维普资讯 http://www.cqvip.com

第6卷第3期 1991年9月
中国病毒学 VIROLOGICA SINICA
Vol.6 No.3 Sept.1991

计算机在确定牛病毒性腹泻病毒基因组常用限制性内切酶位点中的应用

陆家骎 王薇 糜克永

（武汉纺织工学院，武汉430073） （中国科学院武汉病毒研究所，武汉430071）

提 要

应用计算机测定71种限制性内切酶在牛病毒性腹泻病毒NADL株(BVDV NADL)基因组的cDNA核苷酸序列中酶切位点，结果表明其中58种酶的酶切位点所在的核苷酸序列数，也表明了其中13种限制性内切酶在核苷酸序列中无酶切位点。这些结果有助于进一步研究BVDV InL株的cDNA基因库。

关键词：牛病毒性腹泻病毒 限制性内切酶 计算机

牛病毒性腹泻病毒(BVDV)在世界范围内严重影响养牛业的发展[1]。糜克永等人在内蒙古羔羊下痢病流行区分离到一株BVDV[2]，并于1984—1986年间在该病区作了病原调查，证明该病毒病原为羔羊下痢病重要的病原之一[3]，根据湖南和河南等省的调查结果，该病毒对内地奶牛场的发展影响也很大。为了防治由病毒引起的疾病流行，我们进行了BVDV基因探针及基因工程疫苗的研究，建立了含有编码免疫原性蛋白的核酸片段基因库。国外Collet等人完成了对BVDV NADL株基因组克隆插入cDNA核苷酸序列的分析，并报道有12537个核苷酸，说明该序列具有编码3988个氨基酸，分子量为449KD的蛋白，且具有免疫原性[4]。我们以Collett等人提出的核苷酸序列为基础，应用电子计算机技术将核苷酸序列及71种限制性内切酶作用特殊序列片段编制程序输入电子计算机进行了研究。

材料和方法

计算机型号：Apollo DN460；使用语言是Fortron[5]语言；计算根据为Collett等人提出的核苷酸序列。

测定方法和步骤：利用Collett等人提出的12537个核苷酸序列，把它们作为原始数据按其排列顺序输入到每个数组中去，一个数组分配一个核苷酸，12537种核苷酸分配了12537个数组。利用循环语句从第1号数组开始到第12537号数组作循环计算。根据限制性内切酶的识别序列中核苷酸的个数、取出相同数目的数组进行比较。在相同的情况下，计算机打印出序列的识别号的位置。在不相同的情况下，直接从下一号数组开始循环。这样一直算到第12537号数组为止。例如要确定PvuI限制性内切酶的酶切位点的核苷酸序列号，它的识别序列为CGAT↓CG，如果计算机循环到第i

本文于1990年4月16日收到，1991年元月2日修回。

维普资讯 http://www.cqvip.com

号数组，则取出第 i 号到第 i+5 号共六个数组来，将此六个数组内核苷酸排列顺序与 CGATCG 比较，在相同情况下，计算机打出 i+3 号数组号来，这就是 PvuI 识别位置的序号。再从第 i+1 号数组开始循环，以便找出另一个序列号来。

按以上的方法和步骤，我们对71种内切酶的酶切位点的核苷酸序列号进行测定，得到58种内切酶的酶切位点序列号，并说明另外13种内切酶无酶切位点。

结果和讨论

Apollo 计算机计算结果获得前58种内切酶作用的核苷酸序列号，列于表 1。

表 1　BVDV基因组的cDNA核苷酸序列[1]酶切位点的测定*

Table 1. The determination of the restriction enzyme sites in the cDNA nucleotide sequences from BVDV NADL genome

限制性内切酶 Restriction enzymes	识别序列 Distinguishable sequence	酶切位点的核苷酸序列号 Numbers of the restriction enzyme sites in the nucleotide sequences
PstI	CTGCA↓C	340, 593, 9835, 12225
EcoR I	G↓AATTC	2450, 5843, 6524, 9745
BamH I	G↓GATCC	710, 5412, 5801, 11517
Hind II	GTR↓PuAC	1744, 1752, 2519, 3958, 4198, 4381, 5623, 5706, 5791, 7387, 8983, 9868, 11032
Hind III	A↓AGCTT	8895, 10401
EcoR V	GAT↓ATC	2654, 5176, 9490, 12150
Pvu I	CGAT↓CG	1900
Xho I	C↓TCGAG	224
Bal	TGG↓CCA	1294, 7193, 9119, 12165
Bgl I	GCC(NNNN↓N)GGC	5269, 8880, 9929, 12142
Xho II	Pu↓GACPy	4584, 5644, 6429, 6718, 10014
Bcl I	T↓GATCA	393, 460, 7201
Hpa I	GTT↓AAC	5791
Pvu II	CAG↓CTG	2198, 7129, 7712, 12348
Bgl II	A↓GATCT	4584, 5644
Sau961	G↓G(N)CC	343, 611, 641, 668, 684, 685, 847, 1651, 1652, 1973, 1989, 2407, 2717, 2828, 3281, 3323, 3802, 4257, 5258, 4644, 4858, 5035, 5128, 5423, 5756, 7249, 7255, 7526, 8421, 8784, 8785, 8811, 8942, 8975, 9717, 9922, 10131, 10333, 10662, 11020, 11709, 11887, 11958, 12281, 12414
Ave II	G↓G(A/T)CC	611, 668, 847, 1989, 2828, 3281, 4644, 5035, 5756, 7249, 7255, 7576, 8421, 8811, 8975, 10333, 11020, 11709, 11887, 12282, 12414

