

丙型肝炎病毒基因组多样性的 分子基础及临床意义

岳莉莉 齐义鹏*

(武汉大学病毒学研究所, 武汉 430072)

摘要 丙型肝炎病毒 (HCV) 是一种有包膜的单股正链 RNA 病毒, 其编码的前体多聚大蛋白可加工成具有不同功能的结构蛋白和非结构蛋白。研究结果表明, HCV 核酸序列的变异很大, 因此存在许多不同的基因型和亚型。了解 HCV 基因组多样性的分子基础, 对 HCV 的病原学、流行病学、诊断和治疗都具有重要的指导意义。

关键词 HCV, 基因多变性, 临床意义

Molecular foundation and clinical significance of Hepatitis C virus genomic variability/Yue Lili, Qi Yipeng// CHINESE BIODIVERSITY. —1997, 5 (3): 179~184

Hepatitis C virus (HCV) is an enveloped virus with a plus-stranded RNA genome. It contains a large open reading frame (ORF) encoding a polyprotein which can be proteolytically cleaved to structural and nonstructural proteins with different functions. The isolates of HCV show substantial nucleotide sequence variability distributed throughout the viral genome. HCV isolates can be classified into many genotypes and subtypes. Studies on HCV heterogeneity has an important guiding significance for the pathogenesis, epidemiology, diagnose and therapy of HCV.

Key words HCV, genomic variability, clinical significance

Author's address Institute of Virology, Wuhan University, Wuhan 430072

丙型肝炎病毒 (*Hepatitis C virus*, HCV) 是一种传播广泛的血源性肝炎病原体。自 1989 年美国 Chiron 公司首先克隆了 HCV cDNA 以来, 已报道了数株 HCV 全基因序列及许多 HCV 株的 cDNA 克隆片段。目前, HCV 分子水平上的研究已成为国内外病毒性肝病研究领域的热点, 进展十分迅速, 并大大促进病毒性肝炎预防和治疗上的研究。

1 HCV 基因组的结构和功能

HCV 是单股正链 RNA 病毒, 为黄病毒科丙型肝炎病毒属^[1], HCV 颗粒外有脂质包膜, 直径约 30~60 nm, 内含约 9.4 kb 正链 RNA, 与黄病毒和瘟病毒相似。HCV RNA 有单一的长开放阅读框 (open reading fame, ORF), 编码约 3010~3033 个氨基酸的前体多聚大蛋白, 经细胞和病毒的蛋白酶加工, 修饰成具有不同功能的结构蛋白和非结构蛋白 (见图 1)。

5'端和3'端为非翻译区, 但含有调节因子, 在基因组复制中起重要作用。已发现 HCV 5'端非翻译区 (5' NCR) 含有 3~4 个最长可编码 28 个氨基酸的小 ORF, 可能是翻译的负调节元件。采用无细胞翻译系统或体外细胞培养研究了 HCV RNA 的翻译起始机制, 提示 5' NCR 内

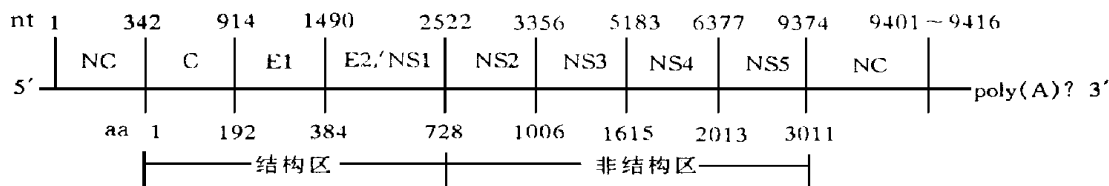


图1 HCV基因组图谱

Fig. 1 Map of HCV genome^[2]

存在内部核糖体进入位点, HCV RNA 在体外细胞中的翻译不依赖于帽子结构。说明 HCV 的翻译起始与小 RNA 病毒相似, 而与多数真核 mRNA 翻译的“核糖体扫描”机制不同。3' NCR 很短且不均一, 含有 27~55 个核苷酸, 由于在 PCR 扩增实验中发现 HCV cDNA 可与 oligo (dT) 结合, 估计 3' 端有 poly(A) 尾, 但已分离的两株 HCV 为 12 mer 的 poly(u) 尾^[3,4]。

