

文章编号: 1000-0615(2008)02-0190-10

异育银鲫 Toll 样受体 3 的 cDNA 序列分析与蛋白质高级结构预测

钱 曦, 黄旭雄, 华雪铭, 冷向军, 周洪琪

(上海水产大学省部共建水产种质资源发掘与利用教育部重点开放实验室, 上海 200090)

摘要: 利用 RT-PCR、RACE 及克隆等方法获得了异育银鲫 Toll 样受体 3(cagTLR3) 基因全长 cDNA 序列。该 cDNA 全长 3170 bp, 5' 端非编码区 181 bp; 3' 端非编码区 283 bp; 开放阅读框 (ORF) 2706 bp, 编码 901 个氨基酸。经序列同源性分析显示, 其编码的氨基酸序列与金鱼、斑马鱼、斑点叉尾鲷、虹鳟和红鳍东方鲀有较高的相似性, 其相似性分别为 92%、88%、79%、75% 和 72%, 说明 TLR3 在长期的进化中具有较高保守性。TLR3 全长序列起始 21 个氨基酸残基为信号肽; 50 至 695 个氨基酸残基间包含 15 个富含亮氨酸的重复序列 (LRR); 704 至 726 个氨基酸残基为跨膜区 (TM); 754 至 897 个氨基酸残基为与白介素-1 受体高度同源的区域 (TIR)。将异育银鲫 TLR3 与人和斑马鱼胞内 TIR 结构的氨基酸比对后发现, Phe728、Leu738、Tyr756 与 Arg736 在 TLR3 的信号转导及胞内定位过程中起到关键作用。异育银鲫 TLR3 胞外配体结合域的三维结构呈 U 型, 与人 TLR3 的相应结构十分相似, 在 N 端具有 1 个二硫键, C 端有 2 个。因 LRRs 上的插入序列和 14 个糖基化位点, 异育银鲫 TLR3 胞外配体结合域内表面的大量 β -折叠结构呈平坦的平面。

关键词: 异育银鲫; Toll 样受体 3; cDNA 末端快速扩增; 序列分析; 蛋白结构预测; 生物信息学
中图分类号: Q 517; S 917 **文献标识码:** A

Toll 样受体 (Toll-like receptors, TLRs) 是参与固有免疫的细胞跨膜受体及病原模式识别受体, 是一个广泛存在于昆虫、脊椎动物和植物中的进化上高度保守而古老的蛋白质家族。TLRs 作为参与固有免疫的模式识别受体 (pattern recognition receptors, PRR), 能够识别某种高度保守的结构基序, 即病原相关的分子模式 (pathogen-associated molecular patterns, PAMP)。TLRs 家族在固有免疫中发挥核心作用, 在诱导获得性免疫中也具有重要作用^[1]。TLRs 均属于 I 型跨膜糖蛋白 (type I integral membrane glycoprotein), 由胞外区、跨膜段和胞内区 3 部分组成, 其胞外 N 端氨基酸序列含有数目不等且富含亮氨酸的重复序列 (leucine-rich repeat, LRR), 每个 LRR 由 18~32 个氨基酸残基组成, 含有 XXLXLXX 结构域及其它保守的亮

氨酸, 这些 LRR 与 PAMP 的识别有直接的联系^[2]; 胞内区 C 端包含一个与白介素-1 受体 (IL-1R) 高度同源的结构域, 即 Toll/白介素-1 受体结构域 (Toll/interleukin-1 receptor, TIR), 约由 200 个氨基酸构成, 在 TLRs 信号转导中起重要作用^[3-4]。

在哺乳动物中, 许多病毒感染细胞后, 在复制或转录的过程中就启动了双链 RNA (dsRNA) 的合成, dsRNA 作为一种 PAMP 能特异性地被 TLR3 识别^[5]。TLR3 的激活能引起细胞因子的分泌—尤其是 I 型干扰素, 并最终导致细胞的抗病毒应答^[6]。

迄今为止, 国内外已有学者对人 (*Homo sapiens*)^[7]、小鼠 (*Mus musculus*)^[8]、斑马鱼 (*Danio rerio*)^[9]、虹鳟 (*Oncorhynchus mykiss*)^[10]、斑点叉尾

收稿日期: 2007-07-06

资助项目: 上海市科委重点科技攻关计划 (073205111); 上海市重点学科建设项目 (Y1101)

作者简介: 钱 曦 (1982-), 男, 江苏溧阳人, 硕士研究生, 从事水产动物营养与饲料学研究。E-mail: qxkillgre@yahoo.com.cn

通讯作者: 周洪琪, E-mail: hqzhou@shfu.edu.cn

鲮(*Ictalurus punctatus*)^[11]等的 TLR3 基因进行研究,但有关异育银鲫(*Carassius auratus gibelio*) TLR3 基因(cagTLR3)分子水平的研究在国内外均未见报道,GenBank 中也未见 cagTLR3 的核酸序列。因此本文运用 RACE(rapid amplification of cDNA ends, cDNA 末端快速扩增)法分离并克隆了 cagTLR3 基因全长序列,并利用生物信息学方法对该基因序列的结构进行比较分析并预测了 cagTLR3 蛋白质胞外配体结合域的高级结构,为 TLR3 的分子进化和同源比较研究提供新的资料;也为研究异育银鲫免疫反应的机理奠定基础。

1 材料与方法

1.1 材料和仪器

Trizol[®] Reagent 购自 Invitrogen[™] life technology 公司;Takara RT PCR[™] Kit (AMV) Ver1.6、pMD19-T 购自宝生生物工程有限公司;琼脂糖凝胶 DNA 回收试剂盒、DNA marker、*E. coli* (DH5 α)购自天根生物有限公司;BD SMART[™] RACE cDNA Amplification Kit 购于 BD Biosciences 公司。电泳仪为 BG-Power600i;采用生物分光光度计 BioPhotometer (eppendorf)检测核酸浓度;PCR 仪为 Mastercycler[®] ep PCR(eppendorf)。