维普资讯 http://www.cqvip.com

限制性内切酶 Restriction enzymes	识别序列 Distinguishable sequence	酶切位点的核苷酸序列号 Numbers of the restriction enzyme sites in the nucleotide sequences
Fnu4H I	GC↓(N)GC	336, 1674, 1814, 1875, 2361, 2819, 2987, 3665, 4517, 4623, 4999, 5081, 6195, 6198, 6351, 6926, 7173, 7710, 7788, 7896, 7989, 8213, 8313, 8385, 8466, 9831, 10854, 11117, 11407, 12242, 12299, 12346
Dde I	C↓T(N)AG	80, 113, 177, 615, 1080, 1113, 1500, 1762, 2333, 2932, 3105, 3360, 3483, 3577, 3680, 3727, 4049, 4588, 4817, 5046, 5067, 5091, 5321, 5417, 5740, 6254, 6344, 6370, 6681, 6731, 7156, 7393, 8234, 8549, 8810, 9017, 9165, 9202, 9857, 9873, 9935, 10084, 10278, 10088, 11619, 12035, 12083, 12227, 12264, 12302, 12349,
Hinf I	G↓A(N)TC	977, 1032, 1118, 1387, 1465, 2224, 2766, 2977, 3068, 3485, 3583, 3794, 4101, 4382, 4962, 5114, 5208, 5229, 5738, 5955, 5978, 6006, 6203, 6785, 7050, 7817, 7961, 8042, 8070, 8492, 8678, 8991, 10895, 11128, 11192, 11598, 11773, 11983
Atu I	CC↓$\binom{A}{T}$GG	286, 639, 656, 1043, 1462, 1721, 1801, 1977, 2306, 2565, 3033, 3449, 3515, 3951, 4055, 4614, 5009, 6920, 6977, 7195, 7247, 7416, 7474, 7523, 8033, 8182, 8869, 8967, 9157, 9334, 9660, 9734, 9808, 9926, 10545, 10865, 10910, 11024, 11133, 11236, 11854
Taq I	T↓CGA	203, 225, 512, 1896, 2739, 4104, 5117, 5822, 5883, 7572, 8661, 9342, 9867, 9891
Rsa I	GT↓AC	182, 376, 770, 1094, 1309, 1448, 2046, 2105, 2280, 4182, 2454, 2624, 2881, 2919, 3178, 3252, 3346, 3531, 3671, 3755, 3812, 3849, 4556, 4597, 4757, 4813, 5292, 5337, 6317, 6638, 8299, 8775, 9270, 9351, 9453, 9791, 10079, 10363, 10612, 11352, 11462, 11487, 11492, 11803, 11826, 12389
Hae III	GG↓CC	61, 90, 344, 643, 686, 1264, 1653, 1974, 2283, 2409, 2535, 2719, 2834, 3325, 3575, 3791, 3804, 4209, 4259, 4808, 4880, 5130, 5425, 6888, 8902, 7193, 7408, 8513, 8597, 8736, 8788, 8944, 9119, 9718, 9923, 10061, 10133, 10863, 10940, 11041, 11213, 11409, 11735, 11960, 12165, 12332
Sau3A	↓GATC	393, 460, 472, 624, 710, 1897, 3593, 3759, 4164, 4281, 4584, 4788, 5412, 5442, 5502, 5644, 5901, 6429, 6556, 6718, 7201, 7375, 7411, 8458, 8629, 9534, 9633, 9729, 10014, 10111, 10138, 10930, 11280, 11517, 11947
Tha I	CG↓CG	8212
Hpa II	C↓CGG	621, 1007, 2084, 2831, 3469, 5038, 5534, 5634, 5873, 6291, 6957, 7253, 7464, 8706, 10726, 10941, 11736, 12245

