新DNA转座子*ProtoRAG*。进一步功能研究表明, *ProtoRAG*编码的蛋白能够介导自身的转座和宿主DNA的重排, 其作用机制与人类抗体蛋白介导的抗体重排机制基本一致。因此, *ProtoRAG*就是研究人员长期搜寻的祖先RAG转座子的“分子活化石”, 该发现为诺贝尔生理学或医学奖获得者利根川进提出的“抗体重排机制转座子起源”假说提供了最有力和最直接的证据。

关键词 抗体重排; RAG; 文昌鱼; *ProtoRAG*转座子

日常生活中, 细菌、病毒和寄生虫等各种病原体威胁着身体的健康, 加上时刻发生着变异, 病原体具有非常高的多样性, 这对人体免疫系统提出了严峻的考验。相应地, 为了抵御多种多样的病原体入侵, 人体通过适应性免疫系统产生了数目庞大的高度多样化的特异性抗体或T细胞受体, 分别行使着人体的体液免疫或细胞免疫的功能。适应性免疫系统所能够产生的抗体(或T细胞受体)具有近乎无限的体细胞多样性(理论上可以达到 10^{18} 种), 这使得人体能够应对许多潜在的未知病原体的威胁。但是基因组所能承载的遗传信息是有限的, 如何通过有限的遗传信息产生近乎无限的多样性抗体, 抗体重排机制为我们提供了解决方案, 如同从不同色系的拼图块中各选出一块再拼成一个整体(图1), 抗体重组激活基因(RAG1和RAG2)通过将不同的抗体基因座片段重新排列组合形成具有不同特异性



3种色系的拼图块(蓝、黄、红)分别代表抗体基因座上3种不同的基因片段(V、D、J), 拼接的三色拼图代表了一个特异的抗体基因

图1 抗体基因的V(D)J重排类似拼图机制

Fig. 1 Jigsaw puzzle represents the mechanism of V(D)J recombination

收稿日期: 2017-10-12; 修回日期: 2018-03-19

基金项目: 国家自然科学基金项目(31722052, 91231206, 31171193); 国家重点基础研究发展计划(973计划)项目(2013CB835305)

作者简介: 严信宇, 副研究员, 研究方向为分子免疫学, 电子信箱: yanxy9@mail.sysu.edu.cn; 黄盛丰(共同第一作者), 教授, 研究方向为生物信息学和比较免疫学, 电子信箱: hshengf2@mail.sysu.edu.cn; 徐安龙(通信作者), 教授, 研究方向为分子免疫学, 电子信箱: lssxal@mail.sysu.edu.cn

引用格式: 严信宇, 黄盛丰, 徐安龙. 人类抗体重排机制的起源[J]. 科技导报, 2018, 36(7): 43-47; doi: 10.3981/j.issn.1000-7857.2018.07.007

的抗体蛋白。因此抗体重排机制是适应性免疫系统的核心机制,它的出现是脊椎动物适应性免疫起源的核心事件。在抗体重排过程中,重组激活基因(RAG)扮演着至关重要的角色,它介导了体细胞基因组中不同基因片段间DNA的切割与重新组合。因此,寻找RAG同源蛋白并研究RAG蛋白功能演化将可能为我们揭示抗体重排机制的起源。