1.2 方法

总 RNA 的提取 实验室活体解剖健康的异育银鲫,取出鳃片,液氮研磨后取 50~100 mg 于 1.5 mL 除 RNase 的离心管中,按照购自 Invitrogen 的 Trizol[®] Reagent 的说明书提取总 RNA,然后通过琼脂糖凝胶电泳溴化乙锭(EB)染色显示 28 s 和 18 s,检测 RNA 的完整性;使用生物分光光度计测定 RNA 溶度,并根据 OD_{260/280} 判断 RNA 的质量,以 1.8~2.0 为好。

引物设计与合成 根据 GeneBank 中已报道的鱼类 TLR3 cDNA 开放阅读框(ORF)保守区域,运用 Primer Primer 5 及 BLASTn 程序设计并由上海生工合成以下一组扩增部分序列的引物,引物序列如下:

P1 上游:5'-TTTGACTGCACCTGCGAGAGC-3'

P2 下游:5'-AGCAAGCTTCGTTCCACCCA-3'

然后以扩增出的 cagTLR3 基因的部分序列为基础,分别设计了扩增 cagTLR3 5'端和 3'端的基因特异性引物 P3、P4,引物序列如下:

P3:5'-CATCCCTGGAAGTGCACCAGGAAAGC-3'

P4:5'-GCTTTCCTGGTGCACCTCCAGGGATG-3'

RT 和 RACE 扩增 (1)RT 及 3'RACE。RT 及 RACE 反应按照 BD SMART[™] RACE cDNA Amplification Kit 的说明书进行。取 1 μ g 总 RNA 为模板,采用反转录试剂盒内引物 3'CDS primer 和反转录酶 BD Powerscript Reverse transcriptase 在 10 μ L 的反应体系中合成 3'RACE cDNA。RT 反应产物用 100 μ L Tricine-EDTA buffer 稀释后,5 μ L 稀释液为模板,采用聚合酶 50 $\times$ BD Advantage 2 Polymerase Mix,通用引物 UPM 和 3'端基因特异性引物 P3 进行 3'端扩增。反应条件:94 $^{\circ}$ C 30 s,72 $^{\circ}$ C 3 min,5 个循环;94 $^{\circ}$ C 30 s,70 $^{\circ}$ C 30 s,72 $^{\circ}$ C 3 min,5 个循环;94 $^{\circ}$ C 30s,68 $^{\circ}$ C 30 s,72 $^{\circ}$ C 3 min,30 个循环;最后,4 $^{\circ}$ C 保存。

(2)RT 及 5'RACE。操作流程基本同 3'RACE。不同之处在于以 5'CDS-Primer 及 BD SMART II A oligo 为反转录引物,采用 5'RACE cDNA 为模板,通用引物 UPM 和 5'端基因特异性引物 P4 进行扩增。

RACE 产物的纯化、克隆及测定分析

RACE 产物经 1.2% 琼脂糖凝胶电泳检测后,用琼脂糖凝胶 DNA 回收试剂盒对 PCR 产物进行回收、纯化。并将纯化后的 RACE PCR 产物与 pMD19-T 载体连接构建重组质粒,转化感受态大肠杆菌(*E. coli*) DH5 α ,经 LB 平板(含 Amp⁺、IPTG 和 X-gal)培养后,筛选重组子,并进行插入片段分析。序列测定由上海生工生物科技有限公司完成。所得序列用 DNASTar 分析软件进行分析,用 Edit Seq 推测 ORF 及氨基酸序列,用 MegAlign 对推导的氨基酸序列进行分析;并用蛋白分析程序组 EXPASY([Http://www.expasy.ch](http://www.expasy.ch))预测蛋白特性、信号肽和结构功能域。

cagTLR3 蛋白胞外配体结合域高级结构的预测 将异育银鲫 TLR3 蛋白序列提交瑞士生物信息研究所(SIB)维护的 SWISS-MODEL 服务器(<http://swissmodel.expasy.org/>),利用同源建模(homology modeling)的方法^[12-14]预测异育银鲫 TLR3 蛋白胞外配体结合域的三维(three-dimensional,3D)高级结构。人 TLR3 蛋白胞外配体结合域(PDB 2aOzA)的三维结构来自 ExPDB 数据库(<http://www.rcsb.org/pdb/>),使用 Swiss-pdb Viewer 软件将上面预测的异育银鲫 TLR3 蛋白胞外结构和数据库中人 TLR3 相应结构进行叠合比

较,根据人 TLR3 中结构半胱氨酸以及插入序列的空间位置,找到异育银鲫 TLR3 相应的氨基酸残基,从而进行测量和比较。并根据 Prosit 数据库 (<http://cn.expasy.org/prosit/>) 提供的由天冬酰胺形成的 N-联接糖基化位点 (N-glycosylation site) 的氨基酸基序 (Consensus pattern, PDOC00001), N-{P}-[ST]-{P}, 利用 pdb Viewer 自带的查找工具预测异育银鲫 TLR3 胞外配体结合域中潜在的 N-链接糖基化位点。

TLR3 分子进化分析 将获得的异育银鲫 TLR3 蛋白序列和 GenBank 中其它多个物种的 TLR3 氨基酸序列进行 ClustalW (<http://www.ebi.ac.uk/clustalw/>) 比对,然后用 MEGA 3.1 程序作出 TLR3 的分子系统进化树。

2 结果

2.1 部分序列及 RACE 扩增结果

cagTLR3 部分序列扩增测序结果 用引物 P1、P2 进行部分序列扩增,产物电泳后获得条明带(图 1),其长度 389 bp,与预期长度大小相同。将异育银鲫部分序列提交 blastn (<http://www.ncbi.nlm.nih.gov/blast/>),发现其与金鱼、斑马鱼、虹鳟、斑点叉尾鲶和红鳍东方鲷的相似性分别为:91%、85%、69%、66% 和 68%。同源性很高,故推测是异育银鲫 TLR3 的部分序列。