HCV 基因组 C 区编码核心蛋白, 分子量为 22 KD, 该蛋白的特性与黄病毒和瘟病毒的核壳蛋白相似: 属碱性蛋白, 不含糖基化位点, 主要出现于胞质, 并与内质网紧密相连^[5]。在除去其 C-端疏水区后, 核心蛋白出现于核内, 故推测 C-端疏水区与该蛋白在细胞中的定位有关。在昆虫细胞中表达的 HCV 核心蛋白, 无 C-端疏水区, 可在细胞内形成包涵体, 并磷酸化。另据推测, 核心蛋白的 99~102 位是一个 DNA 结合基序 SPRG (Ser-Pro-Arg-Gly), 在 9~13、38~43、58~64、66~71 和 113~117 等区段则含有核定位信号, 此外还能找到蛋白激酶 A 和蛋白激酶 C 的若干个识别位点, 核心蛋白的 N 端 122 个氨基酸有结合 HBV RNA 的能力, 可在人肝癌细胞系 HuH-7 中抑制 HBV 的表达和复制, 说明它可能是一种反式调控蛋白。这一领域的研究工作将为阐明 HBV 与 HCV 的相互作用及 HCV 导致 HCC (肝细胞癌) 的分子机理奠定基础^[6]。HCV 包膜糖蛋白 E1 的 cDNA 片段可表达 32~35 KD 的糖蛋白 (gp35), 有 5~6 个 N-糖基化位点, 该蛋白 N-端 175~191 区段内发现潜在的信号肽序列, 推测其 C-端还有一个跨膜疏水区。包膜糖蛋白 E2 约 68~72 KD (gp70), 一般不分泌, 可能是包膜的组成, 其 N-端也有潜在的信号肽序列, C 端可能有跨膜疏水区, 在昆虫细胞表达系统中发现, 除去该疏水区, E2 蛋白的水溶性提高^[7]。

体外表达系统证明, E2 下游可表达 21~23 KD 的 NS 2 蛋白, 是 NS 2~3 蛋白酶的组分, 经生物化学和突变分析表明, NS 2~3 蛋白酶是锌依赖性金属蛋白酶 (Zinc-dependent metalloproteinase)。完整的 NS 3 蛋白为 70~72 KD, 它除作为 NS 2~3 蛋白酶的成份外, 还有丝氨酸蛋白酶和解旋酶活性, 其 N-端含有丝氨酸蛋白酶所特有的天组丝氨酸催化活性中心, 据认为活性中心的保守残基与电荷传递有关。缺失诱变的结果也表明, NS 3 的丝氨酸蛋白酶 N-端最小活性单位位于 1049~1065, 它对 NS 3/NS 4 位点的切割为 cis 调控, 而对其它位点的切割则是 trans 调控, 因而在 NS 4 及 NS 5 的翻译后加工过程中起重要作用。NS 3 的 C-端具有 RNA 解旋酶的活性。NS 4 蛋白的功能尚不明确, 已发现其在多聚大蛋白加工过程中 (尤其是对 NS 4B/5A 位点的切割) 起重要作用^[8]。在体外转染细胞中检测到两种 NS 4 加工产物: NS 4A 和 NS 4B, 它们的分子量分别为 4~10 KD 和 26~31 KD。最大的 NS 5 可能是一个具有依赖于 RNA 的 RNA 聚合酶活性的多功能蛋白 (即复制酶), 其中含有 RNA 病毒复制酶所共有的 GDD 基序, 在体外转染细胞中检测到至少两种 NS 5 加工产物: NS 5A 和 NS 5B, 分别为 56~58 KD 和 65~70 KD, 具有 RNA 聚合酶活性的可能是 NS 5B 蛋白, 它可能也是复制复合物的组分。在 NS 5A 的 N-端还有一个大约 2 KD 的 NS 5C 蛋白, 该蛋白的翻译后加工也是由 NS 3 的丝氨酸蛋白酶所催化的, HCV 的蛋白加工过程见图 2:

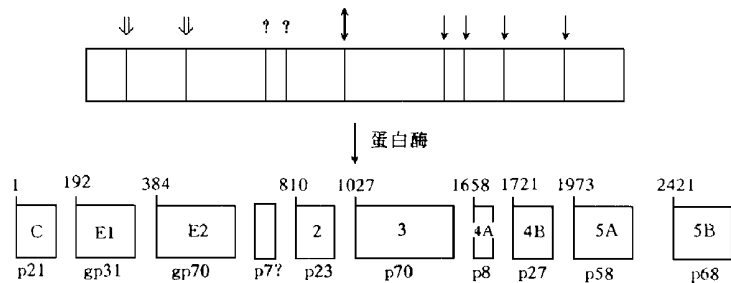


图 2 HCV H多聚大蛋白及其加工后产物^[9]

Fig. 2 Diagram of the HCV-Hpolyprotein and its cleavage products
注：宿主内质网的信号肽酶切割 C/ E1， E1/ E2 (↓)， HCV NS 2 ~ 3 蛋白酶切割 NS 2/ NS 3 (↕)， NS 3 丝氨酸蛋白酶切割 3/ 4A， 4A/ 4B， 4B/ 5A ()， 由未知的蛋白酶切割 E2/ NS 2 (?)
The number at the top of each cleavage products indicates the position of its N-terminal residul in the polyprotein sequence the putative cleavage sites for host signalase (↓)， the HCV NS 2 ~ 3 proteinase (↕)， the NS 3 serine proteinase ()， unknown proteinase (?)

2 HCV 基因的变异

HCV RNA 的突变率每年每个基因位置为 1.9×10^{-3} 碱基替换^[10]，是原核和真核 DNA 突变率的 100×10^4 倍，使得各地分离的 HCV 株、甚至从同一患者体内分离的 HCV 都有一定的差异。

HCV 基因组以 5' NCR 最保守，各株间同源性 95%，该区还有两个不变区，常作为 PCR 引物的核苷酸序列。C、NS 3 和 NS 5' 区相对保守，包膜蛋白编码区（E1，E2）变异较大，E1 可分成两组，同组内高度保守，而两组间同源性仅 50% ~ 60%。E2 区变异最大，在氨基末端有两个高变区（HV1 和 HV2），两者编码的蛋白有相同的亲水区和二级结构，可能在免疫中起重要作用。Simmonds 的分析结果表明，HCV 基因组变异率高达 33%，其中核心区为 8%，E 区为 54%，NS 区约 5 ~ 6%，这种序列上的变化足以使 HCV 的抗原性和生物学特性发生明显的改变^[11]。

由于 RNA 病毒复制的多聚酶缺乏“校读”（proofreading）功能，加上 RNA 病毒为适应环境和逃避宿主的免疫清除而发生的突变，逐渐形成了新的 HCV 突变株，并各自独立演化同时又保证 HCV 重要功能基因的相对稳定和某些基因的相对易变，从而使 HCV 既能表现为高度易变，又能区分为一定的基因型和许多亚型，这也是 HCV 分型的机理。

3 HCV 多样性与分型

由于 HCV 的多样性，加之用于分型的基因部位和采用的方法不同，导致 HCV 基因分型十分混乱。Okamoto 根据 C 区序列设计了 1 对公用引物和 5 对分型引物，用 PCR 扩增后，可将 HCV 分为 ~ 型^[12]，Enomoto 从 NS 5' 区分离的片段中，将 HCV 分成 K 1、K 22 型^[13]，美国分离株 PT-1 是 K 1 的亚型，K 2 可进一步分为 K 2a 和 K 2b 亚型，各亚型间的核苷酸同源性均为 80%^[14]。通过 5' NCR 序列的变异可将 HCV 分成 ~ 等 6 大类型^[11,15]。目前缺少统一的分型方法和标准，因此有些学者提议根据 NS 5' 区 221 bp 片段序列的同源性，可将各国分离的 HCV 株分为 6 个基因型（见表 1）。