1 RAG 蛋白的研究历史

1.1 脊椎动物 RAG 蛋白的发现

20世纪70年代,研究人员发现抗体基因座是由一系列独立的V、D和J基因片段组成,并不能正确编码出有功能的蛋白,这些片段需要经历V(D)J基因片段间的重排/重组才能形成正确编码功能蛋白的抗体基因。随后,进一步研究发现在V(D)J基因片段的边缘存在着DNA重复元件,这些重复元件是间隔着12碱基或者23碱基的保守七聚体和九聚体序列(重组信号序列,RSS)。1979年,利根川进(1987年诺贝尔生理学或医学奖获得者)等注意到这些重复元件与转座子末端的反向重复序列(TIR)十分相似,他们认为,这些转座子元件偶然插入到祖先物种抗体基因中可能促使了重排机制的诞生,并因此提出了重排机制的转座子起源假说^[1],但是当时并不清楚具体是什么蛋白催化了抗体基因的重排。直到1989年,Sakano等通过对提取自发育中的B细胞和T细胞的基因组DNA以及cDNA的重排筛查找到了一个重组激活基因——*RAG1*^[2]。随后对*RAG1*的研究发现,*RAG1*蛋白所介导的重排反应效率很低,但是在向细胞中转入其基因组DNA后却能够极大提升重排反应效率,这也使得另一个重组激活基因*RAG2*得以被发现^[3]。*RAG1*和*RAG2*基因以“尾对尾”的方式在基因组上相邻而居,并且它们的编码区不含有内含子,这种特性让我们联想到原核生物转座子编码的转座酶基因。随后的研究也证实,*RAG*蛋白能够在体外转座具有末端RSS序列的线性DNA^[4],展示了抗体基因的原始插入模型,同时也提示我们祖先*RAG*分子有可能是一个从转座子水平转移而来的转座酶,当然也有可能是来自细菌的重组酶或者病毒重组酶等。

1.2 低等后口动物的 RAG 同源序列

很长一段时间内,人们仅在有颌类脊椎动物中发现了*RAG*蛋白,而在低等后口动物和原口动物中没有

发现*RAG*同源蛋白的踪迹,因此一般认为*RAG*是从软骨鱼类开始才具有的基因。近年来,通过对基因组测序数据的分析,研究人员在低等后口动物中也发现了一些与有颌类脊椎动物*RAG*蛋白有同源性的序列元件。例如,2005年Kapitonov等发现*Transib*转座酶与*RAG1*蛋白中催化活性相关的核心氨基酸序列十分相似,同时在海胆、水螅、果蝇、非洲疟蚊等物种基因组中都发现有*Transib*转座子的存在,推测*RAG1*蛋白的核心序列可能是由*Transib*转座酶演化而来^[5]。但是,或许*Transib*编码的转座酶和*RAG1*有同源关系,但它们仍然相距甚远。2006年,Fugmann等在棘皮动物紫海胆中发现了2个紧邻的*RAG*样基因*SpRAG1L*和*SpRAG2L*,它们在序列和基因组结构上与脊椎动物的*RAG1*和*RAG2*基因都具有相似性^[6],但这个基因座缺乏其他典型的转座子元件,看起来更像是宿主基因。考虑到在棘皮动物中没有发现抗体的存在,因此尚不清楚这些*RAG*样蛋白承担的宿主重排功能是什么,后经研究推断,*RAG1*和*RAG2*可能具有不同的起源,*RAG2*是2个宿主蛋白结构域通过偶然的结构域重排产生的。无论如何,这一研究表明在现存后口动物共同的祖先中可能已经存在紧密相连的*RAG1/2*样基因;它们在当时可能执行着其他功能,直到有颌类脊椎动物演化初期才演化出介导抗体重排反应的功能,这一发现将*RAG*样蛋白的起源推前到6亿年前。

1.3 文昌鱼 RAG 同源分子的寻找

海胆之后,研究人员把寻找*RAG*同源蛋白的目光投向了文昌鱼。文昌鱼是现存与脊椎动物亲缘关系最近的物种之一,经历5亿多年的演化依然保持着许多古老的特性和原始的性状,是研究脊椎动物免疫系统起源与演化的理想模型。2005年,徐安龙研究团队的董美玲等在文昌鱼中发现了哺乳动物*RAG1*激活基因(*RAG1 gene activator, RGA*)的同源蛋白,该分子能够调控*RAG*表达^[7],这表明在低等脊索动物文昌鱼中表达*RAG*蛋白的信号通路可能已经存在,同时,尽管完整的V(D)J重排机制还没有形成,但是一些相关的分子已经在等待着被招募。2008年,徐安龙研究团队的黄盛丰等在文昌鱼基因组中发现了*RAG1*样蛋白片段^[8],提示在脊索动物祖先中就存在*RAG*蛋白。2013年,刘小龙研究团队的张燕妮等对文昌鱼*RAG1*样蛋白片段进行了深入的功能研究,发现它在序列及功能上都与脊椎动物*RAG1*蛋白的核心结构域相似,用它替换小鼠