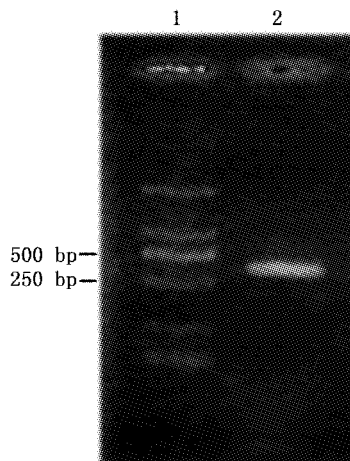


图 1 部分序列扩增结果

Fig.1 Partial PCR product of TLR3

1. 100 bp 分子量标记;2. 部分序列扩增结果

1. 100 bp molecular marker;2. partial product

3'RACE 及 5'RACE 扩增测序结果 以特

定引物 UPM 和基因特异性引物 P4 进行 3'RACE 扩增产物只有一条明带,序列长度为 838 bp(图 2);同样以特定引物 UPM 和基因特异性引物 P3 进行 5'RACE 扩增产物也只有一条明带,长度为 2358 bp(图 3)。

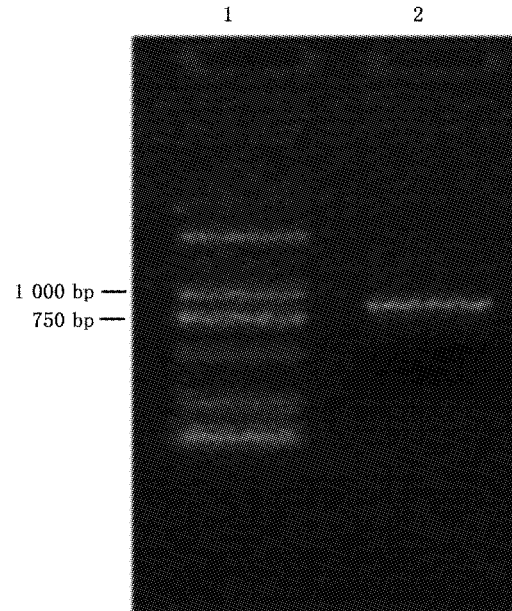


图 2 3'RACE 扩增结果

Fig.2 3'RACE of TLR3

1. 100 bp 分子量标记;2. 3'RACE 扩增结果

1. 100 bp molecular marker;2. 3'RACE product

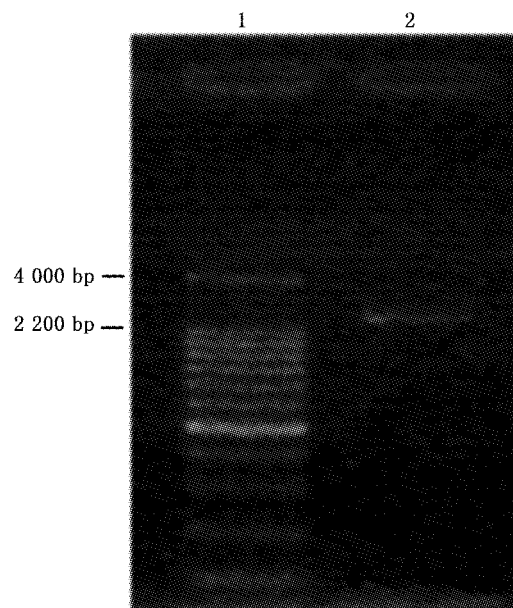


图 3 5'RACE 扩增结果

Fig.3 5'RACE of TLR3

1. 100 bp 分子量标记;2. 5'RACE 扩增结果

1. 100 bp molecular marker;2. 5'RACE product

2.2 序列分析

根据 5'RACE 以及 3'RACE 的测序结果,运用 DNASTar 分析软件中的 EditSeq 工具拼接得到 cagTLR3 基因 cDNA 全长序列(图 4)。序列分析显示,cagTLR3 全长为 3170 bp,包括 181 bp 的 5'端非翻译区(5'UTR);283 bp 的 3'端非翻译区(3'UTR);2706 bp 的开放阅读框(ORF),其编码 901 个氨基酸。

利用在线分析软件组 EXPASY 对所推导的

氨基酸序列进行蛋白质性质与结构的分析,cagTLR3 氨基酸序列的分子量为 102 019.8 u,等电点为 8.06,其蛋白质的不稳定系数为 41.98,可认为是不稳定的蛋白质;其二级结构组成为 α -螺旋 (Alpha helix) 44.17%,延伸主链 (extended strand) 11.88%,随机卷曲 (random coil) 43.95%。

