

文章编号:1000-0615(2012)06-0849-10

DOI:10.3724/SP.J.1231.2012.27768

## 草鱼 I 型胶原蛋白 $\alpha 1$ 基因 cDNA 全序列克隆、组织分布及其生物信息学分析

刘邦辉<sup>1,2</sup>, 郁二蒙<sup>1</sup>, 王广军<sup>1</sup>, 余德光<sup>1</sup>, 谢 骏<sup>1\*</sup>, 王海英<sup>1</sup>, 龚望宝<sup>1</sup>

(1. 中国水产科学研究院珠江水产研究所,  
农业部热带亚热带水产资源利用与养殖重点实验室, 广东 广州 510380;  
2. 上海海洋大学水产与生命学院, 上海 201306)

摘要: 利用 PCR 和 RACE 方法首次克隆了编码草鱼肌肉 I 型胶原蛋白的  $\alpha 1$  基因(COL1A1) 的 cDNA 全长序列, 为 5 772 bp, 其开放阅读框为 4 347 bp, 编码 1 448 个氨基酸。BLAST 同源性分析结果显示, 草鱼 COL1A1 基因的氨基酸序列与斑马鱼、金鱼同源性较高, 分别为 93.90%和 93.60%, 呈现出较高的保守性。系统进化树分析表明, 该基因与斑马鱼、金鱼处于同一支, 亲缘性最近。生物信息学分析显示, 草鱼 COL1A1 蛋白质的相对分子质量为 137.2 ku, 理论等电点是 5.44, 为  $\alpha$  螺旋  $\beta$  折叠三重螺旋结构蛋白。有 18 段三重螺旋重复域, 22 段低复杂度域, 17 个功能域。COL1A1 蛋白有 33~69, 1 249~1 355 两个绑定钙的区域和 62~104 一个绑定锌的区域。利用半定量 RT-PCR 方法检测的组织表达结果表明, COL1A1 基因在草鱼肌肉、肠道、肝胰脏、鳃、皮肤、鳍条、肾脏和脾脏 8 个组织中均有表达, 其中在皮肤、鳃、肾脏、鳍条组织中的 mRNA 表达量高于其他 4 个组织( $P < 0.05$ )。

关键词: 草鱼; I 型胶原蛋白; COL1A1; 组织分布; 生物信息学

中图分类号: Q 781; S 917

文献标志码: A

I 型胶原蛋白(type I collagen)是胶原蛋白家族(collagen family)成员之一, 广泛分布于动物的肌肉、皮肤等组织中, 为多细胞动物中含量最丰富的胶原蛋白和细胞外基质主要组成部分, 具有调控胶原纤维形成、组织分化、损伤修复等多种功能<sup>[1-12]</sup>。已有部分学者研究了 I 型胶原蛋白的主要成分  $\alpha 1$  链(COL1A1)在损伤组织修复过程中的生物学功能。在损伤的组织中 COL1A1 的合成受到了抑制, 且该损伤组织的再生亦被阻止<sup>[7-8]</sup>, COL1A1 的合成能促进损伤组织血管壁的修复<sup>[9]</sup>。

草鱼(*Ctenopharyngodon idellus*) 隶属鲤形目(Cypriniformes)、鲤科(Cyprinid)、雅罗鱼亚科(Leuciscinae), 是我国重要的淡水养殖鱼类之一。在草鱼养殖或运输过程中, 常因拉网、运输等原

因造成一定的机械损伤, 从而出现相当的经济损失<sup>[13-15]</sup>。本研究以大宗养殖的淡水鲤科鱼类草鱼为研究对象, 通过 PCR 和 RACE 技术获得 COL1A1 基因的 cDNA 全长序列, 利用分子信息手段分析基因的结构特征、预测生物学功能, 利用半定量 PCR 技术测定 COL1A1 在肌肉、肝胰脏、肠道、鳃、脾脏、肾脏、皮肤和鳍条组织中的表达情况, 以期为进一步了解 COL1A1 基因在鱼类损伤组织修复中的作用提供依据。