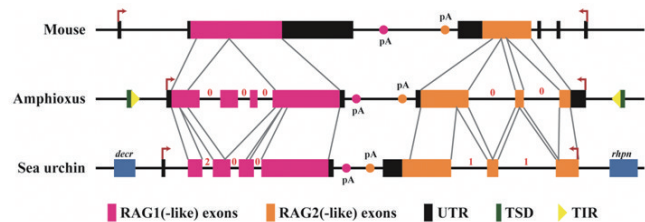
RAG1 核心序列后,依然能与小鼠 RAG2 一起识别重组信号并介导 V(D)J 重排反应^[9],表明文昌鱼这一 RAG1 样蛋白片段已经具备进一步演化形成介导重排反应功能的潜力。由于这一片段具有核酸酶活性,研究认为它可能来自逆转录病毒。2014 年,徐安龙研究团队的黄盛丰等对多个文昌鱼基因组草图进行了比较基因组学分析,通过对不同文昌鱼单倍体基因组序列、EST 序列及 BAC 文库序列的精确比对,发现了大量由转座子插入造成的长片段序列的插入和缺失(indel),其中包含不少完整的转座子序列,并且它们中多数依然具有转座活性。这些转座子属于不同的超家族,其中多数是在哺乳动物中所没有发现的,说明现存脊椎动物可能在很大程度上丢失了转座子多样性,而文昌鱼则保留了这些古老的有活性的转座子。在这其中发现了一个能够编码 RAG1/2 样蛋白的转座子,它的末端包含有经典的靶位点重复序列(TSD)和末端反向重复序列(TIR),是一个有活性的“剪切—黏贴”DNA 转座子,因此这个转座子很有可能就是我们一直寻找的古老的 RAG 转座子,被定名为 *ProtoRAG*^[10]。

2 文昌鱼 *ProtoRAG* 转座子

2.1 *ProtoRAG* 转座子的序列特征

2016 年,徐安龙研究团队的黄盛丰等对文昌鱼 *ProtoRAG* 的进一步序列分析显示,3 种文昌鱼(白氏文昌鱼、日本文昌鱼和佛罗里达文昌鱼)均具有低拷贝数的 *ProtoRAG* 基因,完整的 *ProtoRAG* 基因总长度为 7639 bp^[11],在其两侧是 5 bp 的 TSD 序列以及一对与 V(D)J 重排反应 RSS 序列相类似的 TIR 序列,在 TIR 序列两端同样有类似的保守七聚体和九聚体序列,而在 TIR 序列之间则是 RAG 样基因 *BbRAG1L* 和 *BbRAG2L*,它们在基因组上同样以“尾对尾”的方式相邻而居(图 2)。与典型的转座子基因不同的是,*BbRAG1L/2L* 基因具有多个内含子,同时这些内含子的相位以及所处位点与脊椎动物 RAG 基因和海胆 *SpRAG1L/2L* 基因都不相同(图 2)。而哺乳动物 *RAG1/2* 基因在其编码区则不含有内含子,因此,可能哺乳动物 *RAG1/2* 基因在演化过程中丢失了内含子,或者在祖先 *RAG1/2* 基因编码区中就不含有内含子。文昌鱼 *BbRAG1L* 蛋白序列全长 1136 个氨基酸,与脊椎动物 RAG1 和海胆 *SpRAG1L* 相似度分别为 29% 和 36%,与此相比,海胆 *SpRAG1L* 和 Transib 转座酶与