维普资讯 http://www.cqvip.com

限制性内切酶 Restriction enzymes	识别序列 Distinguishable sequence	酶切位点的核苷酸序列号 Numbers of the restriction enzyme sites in the nucleotide sequences
Alu I	AG↓CT	523, 697, 766, 1166, 1577, 1673, 1816, 2093, 2188, 2198, 2575, 2877, 3263, 3399, 3875, 4002, 4097, 4519, 4636, 5553, 6343, 6350, 6673, 6680, 6944, 7129, 7712, 7760, 7895, 7949, 8479, 8897, 9164, 9410, 9877, 9977, 10105, 10118, 10189, 10197, 10307, 10403, 12226, 12348, 12455, 12519
FocR I*	AA↓TT	11, 43, 56, 1355, 1907, 1932, 1967, 1984, 2256, 2482, 2513, 2761, 3018, 3094, 3142, 3755, 4178, 4300, 4309, 4465, 5387, 5845, 6419, 6526, 7083, 7305, 7810, 7826, 8127, 8504, 9747, 9987, 10151, 10252, 10315, 10648, 10741, 10753, 10921, 11576, 12019, 12086, 12317, 12378
Asc I	GT↓$\binom{AG}{CT}$AC	2, 28, 1922, 3171, 4773, 5622, 6210, 9867
Apy I	CC↓$\binom{A}{T}$GG	286, 639, 656, 1043, 1462, 1721, 1801, 1977, 2396, 2565, 3033, 3449, 3515, 3951, 4055, 4612, 5009, 6920, 6977, 7195, 7247, 7416, 7474, 7523, 8037, 8182, 8869, 8967, 9157, 9334, 9660, 9734, 9807, 9926, 10545, 10865, 10910, 11024, 11133, 11236, 11854
Aha Ⅲ	TTT↓AAA	3097, 7771
Ape I	GGGCC↓C	688, 1665, 4261, 8788
Ava I	G↓$\binom{C}{T}$CG$\binom{A}{G}$G	224, 1817, 6536, 6956, 9941
Bst I	G↓GATCC	710, 5412, 5801, 11517
Ban I	G↓G$\binom{CA}{TG}$CC	1769, 2601, 4595, 4877, 5745, 8871, 10263, 11485, 11801, 11824,
Ban Ⅱ	G$\binom{A}{G}$GC$\binom{C}{T}$↓C	688, 699, 1655, 3062, 3390, 4261, 5583, 5787, 8788, 9216, 9740
Asp718	GG↓TACC	4595, 11485, 11801, 11824
Dpn I	GA↓TC	395, 462, 474, 626, 712, 1899, 3595, 3761, 4166, 4283, 4586, 4790, 5414, 5444, 5504, 5646, 5803, 6431, 6558, 8720, 7203, 7377, 7413, 8460, 8831, 9536, 9635, 9731, 10018, 10113, 10140, 10932, 11262, 11519, 11949
Dra I	TTT↓AAA	3097, 7771
Hha I	GCG↓C	1526, 5152, 6925, 8189, 8389, 10581
Hinc Ⅱ	GT$\binom{T}{C}$↓$\binom{G}{A}$AC	1741, 1752, 2519, 3958, 4198, 4381, 5623, 5706, 5791, 7387, 8983, 9868, 11032
Kpn I	GGTAC↓C	4599, 11489, 11805, 11828
Mbo I	↓GATC	393, 460, 472, 624, 710, 1897, 3593, 3759, 4164, 4281, 4584, 4788, 5412, 5442, 5502, 5844, 5801, 6429, 6556, 6718, 7201, 7375, 7411, 8458, 8829, 9534, 9633, 9729, 10014, 10131, 10138, 10930, 11260, 11517, 11947

维普资讯 http://www.cqvip.com

限制性内切酶 Restriction enzymes	识别序列 Distinguishable sequence	酶切位点的核苷酸序列号 Numbers of the restriction enzyme sites in the nucleotide sequences
Msp I	C↓CGG	621, 1007, 2084, 2831, 3469, 5038, 5534, 5634, 5873, 6291, 6957, 7253, 7464, 8706, 10726, 10941, 11736, 12245
Nsp II	A T G(G)GC(C)↓C T A	688, 699, 972, 1655, 2332, 2463, 3062, 3390, 4261, 5454, 5533, 5750, 5787, 7347, 8240, 8768, 8876, 9063, 9216, 9519, 9740, 10268, 10296, 10356, 10479
Nsp III	C↓(C/T)CG(A/G)G	224, 1817, 6536, 6956, 994[illegible]
Pal I	GG↓CC	61, 90, 344, 643, 686, 1294, 1653, 1974, 2263, 2409, 2535, 2719, 2834, 3325, 3575, 3791, 3804, 4209, 4259, 4808, 4860, 5130, 5425, 6888, 6902, 7193, 7408, 8513, 8597, 8736, 8786, 8944, 9119, 9718, 9923, 10061, 10133, 10663, 10940, 11041, 11213, 11409, 11735, 11960, 12165, 12332,
Sac I	GAGCT↓C	699
Sal I	G↓TCGAC	5621, 9866
Sma I	CCC↓GGG	6958
Sph I	GCATG↓C	251
Xma I	C↓CCGGG	6956
Mae I	C↓TAG	63, 102, 123, 944, 1328, 1469, 1551, 1578, 1826, 2094, 2228, 3167, 3876, 4094, 4098, 4355, 4931, 5304, 5676, 6930, 6945, 7400, 7493, 7586, 7931, 7946, 8817, 9230, 9347, 9390, 9407, 9455, 9774, 9848, 10194, 10358, 10449, 11078, *11253*, *12126*, *12554*
Nci I	CC↓(C/G)GG	622, 3470, 5039, 5534, 6292, 6957, 6958, 7253, 7465, 10942, 11737, 12245
Nco I	C↓CATGG	4647, 4900, 8318
Sca I	AGT↓ACT	1309, 3812, 3849, 9453, 10079
Stu I	AGG↓CCT	61, 2535, 3575, 6888