### 1 材料与方法

#### 1.1 实验材料

2009 年 2 月在广东省中山食品水产进出口集团有限公司西穗养殖基地采集 3 尾草鱼体质量

收稿日期: 2011-10-12 修回日期: 2011-12-30

资助项目: 广东省自然科学基金自由申请项目(10151038001000004); 现代农业产业体系建设专项项目(nycytx-49); 广东省海洋渔业科技推广项目(A200901D04)

通讯作者: 谢 骏, E-mail: xiejunhy@yahoo.com.cn

(4.2±0.4) kg, 去鳞解剖后, 采集肌肉、肠道、肝胰脏、鳃、皮肤、鳍条、肾脏、脾脏 8 个组织, 立即放入液氮保存, 运至实验室, -20 °C 冻藏备用。

### 1.2 主要试剂

Trizol 总 RNA 提取试剂盒购自 Nitrogen 公司; 反转录试剂盒、pMD19-T 载体、DL2000 DNA 和 Taq DNA 聚合酶购自大连 TaKaRa 公司; SMARTTM RACE cDNA Amplification Kit 购自 Clontech 生物工程公司; Advantage-2 PCR Kit 购自 Clontech 生物工程公司; 琼脂糖凝胶 DNA 回收试剂盒和 DH5 $\alpha$  购自天根公司。

### 1.3 实验方法

**引物设计** 在 NCBI 上检索有关鱼类的 *COL1A1* cDNA 序列, 根据 GenBank 中金鱼(*Carassius auratus*, AB275454)、斑马鱼(*Brachydanio rerio*, BC161663)、虹鳟(*Salmo gairdneri*, NM\_001124177) 和大西洋鲑(*Salmo salar*, BT072645)的 *COL1A1* 基因 cDNA 全序列, 通过 DNA man 比对他们的同源性, 在保守区间设计引物。

**总 RNA 提取与 cDNA 第一链的合成** RNA 提取参照 Trizol 说明书进行, 利用紫外分光光度计和琼脂糖凝胶电泳检测 RNA 纯度。按照 TaKaRa RNA PCR Kit 方法合成 cDNA 第一链: 第一步, Oligo dT-接头引物 1  $\mu$ L、dNTP mixture 1  $\mu$ L、总 RNA 4  $\mu$ L、加 ddH<sub>2</sub>O 至 10  $\mu$ L。水浴 65 °C 5 min 后, 冰上放置。第二步, 5 $\times$ PrimeScript Buffer 4  $\mu$ L、RNase inhibitor 0.5  $\mu$ L、PrimeScript I Rtas 1  $\mu$ L、第一步反应液 10  $\mu$ L、RNase Free dH<sub>2</sub>O 加至 20  $\mu$ L。在 PCR 仪中进行如下反应: 42 °C 30 min, 99 °C 5 min, 5 °C 5 min。

**草鱼 *COL1A1* 基因 cDNA 核心序列克隆** 根据设计好的引物及克隆策略, 以草鱼肌肉的 cDNA 为模板, PCR 扩增目的基因片段。PCR 反应体系 25  $\mu$ L ddH<sub>2</sub>O 19.8  $\mu$ L、Buffer(已加入 MgCl<sub>2</sub>) 2.5  $\mu$ L、dNTP Mix 1  $\mu$ L、cDNA 1  $\mu$ L、上下游引物各 0.5  $\mu$ L(COL1A1F1: 5'-CAAGAGGACAAGGCGAGGAC-3'与 COL1A1R1: 5'-CGAATGCCATCTTTACCA G-3'; COL1A1F2: 5'-GTGSTGGGTGCTATTGGT G-3'与 COL1A1R2: 5'CACGCTCGCCTCTTTGT C-3'; COL1A1F3: 5'-CTCAAGGTATTGGTGGTC AG-3'与 COL1A1R3: 5'-GCCCTC GCTGCCATAC TC-3')。PCR 反应条件为 94 °C 3 min; 94 °C 1 min,