使用 SMART 软件 (<http://cn.expasy.org>) 对 cagTLR3 进行蛋白质结构分析,在其 N 末端具有一信号肽序列 MELMKLILLPFFCMCFHGHTA,

```

-181
tgaaggttattttccctcatgcaaagattatatagcaatgcaagagctctgtctcga
acgcggggag
gttgcttcagacttgaaccactttaaccattcagatatattgagctgaacactgggaa
1
ATG GAA CTG ATG AAA CTC ATA TTG TTG CCC TTT TTC TGC ATG TGT
1
M E L M K L I L L P F F C M C
46
TTT CAT GGA CAC ACT GCA TAT ACA CAC AAG TCA ACA TGT ACG ATT
16
F H G H T A Y T H K S T C T I
91
AAG AAT GCC AAA GCA GAC TGC AGT CAT ATG AAT CTG GAT GCA GTT
31
K N A K A D C S H M N L D A V
136
CCA ACG GAC CTA CCA AAA AAT ATC ACA ACA TTG GAC GTG TCT CAC
46
P T D L P K N I T T L D V S H
181
AAT AGA CTT AAA ACT CTG TCT TCT CTG CAT TTG TAC TCG AAT CTG
61
N R L K T L S S L H L Y S N L
226
GTG ACT ATA GAC GCC AGC TAC AAC TCT TTA ACT GCG ATA GAG AAG
76
V T I D A S Y N S L T A T E K
271
GAT CTA TGT CTT TCT GTT CCA CAC CTG CAA ATT CTC AAT GTG CAA
91
D L C L S V P H L O I L N V O
316
CAC AAT GAA GTG TAT TTG ATA AGT GAG AAA GAC CTG AAA AAC TGT
106
H N E V Y L I S E K D L K N C
361
TTT CAT TTA ACA AGA CTT GAC CTG TCT GAC AAC AGA CTG AAG CTA
121
F H L T R L D L S D N R L K L
406
AAT GGA GAG CCG TTC TCT GTT CTA AAG AGT TTG ACA TGG TTG GAT
136
N G E P F S V L K S L T W L D
451
GTA TCT CGA AAC AAA CTG AAA TCA GCG AAA CTG GCG ACA CAC CCT
151
V S R N K L K S A K L G T H P
496
CAG CTG CCA AAC CTA ATC ACC CTC GTT CTC TCT GGA AAT CAA ATT
166
Q L P N L I T L V L S G N Q I
541
TCT GTG CTG CAA AAG AAT GAC TTC TCA TTC CTC AGT AAT TCT TCT
181
S V L Q K N D F S E L S N S S
586
GCA TTT CGG GTT CTG ATA CTC TCG TCT CTG TCT CTT AAA AAG GTA
196
A F R V L I L S S L S L K K V
631
GAG AGT GGC TGT TTC CAG GCT ATT GCT AGA CTT TCT GAC TTA GTC
211
E S G C F Q A I A R L S D L V
676
CTG GAT TAC TGC AAG GTC AGC CCT CAG TTG ACC ACC AGT CTT TAT
226
L D Y C K V S P Q L T T S L Y
721
GAA GAA CTT GCT GGC ACA GCT TTG AGA AAC CTT TCC CTT AAG AAC
241
E E L A G T A L R N L S L K N
766
ATT CAA CAG GTG ACT CTC TCA AAC ACA ACT TTC CAA GGT CTA GAG
256
I Q Q V T L S N T T F Q G L E
811
AAG ACT AAC ATC ACA CTG CTC GAC CTC AGC AGC AAC ACA ATG TCA
811
K T N I T V L D L S S N T M S
856
AAG ATT GCA GAA GAA GCC TTT CAG TGG TTT CCC CGA CTG GAA ATT
286
K I A E E A F Q W F P R L E I
901
TTA TCC CTG GAG CAT AAC ACT CTT AGA CAC CTA ACT AAT GGC ACT
301
L S L E H N T L R H L T N G T
946
TTC CGT GGA TTG GGA AAC CTG AGG CAA CTT AAC CTG CAG AAA GCA
316
F R G L G N L R Q L N L Q K A
991
CTG ATT AAG AGT CAT ACA TCA TCA TTA CCG AAT ATT GAA GAC TTC
331
L I K S H T S S L P N I E D F
1036
TCT TTC CAC CAT TTA GTC CAA TTA GAG CAT CTA TGC ATG GCA AAT
346
S F H L V Q L E H L C M A N
1081
ACT GCG TTC CGA GGG ATA ACA GGG CAT ATA TTT TCT GGA CTT CTG
361
T A E R G I T G H I F S G L L
1126
AAC CTA AAG ACA CTG GAT TTG AGT TGG AGC ATC ACA GGG TTG AAG
376
N L K T L D L S W S I T G L K
1171
ATG GTC ACA AAC AGA ACA TTT GCT TCT TTG CAA GAA TCT CCA CTC
391
M V T N R T F A S L Q E S P L
1216
CTT CAG ACT CTT AAT CTT ACT GCC ATG GGT ATC AAC AAG TTG GCA
406
L Q T L N L T A M G I N K L A