RAG1 的相似度分别为 26% 和 16%~18%。同时 *BbRAG1L* 与 RAG1 在序列上具有许多保守的区块,尤其是一些活性相关的氨基酸位点,因此推测 *BbRAG1L* 可能具有与 RAG1 相似的活性。文昌鱼 *BbRAG2L* 蛋白序列全长 366 个氨基酸,与脊椎动物 RAG2 和海胆 *SpRAG2L* 相似度分别只有 20% 和 24%。*BbRAG2L* 具有在 RAG2 和 *SpRAG2L* 中保守的 N 端结构域(six-bladed β -propeller domain),但是相比于 RAG2, *BbRAG2L* 缺少 RAG2 的整个 C 端区域,该区域对于 RAG2 的细胞核定位及染色质结合十分重要,同时小鼠 RAG2 的 C 端区域显示能够抑制 RAG 的转座酶活性,因此该区域的缺失可能有利于 *ProtoRAG* 在基因组上的迁移与保留。



红色数字表示内含子相位

图 2 海胆、文昌鱼和小鼠 RAG 基因组结构比对

Fig. 2 Genomic organization of the lancelet *ProtoRAG*, sea urchin *RAG1/2*-like gene locus, and mouse *RAG* locus

2.2 文昌鱼 *ProtoRAG* 转座子的功能研究

ProtoRAG 序列既具有一些典型的转座子特征,又能够编码与脊椎动物 RAG 分子同源的 *BbRAG1L/2L* 蛋白,因此对它是否依然保持转座活性,又是否能够如同 RAG 一样介导重排反应的问题进行了研究。首先,在离体细胞体系中,同时转染 *BbRAG1L* 和 *BbRAG2L* 表达质粒能够引发工程质粒上两侧具有 *ProtoRAG* TIR 序列的片段的切割。同时, *BbRAG1L/2L* 蛋白并不能够切割两侧为 RSS 序列的底物,而小鼠 RAG1/2 蛋白也不能切割两侧为 *ProtoRAG* TIR 序列的底物,这表明在经过 5 亿多年的独立演化后这两套 RAG 系统在功能上已经不能兼容彼此。通过哺乳动物细胞表达纯化的 *BbRAG1L/2L* 蛋白同样能够在体外切割两侧含 *ProtoRAG* TIR 序列的底物,与 RAG 蛋白相似的是其切割活性强烈依赖于 DNA 结合蛋白 HMGB1 (high mobility group box-1) 以及 Mg^{2+} 和 Mn^{2+} 的参与,同时其切割也采用与脊椎动物 RAG 蛋白的“缺口/发夹”相似的机制,能够在 TIR 序列的邻近位置产生缺口并形成发夹样的闭合结构。工程质粒在被 *BbRAG1L/2L* 蛋白切割后留下的质粒 DNA 位点被

称为宿主 DNA 连接(HDJ),对含有 HDJ 序列质粒的分析显示多数 HDJ 序列(>90%)包含小片段缺失,而少数则有短片段的插入,这些插入片段几乎全部为“回文-核苷酸(P-核苷酸)”插入。P-核苷酸插入现象是 V(D)J 重排反应的重要特征,是由 RAG 蛋白介导的闭合发夹结构 DNA 末端在随机碱基位点打开所产生的。进一步对 BbRAG1L/2L 蛋白切割下的 TIR 末端片段的研究显示,它们中多数发生了分子内转座,造成缺失或者倒位;特别是其中有少数(约 11%)发生了精确的“TIR 末端-TIR 末端连接”,这与 V(D)J 重排中的信号连接(signal joint)相似,信号连接的重要性在于防止 RSS 回插基因组,造成细胞基因组失去稳定性乃至癌变,这说明文昌鱼 RAG 样蛋白已经具备了脊椎动物 RAG 的关键功能,是 RAG 转座子被驯化的关键前提之一(图 3)^[11]。同时,BbRAG1L/2L 蛋白也能介导 *ProtoRAG* 转座子在分子间的转移,将含有四环素抗性基因(Tet^r)的线性 *ProtoRAG* 转座子、含有卡那霉素抗性基因(Kan^r)的目的质粒及纯化的 BbRAG1L/2L 蛋白一起孵育后,在转化细菌后能够获得 Kan^r/Tet^r 双抗性的克隆,对从这些双抗性克隆中提取的质粒分析显示,它们都包含完整的转座子序列以及两侧的 TSD 序列(80%长度为 5 bp),并且转座子插入了目的质粒上多个不同的位点,而在离体的人细胞系中进行类似的实验,也得到了一致的结果,因此尽管存在极大的物种差异,文昌鱼 *ProtoRAG* 转座子依然能够真实地在人细胞系中介导转座反应。