49 °C 30 s, 72 °C 1 min, 35 个循环; 72 °C 10 min。PCR 扩增产物经 1.0%琼脂糖凝胶电泳检测、回收后, 与 pMD19-T 载体连接, 并转化入大肠杆菌 DH5 $\alpha$  感受态细胞中。挑选阳性克隆, 由上海英俊生物技术有限公司进行测序, 所得的 cDNA 片段与 GenBank 数据库中的已知序列进行 BLAST 同源性分析。

**草鱼 *COL1A1* 基因 cDNA 3'RACE 的克隆** 3'RACE 的操作参照 TaKaRa RNA PCR Kit(AMV)Ver.3.0 试剂盒方法进行。首先以试剂盒提供的 Oligo dT-Adaptor primer 为引物进行逆转录反应, 反转录反应体系: MgCl<sub>2</sub> 2  $\mu$ L、10 $\times$ RT Buffer 1  $\mu$ L、RNase Free dH<sub>2</sub>O 3.75  $\mu$ L、dNTP Mixture 1  $\mu$ L、RNase Inhibitor 0.25  $\mu$ L Oligo dT- Adaptor Primer 0.5  $\mu$ L、实验样品 RNA( 500 ng total RNA)1  $\mu$ L。反转录反应: 42 °C 30 min, 99 °C 5 min, 5 °C 5 min。

用上述的逆转录反应液、5 $\times$ PCR Buffer 10  $\mu$ L、ddH<sub>2</sub>O 28.75  $\mu$ L、TaKaRa Ex Taq<sup>®</sup>HS 0.25  $\mu$ L、COL1A1-3'RACE(5'-GAGGCCTCCCAGAACATT AC-3')0.5  $\mu$ L、M13 Primer M4 0.5  $\mu$ L 进行首次 PCR 反应, 反应条件为 94 °C 3 min; 94 °C 30 s, 55 °C 30 s, 72 °C 1 min, 共 30 个循环。巢式 PCR 所用引物 M13 Primer M4 和 COL1A1-3'2RACE (5'-TAA GAGCCAGACAGCAGAG -3')。PCR 产物在 1.5%琼脂糖凝胶中电泳检测, 并回收、克隆后测序。

**草鱼 *COL1A1* 基因 cDNA 5'RACE 的克隆** 根据已克隆的 *COL1A1* 基因的 cDNA 核心片段, 以 SMART<sup>™</sup> RACE cDNA Amplification Kit 引物设计原则设计 2 个下游引物 COL1A1-5'RACE(5'-CAG CAGAATATCCACAAAGC-3')与 COL1A1-5'2RAC E(5'-CAGACCCCAGC AGAAA CAG-3'), 上游引物 5'-CDS Primer A 与 SMART II A Oligo 由试剂盒提供。取总 RNA 2  $\mu$ L, 合成 5'RACE cDNA 方法参照试剂盒进行。首次 PCR 反应体系为: 10 $\times$  Advantage 2 PCR Buffer 5  $\mu$ L、dNTP (10 mmol/L) 1  $\mu$ L、10 $\times$ UPM 5  $\mu$ L、P5 1  $\mu$ L、5'-RACE-Ready cDNA 2.5  $\mu$ L、50 $\times$ Advantage 2 Polymerase Mix 1  $\mu$ L、加 PCR-Grade water 至 50  $\mu$ L。反应条件为 94 °C 30 s, 65 °C 30 s, 72 °C 3 min, 共 25 个循环。取首次 PCR 产物 5  $\mu$ L, 加入 245  $\mu$ L 的 Tricine-EDTA buffer。然后取 5  $\mu$ L 作为巢式 PCR 的 cDNA。反应体系: 10 $\times$ Advantage 2 PCR Buffer 5  $\mu$ L、dNTP (10

mmol/L)1  $\mu$ L、引物 P6 及 NUP 各 1  $\mu$ L、50 $\times$ Advantage 2、Polymerase Mix 1  $\mu$ L、cDNA 5  $\mu$ L、加 PCR-Grade water 至 50  $\mu$ L。进行 PCR 反应, 扩增条件为: 94  $^{\circ}$ C 30 s, 65  $^{\circ}$ C 30 s, 72  $^{\circ}$ C 3 min, 共 20 个循环。RACE 产物纯化、克隆和测序方法同上。