```

```

1261 CCT GGG GCC TTT TCA AGT TTG GGC AAC CTC ACT ACA CTC CTG CTT
421 P G A F S S L G N L T T L L L
1306 AGT CAC AAT TTC ATT TAT CAG CAA TTG AAA GGA GAT GAG TTC GAG
436 S H N F I V Q O L K G D E F F
1351 GGC CTC TCT AAC ATT AAA CAG ATT GAC TTA TCT ATG AAC CAG CAA
451 G L S N I K Q I D L S M N Q Q
1396 AGC ATT TCC CTT ACC AAT ACA TCA TTC ATC CAC GTC CCA ACA TTA
466 S I S L T N T S F I H V P T L
1441 AGG AGT CTA AAG CTC GGC CGT GCC CTA AAA GGG ACT TTA GCT ATG
481 R S L K L G R A L K G T L A M
1486 GAA CCA TCT CCA TTC AAG CCA CTT ATC AAC CTC ACA TTT CTA GAT
496 E P S P F K P L I N L T F L D
1531 CTC AGT AAT AAC AAT ATT GCA AAC ATA AAT GCT TGC TTG CTG AAA
511 L S N N N I A N I N A C L L K

1576 GGA CTC AAC CAT CTG AAA GTG CTG AAA ATG CAG CAC AAC AAC TTG
526 G L N H L K V L K M Q H N N L

1621 GCT AGG TTG TGG AAG ACA GCC AAC CCT GGT GGT CCA GTG TTG TTT
541 A R L W K T A A N P G G P V L F
1666 CTC CAG GAT GCC ACA AAA CTG TCT GTC CTG GAT CTG GAT TAC AAT
556 L Q D A T K L S V L D L D Y N
1711 GGT TTA GAT GAG ATT CCA CTC AAT GCT TTG CGC GGC CTC TCA GAG
571 G L D E I P L N A L B G L S E
1756 TTA CGT GAG CTA AGT CTC CAT GGT AAT CTT TTG GAT CAG CTC CAT
586 L R E L S L H G N L L D Q L H
1801 GCC TCT GTT TTT GAT GAC CTA CAG TCT TTA AAG TAC TTG TAT CTT
601 A S V F D D L Q S L K Y L
1846 CAA AAG AAC CTT ATA ACC TCT GTC CAA CGA GTC ACT TTT GGT GTG
616 Q K N L I T S V Q R V T F G V
1891 CCG CTG TCC AAC CTG ACA GAA CTT TAC ATG GAC CAC AAC CCC TTT
631 P L S N L T E L Y M D H N P F
1936 GAC TGC ACC TGC GAG AGC ATC CTG TGG TTC TCT GAG TGG CTT AAC
646 D C T C E S I L W F S E W L N
1981 TCT ACC AGT GCC AGC ATT CCT GGA TTC CCT CAA AGT TAC ATC TGC
661 S T S A S I P G F P Q S Y I C
2026 AAC ACC CCA AAT GCC TAC TTT AAC CAC TCT GTC ATG AAA TTT GAC
676 N T P N A Y F N H S V M K E D
2071 CCA TTG TCC TGC AAG GAT ATG ACA CCT TTT AAG GCA CTG TAC ATT
691 P L S C K D M T P F K A L Y I
2116 TTG AGT AGC ACA GCA GTG TTG ATG TTG TTA TTT GCA GCT TTC CTG
706 L S S T A V L M L F A A L
2161 GTG CAC TTC CAG GGA TGG AGA ATC CAG TTC TTC TGG AAC ATA ATG
721 V H F O G W R I Q F F W N I M
2206 GTA AAC CGT ATG CTG GCG TCA CTG GAG GAT GAA AGC ATT ACT GAA
736 V N R M L A S L E D E S I T E
2251 GGT AGA TAT GCA TAT AAT GCA TAT ATC ATC CAC AGT GCT GAG GAC
751 G R Y A Y N A Y I I H S A E D
2296 AGG CCA TGG GTG GAA CGA AGC TTG CTC CCC TTA GAG GAT GAA AAA
766 R P W V E R S L L P L E D E K
2341 TTC CAT TTT TTT CTT GAA GAC AGA GAT GCA GTA CCT GGC TTT TCT
781 F H F F L E D R D A V P G F S
2386 CAA CTT GAC ACC ATT GTT GAA AAC ATG GGA CAG TCT AGG AAA ATC
796 Q L D T I V E N M G Q S R K I
2431 CTT TTT GTT ATT ACA GAA AAG CTT CTA GAG GAT CCA TGG TGT CGG
811 L F V I T E K L L E D P W C R
2476 CAA TTT AAA GCC CAT CAC GCA CTT CAC CAG GTA ATG GAG GAC AAT
826 Q F K A H H A L H Q V M E D N
2521 CGT GAC TCG CTG ATT CTG ATT TTT TTG CAA GAT GCA ACT GAT TAC
841 R D S L I L I F L O D A T D Y
2566 AGT CTG AAC CGC TCT CTG CAT CTT CGC CGC GGC ATG TTG AAA TCT
856 S L N R S L H L R R G M L K S
2611 CAT TGT ATC CTC CAC TGG CCT TTA CAC AGG GAA CGA ATC CCA GCT
871 H C I L H W P L H R E R I P A
2656 TTC CAT CAG AAA CTC CGC TCA GCA TTA GCC TCT ACC AAC AAG GTC
886 F H Q K L R S A L A S T N K V
2701 AAC TAA
901 N *

gtaagtaaaattgtctttactatctataactgattcaaaaaggcttttctacaattac
atthtagtgaccattttataaaatgcaaaccactaacaagtttaggttctttaagggtg
agatttcagatcctcaagaagcacctcacatattcaccaaacgcgatgcttccttttc
tgccagttactgaaatgcacaggtgttttagcaatttcaagttcaattaagaagtg
ttacactctctctgccttccactctgcagaaaaaa

```

图4 cagTLR3 cDNA 及推导出的氨基酸序列

Fig.4 The cDNA and deduced amino acid sequence of *Carassius auratus gibelio* TLR3

小写字母代表 5',3'端非翻译区;大写字母部分为编码区,上面为核苷酸序列,下面为氨基酸序列;LRR 区用灰色阴影表示, TM 区用粗斜体表示, TIR 区使用粗体表示

5'-,3'- untranslated regions are shown as lowercases. Coding region is shown as uppercases, where the upper sequence indicates the nucleotides and the lower shows the amino acids; Individual LRRs are represented in alternating gray, TM is printed in bold and italic at the same time, and TIR is represented with bold