按照此分型系统，各型之间核苷酸的同源性是 56% ~ 72%，亚型之间是 74% ~ 86%，同一型不同分离株之间 88% ~ 100%。因此与 11 种 221 bp 片段的序列相比较，若某种被鉴定的 HCV 序列同源性 < 72%，则可确认为是一种新的 HCV 基因型；但若与其中任意一株

的同源性 > 88 %，则不能被认定为一种新的基因型^[11]。

表 1 HCV 基因分型

Table 1 Genoic classification of HCV

HCV 变异株 Variant plant of HCV	来源 Origin	建议分型 Suggestion	Chiron 分型	Enomoto 分型	Mori/ Okamoto 分型
HCV-1	Choo	1a	I	K-PT	I
HCV-J-B K	Kato	1b	II	K-I	II
	Takamizawa				
EG-28	Simmonds	1c			
HC-J6	Okamoto	2a	III	K-2a	III
HC-J8	Okamoto	2b	III	K-2b	III
Tor-994	Chan				
E-b1	Chan	3a	IV		V
Ta	Mori				
Tb	Mori	3b			VI
EG-16	Simmonds	4			
29, 33					
SA-1, 7	Bukh				
11	Cha	5	V		
Hk-1, 2	Simmonds	6			
3, 4	Bukh				

4 HCV 变异对宿主免疫反应性的影响

目前对 HCV 感染与免疫的了解还不很清楚，C、E1、E2、NS 4 和 NS 5 蛋白均含有优势抗原表位，从动物实验的结果上看，机体针对 HCV 感染的免疫反应相对较弱。用 HCV 交叉攻击感染黑猩猩，发现无有效的保护性抗体应答^[14]。由于在 HCV 蛋白上发现了几种细胞毒性 T 细胞（CTL）表位，故推测 CTL 在控制感染免疫反应中可能起重要作用。E2HV 区氨基酸的第 384 ~ 414 区有类似于人免疫缺陷病毒 gp120V3 区的中和抗原决定簇，有型、亚型甚至 HCV 克隆间特异性，使 HCV 通过抗原变异而逃避宿主的免疫选择，成为 HCV 持续感染的主要原因之一。NS 4 蛋白的免疫活性区也高度易变，无交叉反应，因此它们的主要抗原决定簇是型特异的^[16、17]。所以在免疫诊断时，因不同基因型 HCV 感染而影响抗体效价的检测。已发现 HCV 有一定的地区和人群分布特征，反映出 HCV 进化和亲缘关系上的差异，不同国家确实存在不同的基因型，如 1、2 型普遍流行，欧美国家流行 3 型，日本、中国主要是 2 型，泰国、印度、中东几乎都是 4 型^[14]，南非和香港分别是 5、6 型。

5 HCV 变异与临床和治疗的关系

HCV 急性感染后转为慢性的约占 50 %左右，这是因为：（1）E 区最易变异，导致包膜蛋白的改变，逃避宿主免疫；（2）肝外组织存在 HCV 感染。HCV 很可能是肝细胞癌（HCC）的病原，抗-HCV 阳性者较阴性者 HCC 发生的危险性增加 10 倍。有人认为持续性 HCV 感染所致的肝细胞损伤、坏死、炎症可引起肝硬化，通过肝细胞再生与间变，发展到肝癌^[18]。最近 HCV 的克隆和分子分析为研究两者之间的关系提供了有力的工具，慢性丙肝患者对干扰素（IFN）的治疗效果相差很大，有关原因不清，贺古真提出 HCV 基因型不