**COL1A1 基因序列分析** 用 Vector NTI Suite 8.0 软件包中的 Contig Express 程序对获得的 cDNA 序列进行序列拼接, 得到 COL1A1 基因的全长 cDNA 序列。用 DNASTar 搜索开放阅读框并推导氨基酸序列, DNAMAN 和 SignalP 软件预测其信号肽序列, 采用 ClustalX1.83 对推导的氨基酸序列与 NCBI GenBank 数据库(www.ncbi.nih.gov)中其它物种的 COL1A1 氨基酸序列进行同源性比对, 在此基础上通过 MEGA 4.0 软件构建系统进化树。

**COL1A1 蛋白的生物信息学分析** 利用 ProtParam(<http://expasy.org/tools/protparam.html>) 软件在线分析 COL1A1 蛋白序列理化性质; 利用 ProtScale (<http://expasy.org/tools/protscale.html>) 软件在线进行 COL1A1 蛋白质亲疏水性分析; 利用 Predict Protein (<http://www.predictproten.org/>) 软件在线进行 COL1A1 蛋白质二级结构预测; 利用 PROSITE (<http://www.expasy.ch/prosite/>) 软件预测 COL1A1 蛋白质结构功能域。

**不同组织中 COL1A1 基因的 mRNA 表达** 通过上述方法得到不同组织的 cDNA 模板, 根据已克隆的 COL1A1 全长 cDNA 序列, 设计 1 对 RT-PCR 特异引物 COL1A1-F(5'-ACGCACACAAA CAATC TCAAGT-3')和 COL1A1-R(5'-GCATGGGG CAAG ACAGTCA-3'), 以草鱼的  $\beta$ -actin 基因序列设计内参引物  $\beta$ -actin-F(5'-GCAACACGCAGCTCG TIGT A-3') 和  $\beta$ -actin-R(5'-TGACGAGGCTCAGAGCAAGA-3')。PCR 反应体系同上。PCR 反应条件 94  $^{\circ}$ C 3 min; 94  $^{\circ}$ C 30 s, 49  $^{\circ}$ C 30 s, 72  $^{\circ}$ C 1 min, 32 个循环; 72  $^{\circ}$ C 10 min, 实验重复 3 次。

## 2 结果

### 2.1 COL1A1 基因全长 cDNA 序列特征分析

测序结果显示: 草鱼 COL1A1 cDNA 全长为 5 772 bp, 开放阅读框为 4 347 bp(从 131 bp 到 4 478 bp), 编码的氨基酸为 1 448 个。基因序列和推测的蛋白质序列已登录入 GenBank, 序列号为

HM363526。

### 2.2 COL1A1 基因全长 cDNA 序列推导的氨基酸序列分析

根据草鱼 COL1A1 基因全长 cDNA 序列推导其氨基酸序列(图 1), 共有氨基酸 1 448 个, 其中包括 24 个氨基酸的信号肽, 位于 M1-G24 (MFSF-VDIRLALLLSATVLLARGQG)区域。初级结构分析显示, 该肽链含有与哺乳类相似的 N 端、C 端和螺旋区域, 含有 17 个 N-末端肽, 146 个 N-前肽, 25 个 C-末端肽, 278 个的 C-前肽, 1 014 个不间断重复的 Gly-X-Y 螺旋区域。在螺旋区域内有 3 个 RGD(精氨酸-甘氨酸-天冬氨酸)细胞粘连位点, 有两个分子内部交错耦合点 GMKGHR 和 GMKGLR, 三重螺旋区域内 G-G(甘氨酸-甘氨酸)或 G-P-P(甘氨酸-脯氨酸-脯氨酸)重复位点共有 34 个, 氨基化位点 1 个(EGKR), ATP/GTP 绑定位点 1 个(ATGEPGKS), C-前肽内二硫键形成位点 Cys 为 8 个, 糖结合位点(NIT)1 个。