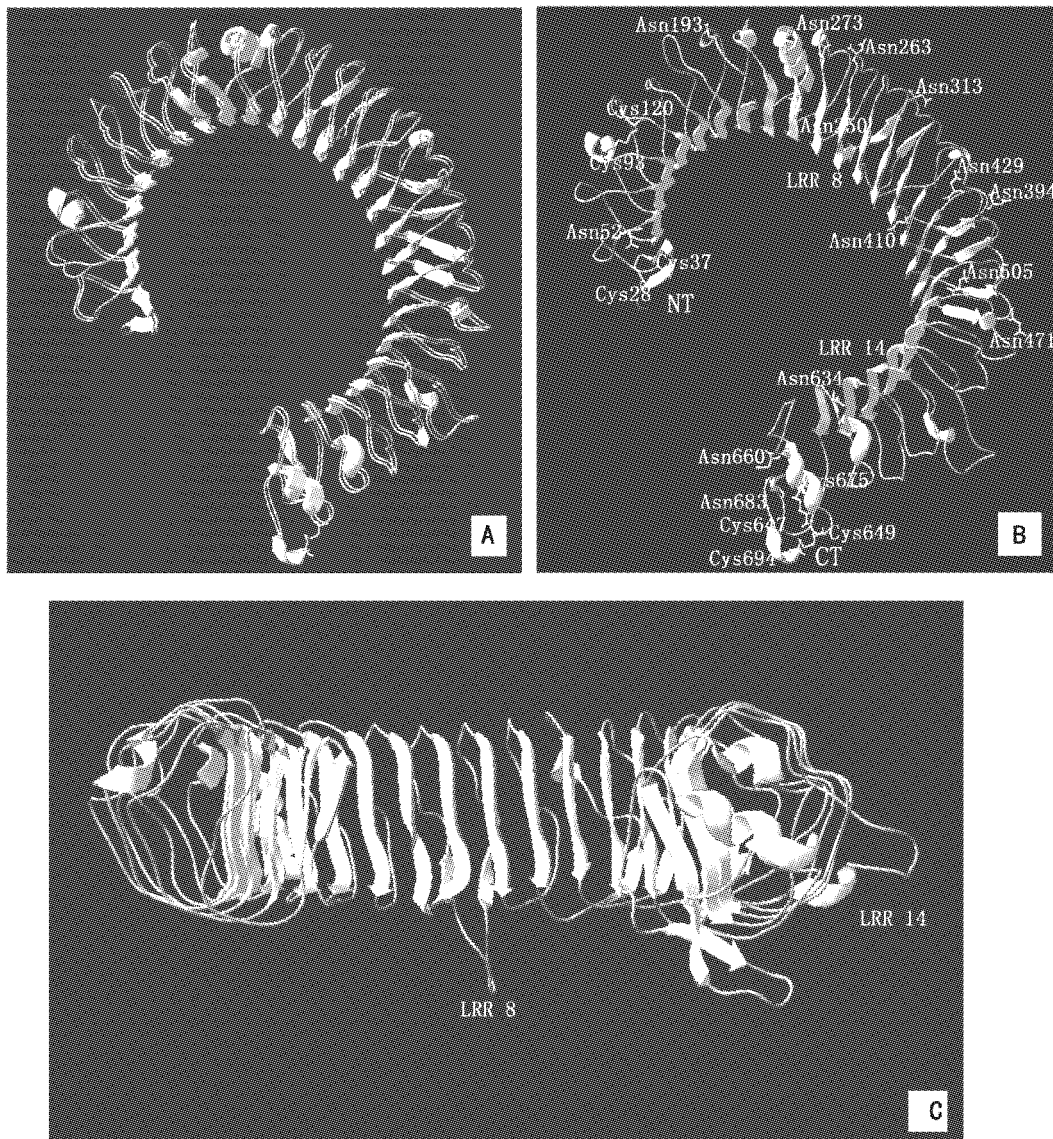


图6 *cag* TLR3 胞外配体结合域的三维结构模型

Fig.6 Three-dimensional structure models of *Carassius auratus gibelio* TLR3 ligand-binding domain

蛋白质三维结构利用比较蛋白建模程序构建;(A)*cag*TLR3 胞外配体结合域和相应人(PDB 2aOzA)的 X-射线衍射所确定三维结构模型的叠合;(B)*cag* TLR3 胞外配体结合域的 4 个二硫键和 14 个 N-连接的糖基化位点;(C)异育银鲫 TLR3 胞外配体结合域的侧面图和 LRR 中的两个插入结构

the 3-D protein models were constructed by comparative protein modeling program SWISS-MODEL. (A) Three-dimensional model of TLR3 ligand-binding domain from *Carassius auratus gibelio* (with green helix) superposed with X-ray structure of human ligand-binding domain (with red helix); (B) indicate the 3-D model of *Carassius auratus gibelio* TLR3 ligand-binding domain including 4 disulfide bonds and 14 N-linked glycosylation sites, respectively; (C) Side view of *Carassius auratus gibelio* TLR3 ligand-binding domain shows the two insertions at LRRs 8 and 14, respectively

外径为 90 Å,平均厚度为 35 Å(图 6-A,B)。该结构所分析的氨基酸序列只包括胞外的成熟肽部分(26 位到 694 位),不含有 N 端的信号肽、膜内和胞内部分。

在 *cag*TLR3 胞外配体结合域肽链的 N 端有 LRR-N-末端结构(LRR-NT),空间结构呈发夹环状(hairpin loop),并被 Cys28 和 Cys37 之间形成的二硫键所固定(图 6-B)。紧接 LRR-N-末端结构后

是一系列平行的 β -折叠结构,这些 β -折叠构成 TLR3 胞外配体结合域凹陷的内表面,15 个 LRR 区即包含在其中,其中每一条 β -链和螺线管的中轴线基本垂直,相邻的 β -链通过不规则环状结构相连(图 6-B)。LRR 8 和 LRR 14 含有两段插入结构(图 6-B,C)。在 LRR 区之后是胞外肽链 C 端基序(LRR-CT),其被 Cys647 和 Cys675 与 Cys649 和 Cys694 所形成的两个二硫键所固定,呈现紧密的空间结构(图 6-B)。除上述三个二硫键,还有一个存在于 Cys93 和 Cys120 之间。

cagTLR3 胞外配体结合域表面含有丰富的分布不均匀的 Asn 通过 N-连接的 14 个糖基化位点(图 6-B)。cagTLR3 胞外配体结合域内表面,尤其靠近 N 端和 C 端的内表面,糖基化位点较少。

2.5 TLR3 分子进化分析

经 Clustal W 比对后,采用 MEGA 3.1 程序中的非加权配对算术平均法(UPGMA),重复 1000 次,gap 处理为缺失,构建了多物种 TLR3 的系统进化树(图 7),上标数据用 Bootstrap 计算每个分支的支持度。图中直观地表示了 cagTLR3 与各种动物 TLR3 氨基酸序列的相似程度。结果表明,同属鲤科的异育银鲫与鲫属的金鱼 TLR3 同源性最高为平衡的内部分支,与鲟属斑马鱼同源性相对较远,与人和鼠差距最远。

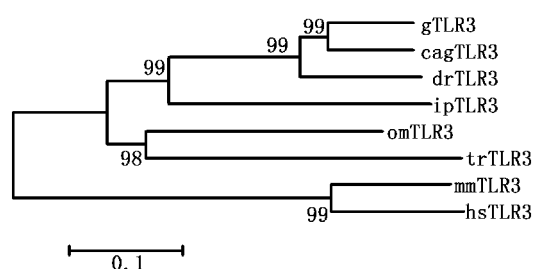


图 7 TLR3 蛋白质的系统树
(使用 Dnastar 选择最大相似方法所构建)