综上所述,*ProtoRAG* 在文昌鱼胚系中依然活跃着,它所编码的 BbRAG1L/2L 蛋白能够介导自身的转座以

及宿主 DNA 的重排,其作用机制与人类 RAG 蛋白介导的抗体重排机制非常相似,同时又具有典型的转座活性,能够介导分子内和分子间转座。因此,*ProtoRAG* 被认为就是长期寻找的祖先 RAG 转座子的分子活化石^[12]。在脊索动物演化过程中 RAG 转座子垂直传播着,它在文昌鱼中依然保持着转座子活性;而在有颌类脊椎动物中则被驯化为介导 V(D)J 重排过程的重要分子,最终促成适应性免疫的诞生。*ProtoRAG* 的发现被认为是多样性生成奇迹的证据(G.O.D.)^[13],再次说明了转座子对物种起源的重要作用。

3 RAG 转座子的演化

文昌鱼之后,研究人员在其他物种中也陆续发现了 RAG 转座子的存在,包括海胆、半索动物等,在一些半索动物中甚至多达上千拷贝,对多个 RAG 转座子家族的序列分析发现,它们中的一些可能已经被驯化为宿主基因,有可能介导免疫或者其他方面的宿主重排功能。RAG 转座子在后口动物谱系中的广泛存在说明活跃的 RAG 转座子可能插入后口动物共同祖先的基因组中,而在随后漫长的演化过程中,在不同物种中,有的慢慢不再活跃,而成为基因组上的“化石”;有的依然活跃至今;有的则被驯化为宿主基因,参与宿主免疫反应。在数百万年演化过程中,许多后口动物谱系中一直活跃着不同拷贝数的 RAG 转座子,此外,RAG 转座子属于 DDE 转座酶超家族,而 DDE 转座酶则被报道与 DNA 修复及复制过程相关,这显然增加了 RAG 转座子演化成为 V(D)J 重组酶的可能性。因此,在有颌类脊椎动物中 V(D)J 重排机制的出现可能是一个可预见的遗传事件。

4 结论与展望

文昌鱼 RAG 转座子的发现为抗体重排的转座子起源假说提供了最有力和最直接的证据。在生化、结构以及功能等方面对 *ProtoRAG* 的深入研究,将能够进一步加深对于 RAG 分子以及脊椎动物适应性免疫起源的认知。同时,基于文昌鱼原始 RAG 蛋白的功能,有望设计出高效且精准的抗体编辑工具酶,为利用重排设计新的免疫抗体提供崭新的基因编辑思路和技术。

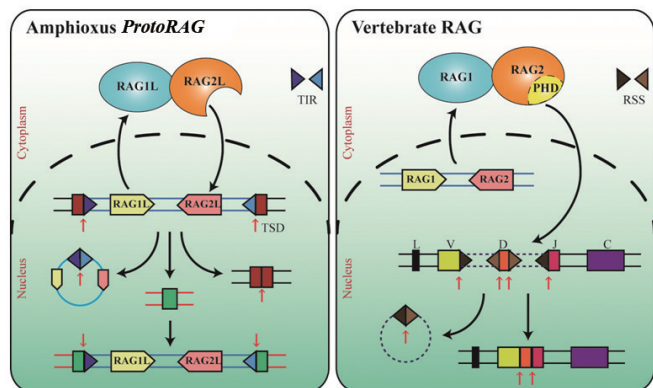


图3 文昌鱼 *ProtoRAG* 与脊椎动物 RAG 重排机制对比

Fig. 3 The rearrangement mechanism of lancelet *ProtoRAG* and vertebrate RAG proteins