### 2.3 COL1A1 基因推导的氨基酸序列的同源性分析与系统进化树分析

COL1A1 的系统发育树有 3 种聚类类群: 鱼类、两栖类、哺乳类, 草鱼与金鱼、斑马鱼等鱼类首先聚类, 属于同一分支, 具有较高的亲源性。然后与两栖类聚类, 最后与哺乳类聚类。所构建的进化树与传统生物界的划分比较一致(图 2)。

### 2.4 COL1A1 蛋白序列的生物信息学分析

COL1A1 蛋白序列的理化参数分析 基因 COL1A1 的蛋白序列理化参数分析结果表明(表 1): 基因 COL1A1 编码 1 448 个氨基酸, 相对分子质量为 137.2 ku, 分子式为 C5863H9144N1804O1928S44, 理论等电点为 5.44。该蛋白在水溶液中 280 nm 处的摩尔消光系数为 51 840 mol/(L $\cdot$ cm)。若其成熟肽 N 端为蛋氨酸时, 在哺乳动物网状红细胞体外表达的半衰期为 30 h, 在酵母和大肠埃希菌中表达的半衰期分别大于 20 h 和 10 h。在溶液中的不稳定指数为 25.15, 在溶液中性质稳定。疏水指数为-0.739, 表明其具有较高亲水性。

**COL1A1 蛋白质亲疏水性分析** 利用 ProtScale 软件预测结果如图 3 所示, 该蛋白质存在 2 个高分值(Scare>1.5)峰, 分布在 8~17, 1 442~1 444, 属于高疏水性域; 32 个低分值(Scare<-1.5), 分布在 23~29, 39~48, 72~74, 92~95, 103~112, 174~191, 212~232, 291~198, 353~359, 446~457, 468~469,

```

1  MFSFVDIRLALLLSATVLLARGQGEDDRDTGGSC TLDGQVY
41  NDRD VWKPEPCQICVCD SGTVMCDEVICEDTTDCPNPVI P
81  HDECCPVCPDDDFQEPSVEGPRGTPGEKGD RGPAGPPGND
121 GIPGQPGLPGPPGPPGPPGLGGNFSPQMSGGFDEKSGGAM
161 AVP{GPMGPMGPRGPPGPPGTPGPQGFTGPPGEPGEAGAPG
201 PMGPRGAAGPPGKNGEDGESGKPGRPGER GPPGPQGARGF
241 PGTPLPGIKGHRGF SGLDGA KGDTG P SGP KGEAGAPGEN
281 GTPGAMGPRGLPGERGRAGPPGAAGARGNDGAAGAAGPPG
321 PTGPAGPPGFPGGPGAKGEVGPQGARGAEGPQGARGEAGN
361 PGPAGPAGPAGNNGADGAAGPKGSPGTPGIAGAPGFPGPR
401 GPPGPSGAAGAPGPKGNTGEVGA PGAKGEAGAKGEAGA QG
441 VQGP PGPPEEGKR GAR GEPGAAGGRGPPGERGAPGARGF
481 PGADGSAGPKGAPGERGGPGVVGPKGATGEPGRNGEPGMP
521 GSKGMTGSPGSPGPDGKTGPSGT PGQDGRPGPPGPVARG
561 QPGVMGFPGPKGAAGEAGKPGERGV MGA VGATGAPGKDGD
601 VGAPGAPGPAGPAGERGEQGPAGPPGFQGLPGPQGATGEP
641 GKSGEQGV PGEAGAPGPAGS RGD RGF PGERGAPGPAGPAG
681 ARGSPGSAGNDGAKGDAGAPGAPGAQGP PGLQGM PGERGA
721 AGLPLKGD RGD QGAKGTDGAPGKDGIRGMTGPIGPPGPA
761 GAPGDKGETGAPGLVGP TGAR GPPGERGETGAPGPAGFAG
801 PPGADGLPGA KGEAGDNGAKGDAGSPGPAGATGAPGPQGP
841 VGATGPKGARGAAGPPGATGFPGAAGR VGP PGPAGNAGPP
881 GPPGPGGKEGQKGNRGETGPAGRTGEV GAGP PPGAPGEKG
921 TPGAEGPTGPAGIPGPQGIAGQRGIVGLPGQRGERGF PGL
961 PGPSGEPGKQGSPSGPSGERGPPGPMGPPGLAGPPGEPGRE
1001 GTPGNEGSAGR DGAPGPKGDRGETGAAGTPGAPGPPGAPG
1041 PVGPAGKTGDRGESGPAGPAGAAGPAGPRGPAGPAGARGD
1081 KGETGEAGERGMKGHRGFTGMQGP PPGPPGPSGEPGPAGAS
1121 GPAGPRGPAGSSGPA GKDGMSGLPGPIGPPGPRGRNGEIG
1161 PGP PPGAPGPPGPPGPS}GGGFDIGFIAQPQE KAPDPFRHF
1201 RADDANVMRDRDLEVD TLLKSL SQQIESIMSPDGT KKNPA
1241 RTORDLKM HPDWKSGEY WIDPDQG NQDAIKVY CNMETG
1281 ETVYPT ESTIPKKNWYTSKNIKEK KHVWFGEAMTDGFQF
1321 EYGSEGSKAEDVNIQLTFLRLMSTEASQ NITYH CNKNSIAY
1361 MDQASGNLKKALLLQGSNEIEIRAEGNSRFTYSVTE DGCOT
1401 SH TGA WGKTVIDYKTTKTSRLPIIDIA PMDVGAPNQEFGI
1441 EVG PV F L *