无酶切位点的酶是Xba I，Atu I，Hae II，Aat II，BstE II，Asp770，Cla I，Mlu I，Teh III，Mae III，Nac I，Sau I，Nra I等13种。从表1中可以查知FcoR I，Alu I，Sau3A，Hae III，Atul，Hinf I，Apy I等14种酶有30个以上酶切位点，这样多的酶切位点对于研究基因探针和基因库中各种片段的核酸密码含义是困难的，而Pun II，Pst I，EcoR I，BamH I，Bg I I，Bal I，EcoRV，Ava I，Bst I，Asp718，Kpn I，Nsp III，Stu I，Sca I等仅有4至5个酶切位点，这给我们上述的研究带来了方便。这些酶中的Pstl酶的酶切位点分布在序列号在340，593，9835，12225处，分布比较均匀，这对于确定有特异性的基因探针片段以及精确测定12573个核苷酸序列的病毒敏感区[2]和确定编码免疫原性蛋白的基因片段、研制基因工程疫苗是有很大价值的。

夏志清[6]和杨学楼等人[7]利用计算机技术研究人干扰素编码基因的酶切位点作了报

维普资讯 http://www.cqvip.com

道。我们采用相似方法进行处理，但由于我们所处理的核苷酸序列比干扰素编码基因的核苷酸序列大得多，所以我们对于12357个长核苷酸序列采用分段处理方式，解决了因容量不足而引起的困难。为了提高计算机检测酶切位点的速度，还作了特殊处理。我们编制的程序可以处理无论多长的序列的编码基因，且在较短时间内迅速而正确地完成71种内切酶位点的测定。若不用电子计算机，以71种内切酶对这样多的核苷酸进行酶切位点的分析，工作量是难以想像的，因此使用计算机技术发展分子生物学的前景是乐观的。

我们用来建立BVDV基因库及cDNA探针的出发毒株是BVDV InL株，该株与Collett等人研究的毒株BVDV NADL株在抗原上是相似的[2]，所以我们认为以电子计算机对BVDV NADL株的核苷酸序列酶切位点的分析结果对我们正在进行的研究是有参考价值的。由于这样，我们采用Pst I酶以及选用含有Pst I酶切位点的pUC8质粒与BVDV核酸的双链cDNA进行重组并进行克隆，建立了该病毒株的cDNA基因库。

参考文献

〔1〕 Marca Collett等，1988，Virology，165：191—199.

〔2〕 糜克永等，1989，病毒学杂志，1：69—75。

〔3〕 王蕾等，1989，内蒙古畜牧科学，2：31—33。

〔4〕 Marca Collett等，1988，Virology，165：200—208.

〔5〕 谭浩强，1986，FORTRAN语言程序设计，高等教育出版社。

〔6〕 夏志清，1984，生物化学与生物物理进展，6：83。

〔7〕 杨学楼等，1986，科学进展，4：75—76。

Applicaion of the Computer to Determine Sites of Most Restriction Enzymes Used to Assay the Genome of Bovine Viral Diarrhea Virus

Lu Jia-ti　Wan wei　　Mi Ke-yong

(Wuhan College of Textile industry, Wuhan, 430073)　　*(Wuhan Institute of Virology, Academia Sinica, Wuhan, 430071)*

Using computer to determine sites of 71 species of restriction enzymes in the sequence of cDNA from Bovine viral diarrhea virus NADL genome, we get the results indicating numbers of nucleotides sequences of 58 species. The others 13 species have no sites of restriction enzymes in the nucleotide sequences. These results are of help to study gene labraries of cDNA from BVDV InL.

Key words: Bovine viral diarrhea virus　Restriction enzyme　Computer