Fig. 7 Phylogenetic tree of TLR3 proteins

(The tree was constructed by Dnastar using
Maximum likelihood of the TLR3 proteins)

ipTLR3, omTLR3, trTLR3 分别代表斑点叉尾鲟、虹鳟和红鳍东方鲃的 TLR3 氨基酸序列

ipTLR3, omTLR3, trTLR3 represent amino acid sequence of TLR3 of *Ictalurus punctatus*, *Oncorhynchus mykiss* and *Takifugu rubripes*, respectively

3 讨论

本试验运用 RACE 以及克隆等方法快速扩增获得了 3170 bp 的 cagTLR3 基因 cDNA 全长序列。对其氨基酸序列的同源性比较后发现, cagTLR3 与金鱼同源性最高,与斑马鱼、虹鳟、红鳍东方鲃(*Takifugu rubripes*)和斑点叉尾鲟次之,与人和鼠最远。这说明在长期的生物进化过程中 TLR3 具有较强保守性,同时说明了他们的分类地位。

异育银鲫的 3'UTR 比较短,虽然尚无法确定 3'UTR 的大小对 TLR3 功能是否有影响,但根据已有的资料表明,3'UTR 的大小和序列组成对 mRNA 稳定性、转录调控和蛋白定位都有影响^[15-17]。不稳定的 mRNAs 3'UTR 通常含有 AU 富含基序(AU-rich elements, AREs),这种情况尤其发生在那些与炎症反应和免疫调节有关的基因转录形成的 mRNA 中^[18],如 TNF、IL-1、IL-2、IL-4、IL-6 和 IFN 等基因的 mRNA 中均含有该种结构^[19-20]。ARE 结构中最典型的基序为 ATTAA,但 ARE 结构的长度、AU 的含量、典型基序的拷贝数及其组成均是可变的^[18],如本研究中异育银鲫 3'UTR 中含有 2 个 ATTAA 基序,而虹鳟有 8 个该种基序、人只有 3 个、斑马鱼中没有^[10]。该结果表明, cagTLR3 mRNA 比人和虹鳟更稳定,但不及斑马鱼。此外,根据 SMART 软件在线分析软件得知: cagTLR3 的胞外结构域具有 15 个 LRR,斑马鱼(NP-001013287)也具有 15 个,人(NP-003256)有 18 个,这些结果相似或接近。但这与 Bell 等^[2]报道的 hsTLR3 具有 23 个 LRR 不符,或许是使用不同分析软件所引起的。

TIR 结构域中某些特定的氨基酸在受体信号转导和胞内定位方面发挥重要作用, Funami 等^[3]报道 hsTLR3 的 TIR 结构域中 Phe732 和 Leu742 氨基酸残基在激活 NF- κ B 和 IFN- β 转录启动子的信号通路中发挥重要作用, cagTLR3 的 TIR 结构域中也拥有相应的 Phe728 和 Leu738 氨基酸残基(图 5)。Sarkar 等^[4]指出, hsTLR3 的 TIR 结构域中 Tyr759 与 hsTLR3 的胞内信号转导有密切关系,经过氨基酸比对发现, cagTLR3 中同样存在相对应的 Tyr756。hsTLR3 的胞内定位涉及到 Arg740 和 Val741 两个关键的氨基酸残基^[3],其中 Arg 在 cagTLR3 中同样保守,即 Arg736,但是 Val 在异育银鲫和斑马鱼的 TLR3 中都表现为缺失。

因此, cagTLR3 TIR 结构域中上述氨基酸残基是否亦具有相同的作用还有待于进一步研究。

cagTLR3 胞外配体结合域的三维结构呈弯曲的螺旋管状, 其分子内表面 LRR 区富含疏水氨基酸残基, 从而形成一个疏水核心^[21]。cagTLR3 胞外配体结合域不像其它富含 LRR 的蛋白, 如内化素 A(PDB 1O6S)和糖蛋白 IB α (PDB 1M0Z), 在螺旋管的基础上进一步盘绕, 这可能是由于一些不规则 LRR 打断了 LRRs 的连续性所导致的。正是因为这一特点, cagTLR3 胞外配体结合域与人相应结构一样形成了平坦的内表面, 从而有利于 TLR3 的寡聚化和信号转导^[18]。哺乳动物(人、小鼠、大鼠和牛)都在相应的 LRR 区含有两个保守的插入序列, 该结构可能和配体的结合有关^[22], 本文首次发现位于 cagTLR3 胞外配体结合域的 LRR8 和 LRR14 上也具有的这两个插入序列。在异育银鲫 TLR3 胞外配体结合域 14 个糖基化为位点中, Asn250 和 Asn410 位于 β -折叠的 β -链上, 对应与人 Asn252 和 Asn413 两个位点, 与配体识别和结合有关^[2,23]。

参考文献:

- [1] Takeda K, Kaisho T, Akira S. Toll-like receptors[J]. *Annu Rev Immunol*, 2003, 21: 335–376.
- [2] Bell J K, Mullen G E, Leifer C A, *et al.* Lucine-rich repeats and pathogen recognition in Toll-like receptors [J]. *Trends Immunol*, 2003, 24(10): 528–533.
- [3] Funami K, Matsumoto M, Oshiumi H, *et al.* The cytoplasmic linker region in Toll-like receptor 3 controls receptor localization and signaling [J]. *Int Immunol*, 2004, 16 (8): 143–154.
- [4] Sarkar S N, Smith H L, Rowe T M. Double-stranded RNA signaling by Toll-like receptor 3 requires specific tyrosine residues in its cytoplasmic domain[J]. *J Biol Chem*, 2003, 278 (7): 4393–4396.
- [5] Alexopoulou L, Holt A C, Medzhitov R, *et al.* Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3 [J]. *Nature*, 2001, 413 (18): 732–738.
- [6] Hoebe K, Du X, Georgel P, *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signaling[J]. *Nature*, 2003, 424: 743–748.
- [7] Sarkar S N, Elco C P, Peters K L, *et al.* Two tyrosine residues of Toll-like receptor 3 trigger different steps of NF- κ B activation[J]. *J Biol Chem*, 2007, 282 (6), 3423–3427.
- [8] Carpentier P A, Williams B R, Mille S D. Distinct roles of protein kinase R and toll-like receptor 3 in the activation of astrocytes by viral stimuli[J]. *Glia*, 2007, 55 (3): 239–252.
- [9] Meijer A H, Gabby Krens S F, Medina Rodriguez I A, *et al.* Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish [J]. *Mol Immunol*, 2004, 40: 773–783.
- [10] Rodriguez M F, Wiens G D, Purcell M K. Characterization of Toll-like receptor 3 gene in rainbow trout (*Oncorhynchus mykiss*) [J]. *Immunogenetics*, 2005, 57: 510–519.
- [11] Puttharat B E, Peatman B S, Liu Z J. Toll-like receptor 3 and TICAM genes in catfish: species-specific expression profiles following infection with *Edwardsiella ictaluri* [J]. *Immunogenetics*, 2006, 58: 817–830.
- [12] Schwede T, Kopp J, Guex N, *et al.* SWISS-MODEL: an automated protein homology-modeling server [J]. *Nucleic Acids Research*, 2003, 31: 3381–3385.
- [13] Guex N, Peitsch M C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling[J]. *Electrophoresis*, 1997, 18: 2714–2723.
- [14] Arnold K, Bordoli L, Kopp J, *et al.* The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling [J]. *Bioinformatics*, 2006, 22, 195–201.
- [15] Mazumder B, Seshadri V, Fox P L. Translational control by the 3'-UTR: the ends specify the means[J]. *Trends Biochem Sci*, 2003, 28 (2): 91–98.
- [16] Jacobson A. Regulation of mRNA decay: decapping goes solo[J]. *Mol Cell*, 2004, 15 (1): 1–2.
- [17] Hesketh J. 3'-Untranslated regions are important in mRNA localization and translation: lessons from selenium and metallothionein[J]. *Biochem Soc Trans*, 2004, 32: 990–993.
- [18] Saklatvala J, Kracht M. Transcriptional and post-transcriptional control of gene expression in inflammation [J]. *Cytokine*, 2002, 20 (3): 91–106.
- [19] Shaw G, Kamen R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation[J]. *Cell*, 1986, 46 (5): 659–667.
- [20] Bakheet T, Frevel M, Williams B R, *et al.* ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins [J]. *Nucleic Acids Res*, 2001, 29

- (1): 246 – 254.
- [21] Kajava A V, Kobe B. Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information[J]. *Protein Sci*, 2002, 11: 1082 – 1090.
- [22] Jessica K B, Istvan B, Pamela R H, *et al*. The molecular structure of the Toll-like receptor 3 ligand-binding domain[J]. *Proceeding National Academy of Science*, 2005, 31(102): 10976 – 10980.
- [23] Kobe B, Kajava A V. The leucine-rich repeat as a protein recognition motif[J]. *Curr Opin Struct Biol*, 2001, 11 (6): 725 – 732.

cDNA sequence analysis and tertiary structure prediction of Toll-like receptor 3 gene from *Carassius auratus gibelio*

QIAN Xi, HUANG Xu-xiong, HUA Xue-ming, LENG Xiang-jun, ZHOU Hong-qi

(Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources Certificated by the Ministry of Education, Shanghai Fisheries University, Shanghai 200090, China)

Abstract: The full length cDNA of Toll-like receptor 3 were cloned from *Carassius auratus gibelio* gill by means of reverse transcription polymerase chain reaction (RT-PCR) and rapid amplification of cDNA ends (RACE). It was a good data for further studying the function of TLR3 in immunity of *Carassius auratus gibelio*. Sequence analysis revealed a 3170 bp cDNA containing 181 bp 5'-untranslated region, 283 bp 3'-untranslated region and 2706 bp open reading frame (ORF) which encoded 901 amino acids with predicted molecular weight of 10 219.8 u and theoretical isoelectric point of 8.06. The deduced amino acid sequence of *Carassius auratus gibelio* had high similarity with that of *Carassius auratus* (goldfish), *Danio rerio*, *Ictalurus punctatus*, *Oncorhynchus mykiss* and *Takifugu rubripes*. The similarities were 92%, 88%, 79%, 75% and 72%, respectively. Molecular phylogeny tree also showed that TLR3 of *Carassius auratus gibelio* shared high homology with TLR3 of *Carassius auratus gibelio* and *Danio rerio*. The full-length TLR3 of *Carassius auratus gibelio* contained a signal peptide in the first 21 amino acid residues, 15 leucine-rich regions (LRR) from 50th to 695th amino acid residue, a transmembrane region (TM) from 704th to 726th residue, a Toll/IL-1R homologous region (TIR) from 754th to 897th residue. Alignment with amino acids sequence of TIR from *Carassius auratus gibelio*, *Homo sapiens* and *Danio rerio* indicated, Phe728, Leu738, Tyr756 and Arg736 of *Carassius auratus gibelio* played an important role both in the receptor signaling and intracellular localization. The three-dimensional structure of *Carassius auratus gibelio* TLR3 ectodomain (TLR3-ECD) had high similarity with that of human TLR3-ECD and showed horseshoe-shaped solenoid structure with predicted inner diameter of 42 Å, predicted outer diameter of 90 Å and predicted thickness of 35 Å. *Carassius auratus gibelio* TLR3-ECD possessed 1 disulfide bond at N-terminal and 2 at C-terminal. The extensive β -sheet on the molecular's concave surface formed a platform several modifications, including insertions in the LRRs and 14 N-linked glycosylation sites.

Key words: *Carassius auratus gibelio*; Toll-like receptor 3 (TLR3); rapid amplification of cDNA ends (RACE); sequence analysis; protein tertiary structure; bioinformatics