```

图 1 草鱼 COL1A1 基因全长 cDNA 序列推导的氨基酸序列

细下划线表示信号肽; 箭头表示 N-前肽和 C-前肽分离位点; 大括号表示三重螺旋结构域; 双下划线表示分子内部交错耦合点; 黑体表示三重螺旋区域内 G-G 或 G-P-P 重复位点; 方块表示 RGD 细胞粘连位点; 粗下划线表示 NIT 糖基化结合位点; 圆圈表示二硫键形成位点; 阴影表示氨基化位点; 黑体加斜体表示 ATP/GTP 绑定位点。星号表示终止密码子。

Fig. 1 Nucleotide and deduced amino acid sequence of full-length COL1A1 cDNA in grass carp

Single underlined indicates the cleavage site of the putative signal peptide; arrowheads indicate N-propeptide and C-propeptide cleavage site; Gly-Gly and Gly-Pro-Pro repeats inside the triple hical domain (bracketed) are in bold; The putative intermolecular cross-linking sites are double underlined; The three RGD sites which could represent potential cell adhesion sites are boxed; The NIT (Asn-Ile-Thr amino acids) dotted sequence indicates the putative carbohydrate attachment site; Cysteine residues in the C-propetide are circled; shadow indicates the Amidation site; bold and italic indicates ATP/GTP-binding site; asterisk indicates stop codon.

512~517, 545~551, 615~620, 642~643, 660~667, 728~36, 783~788, 885~899, 951~955, 967~980, 996~1004, 1007~1009, 1014~1022, 1047~1054, 1080~1087, 1092~1093, 1152~1157, 1191~1214, 1234~1241, 1253~1266, 1296~1305, COL1A1 蛋白是一种亲水性(Score<0)很强的蛋白。这些低分值区域属于

高亲水性域, 往往富集于表面, 是其线性抗原决定簇的位置, 同时也是蛋白质进化中氨基酸插入的主要位点。

COL1A1 蛋白预测的二级结构特征 用 Predict protein 预测结果如图 4-b 所示, COL1A1 蛋白氨基酸有 6 个 α 螺旋结构域, 分布在 6~18, 1219~1225,