同, IFN 的抗病毒作用也不同, 其中 III 型最有效, II 型较差^[19]。

6 HCV 的实验室诊断

6.1 免疫学诊断 由于血中 HCV 抗原量太少, 无法进行抗原检测, 故只能测其抗体。目前, 从 HCV cDNA 克隆获得的不同类型的基因工程抗原片段不断增加, 促进了 HCV 诊断技术的发展。首先得到的是 HCV C-100, 它是由美国 Chiron 公司将 HCV cDNA 融合于酵母菌中表达的多肽抗原, 约 363 个氨基酸。将 C 100 与人超氧化物歧化酶 (superoxide dismutase, SOD) 基因连接, 在细菌或酵母菌中表达融合蛋白 C 100-3 作为包被抗原, 检测血清中抗-HCV, 成为第一代诊断试剂盒。目前世界各国普遍使用的均是以这些抗原为基础组装的产品。我国也有几家单位生产, 主要用于筛选献血员, 以防输血后肝炎的发生。

第二代诊断试剂盒是将 C-100-3、5-1-1、C 22-3 和 C 33c 结合在一起, 从而增加了试验的敏感性和特异性^[20], 它的诊断标准: 与 4 种抗原中 2 种以上抗原起反应者为阳性, 仅与 1 种抗原起反应者可疑, 均不起反应者为阴性。C 区基因表达产物具有较好的抗原稳定性, NS 3、NS 5 具有优势抗原表位, 这些区域多肽, 可确保反应的广源性和诊断的敏感性。

6.2 HCV RNA 的检测 聚合酶链反应 (PCR) 已用于 HCV RNA 的检测。所用的引物多取自 5' NCR, 其灵敏度可检测出低于黑猩猩最小感染剂量 (CID) 10 倍。PCR 有以下优点: (1) 可作为判断 HCV 感染有无传染的可靠指标; (2) 特异性好; (3) 敏感性高; (4) 有利于早期诊断; (5) 应用广泛。但 PCR 法也有其局限性, 由于 HCV RNA 存在显著的易源性和多变性, 因而引物的设计可影响其敏感性; 由于技术过程较复杂, 若操作不当, 易假阴性; 操作时若污染可呈假阳性。且试剂较昂贵, 这给推广应用带来困难。

从上述不难看出, 目前检测 HCV 的方法虽不少, 但各有其长短。所以尚需进一步寻找特异、敏感、简便、快速的检测方法, 以适应临床需要。

综上所述, HCV 核酸和蛋白质变异性很大, 存在不同的型和亚型。在分子水平上对其变异性的研究对揭示丙型肝炎的病原学、流行病学及在临床和治疗方面都有重要的意义, 同时也有助于了解病毒和宿主间的相互作用, 促进对 HCV 诊断试剂盒和疫苗的研制。但控制 HCV 的前景并不乐观, 其原因如下: (1) HCV 有多种基因型, 异源性较高; (2) 存在类似 HCV 毒株的基因漂移; (3) 对再感染无保护作用; (4) 隐匿性、慢性感染比例较高; (5) 相当比例的高发人群尚未发现。因此, 目前对 HCV 的认识尚欠深入, 还有许多工作要做。随着 HCV 分子生物学研究的不断深入, 在不久的将来, 人们在控制丙型肝炎方面必将有更大的突破。

参 考 文 献

- 1 Miller R H, R H Purcell. *Hepatitis C virus* shares amino acid sequence similarity with pestiviruses and flaviviruses as well as plantvirus supergroups. *Proc. Natl. Acad. Sci. USA*, 1990, **87**: 2057 ~ 2061
- 2 Inehauspe G, S Zebedee, D H Lee, M Sugitani, M Nasoff et al. Genomic structure of the human prototype strain H of *Hepatitis C virus*: comparison with American and Japanese isolates. *Proc. Natl. Acad. Sci. USA*, 1991, **88**: 1029 ~ 1032
- 3 Tsukiyama Koharak, N Iizuka, M Kohara, A Nomoto. Internal ribosome entry site within *Hepatitis C virus* RNA. *J. Virol.*, 1992, **66**: 1476 ~ 1483
- 4 Wang C, P Sarnow, A Siddiqui. Translation of human *Hepatitis C virus* RNA in cultured cells is mediated by an internal ribosomebinding mechanism. *J. Virol.*, 1993, **67**: 3338 ~ 3344
- 5 Harada S, Y Watanabe, K Takeuchi, T Suzuki, T Katayama et al. Expression of processed core protein of *Hepatitis C virus* in mammalian cells. *J. Virol.*, 1993, **65**: 3015 ~ 3021