参考文献 (References)

- [1] Sakano H, Huppi K, Heinrich G, et al. Sequences at the somatic recombination sites of immunoglobulin light-chain genes[J]. Nature, 1979, 280(5720): 288–294.
- [2] Schatz D G, Oettinger M A, Baltimore D. The V(D)J recombination activating gene, RAG-1[J]. Cell, 1989, 59(6): 1035–1048.
- [3] Oettinger M A, Schatz D G, Gorka C, et al. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination[J]. Science, 1990, 248(4962): 1517–1523.
- [4] Agrawal A, Eastman Q M, Schatz D G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system[J]. Nature, 1998, 394(6695): 744–751.
- [5] Kapitonov V V, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons[J]. PLoS Biology, 2005, 3(6): e181.
- [6] Fugmann S D, Messier C, Novack L A, et al. An ancient evolutionary origin of the *RAG1/2* gene locus[J]. PNAS, 2006, 103(10): 3728–3733.
- [7] Dong M, Fu Y, Yu C, et al. Identification and characterisation of a homolog of an activation gene for the recombination activating gene 1 (RAG 1) in amphioxus[J]. Fish and Shellfish Immunology, 2005, 19(2): 165–174.
- [8] Huang S, Yuan S, Guo L, et al. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity[J]. Genome Research, 2008, 18(7): 1112–1126.
- [9] Zhang Y, Xu K, Deng A, et al. An amphioxus RAG1-like DNA fragment encodes a functional central domain of vertebrate core RAG1[J]. PNAS, 2013, 111(1): 397–402.
- [10] Huang S, Chen Z, Yan X, et al. Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes[J]. Nature Communications, 2014, 5: 5896.
- [11] Huang S, Tao X, Yuan S, et al. Discovery of an active RAG transposon illuminates the origins of V(D)J recombination[J]. Cell, 2016, 166(1): 102–114.
- [12] Ding X, Dong C. Evolution of RAG transposon unveiled[J]. Science China Life Sciences, 2016, 59(9): 968–970.
- [13] Flajnik M F. Evidence of G.O.D.'s miracle: Unearthing a RAG transposon[J]. Cell, 2016, 166(1): 11–12.

The origin of human V(D)J recombination mechanism

YAN Xinyu¹, HUANG Shengfeng¹, XU Anlong^{1,2}

1. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

2. School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China

Abstract Vertebrate V(D)J recombination is mediated by RAG proteins and plays a core role in the vertebrate adaptive immune system. Thus, the origin of V(D)J recombination mechanism has always been a hot spot of immunology research. In this study, amphioxus is chosen as the model species, which is called "living fossil" of vertebrate ancestor. Based on in-depth bioinformatics analyses of three amphioxus genomes and systematic functional research, the *ProtoRAG* transposon encoding RAG homologous proteins is found. Functional experiments show that *ProtoRAG* is an active transposon encoding original RAG1/2-like protein, which not only mediates the transfer of itself but also catalyzes the V(D)J rearrangement in a way similar to RAGs in vertebrates. Thus, *ProtoRAG* represents a molecular "living fossil" of the long-sought ancestral RAG transposon. These findings provide direct and compelling evidence that V(D)J recombination evolved subsequent to an ancestral transposon, as proposed by Tonegawa in 1979. The results of this study are published in Cell in June 2016 and were considered to be evidence of G.O.D.'s (Generation Of Diversity) miracle.

Keywords V(D)J recombination; RAG; amphioxus; *ProtoRAG* transposon ●



(责任编辑 傅雪)