- 6 Shih C M, S J Lo, T Miyamura, S Y Chen, Y H W Lee. Suppression of *Hepatitis B virus* expression and replication by *Hepatitis C virus* core protein in HuH-7 cells. *J. Virol.*, 1993, **67**: 5823 ~ 5832
- 7 Matsuura Y, S Harada, R Suzuki, Y Watanabe, I Inoue et al. Expression of processed envelope protein of *Hepatitis C virus* in mammalian and insect cells. *J. Virol.*, 1992, **66**: 1425 ~ 1431
- 8 Bartenschlager R, L Ahlborn-laake, J Mous, H Jacobsen. Kinetic and structural analyses of *Hepatitis C virus* polyprotein processing. *J. Virol.*, 1994, **68**: 5045 ~ 5055
- 9 Lin C, B D Lindenbach, B M Pragai, D W McCourt, C M Rice. Processing in the *Hepatitis C virus* E2-NS 2 region: identification of p7 and two distinct E2-specific products with different C termini. *J. Virol.*, 1994, **68**: 5063 ~ 5073
- 10 Hijikata M, H Mizushima, T Akagi, S Mori, N Kakiuchi et al. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of *Hepatitis C virus*. *J. Virol.*, 1993, **67**: 4665 ~ 4675
- 11 Simmonds P, E C Holmes, T A Cha, S W Chan, F McOmish et al. Classification of *Hepatitis C virus* into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J. Gen. Virol.*, 1993, **74**: 2391 ~ 2399
- 12 Okamoto H, H Tokita, M Sakamoto, M Horikita, Kojima et al. Characterization of the genomic sequence of type V (or 3a) *Hepatitis C virus* isolates and PCR primers for specific detection. *J. Gen. Virol.*, 1993, **74**: 2385 ~ 2390
- 13 Enomoto N, A Takada, T Nakao, T Date. There are two major types of *Hepatitis C virus* in Japan. *Biochem. Biophys. Res. Comm.*, 1990, **170**: 1021 ~ 1025
- 14 Simmonds P, F McOmish, P L Yap, S W Chan, C K Lin et al. Sequence variability in the 5 non-coding region of *Hepatitis C virus*: identification of a new virus type and restrictions on sequence diversity. *J. Gen. Virol.*, 1993, **74**: 661 ~ 668
- 15 Weiner A J, H M Geysen, C Christopherson, I E Hall, T Mason et al. Immune selection of *Hepatitis C virus* (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl. Acad. Sci. USA*, 1992, **89**: 3468 ~ 3472
- 16 Simmonds P, K A Rose, S Graham, S W Chan, F McOmish et al. Mapping of serotype-specific, immunodominant epitopes in the NS-4 region of *Hepatitis C virus* (HCV) —use of type —specific peptides to serologically differentiate infections with HCV type 1, type 2, and type 3. *J. Clin. Microbio.*, 1993, **31**: 1493 ~ 1503
- 17 Lesniewski R R, K B Boardway, J M Casey. Hypervariable 5 terminus of *Hepatitis C virus* E2/NS 1 encodes antigenically distinct variants. *J. Med. Virol.*, 1993, **40**: 150
- 18 Yuki N, N Hayashi, A Kasahara, H Hagiwara et al. *Hepatitis C virus* in Japanese patients with hepatocellular carcinoma. *Dig. Dis. Sci.*, 1992, **37**: 65
- 19 贺古真等, 肝脏. 1992, **33** (增刊 3): 34
- 20 Vanderpoel C L, H T M Cuypers, H W Reesink, A J Weiner, S Quan, R Dinello et al. Confirmation of *Hepatitis C virus* infection by new four-antigen recombinant immunoblot assay. *Lancet.*, 1991, **337**: 317 ~ 319