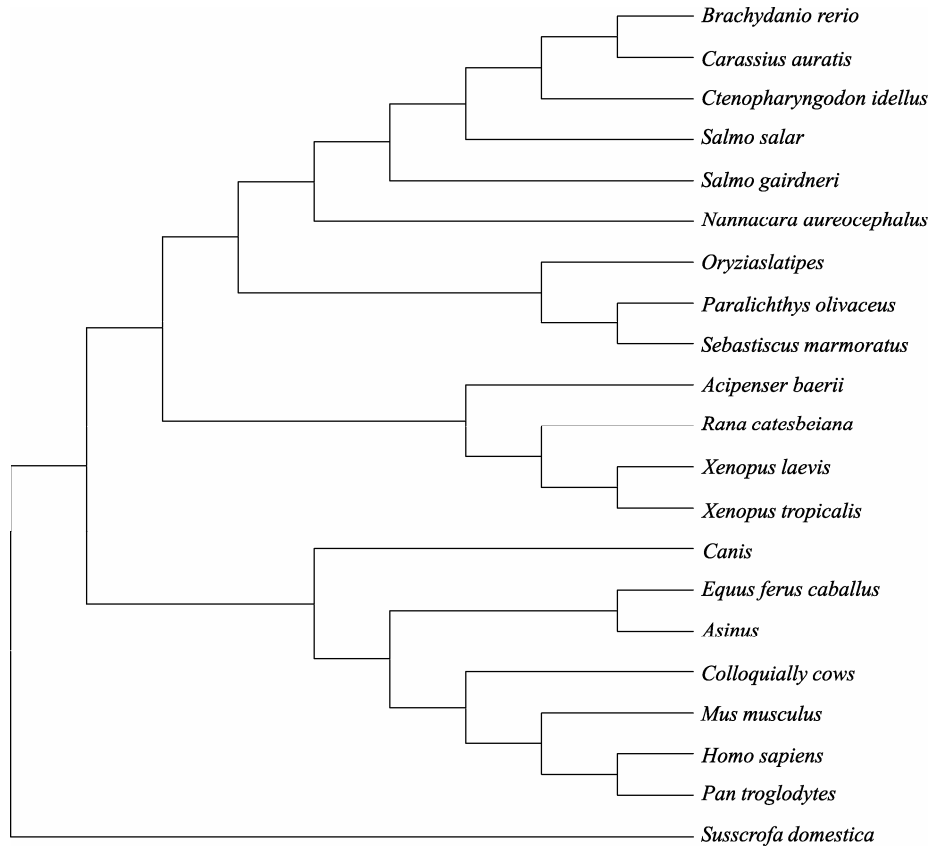


图 2 草鱼 COL1A1 基因推导的氨基酸序列同源性和系统进化树分析

Fig. 2 Multiple alignment and phylogenetic analysis of deduced amino acid sequences of COL1A1

大西洋鲑 *Salmo salar*: (BT072645); 斑马鱼 *Brachydanio rerio*: (BC161663); 金鱼 *Carassius auratis*: (DQ324363); 草鱼 *Ctenopharyngodon idellus*: (HM363526); 虹鳟 *Salmo gairdneri*: (AB275454); 乌鲷 *Nannacara aureocephalus*: (NM001124177); 日本青鳉 *Oryziaslatipes*: (NM001122918); 日本牙鲈 *Paralicthys olivaceus*: (AB196513); 褐菖鲉 *Sebastiscus marmoratus*: (GU067476); 西伯利亚鲟 *Acipenser baerii*: (EU241868); 牛蛙 *Rana catesbeiana*: (AB015440); 非洲爪蛙 *Xenopus laevis*: (NM001087352); 西方爪蛙 *Xenopus tropicalis*: (NM001011005); 狗 *Canis*: (BC050014); 马 *Equus ferus caballus*: (NM001003090); 驴 *Asinus*: (FJ594763); 牛 *Colloquially cows*: (NM001034039); 家鼠 *Mus musculus*: (AF034691); 人 *Homo sapiens*: (NM0000088); 黑猩猩 *Pan troglodytes*: (XM01169052); 猪 *Suscrofa domestica*: (AY165952)。

1311~1313, 1361~1362, 1368~1373, 1380~1383 区域; 有 12 个  $\beta$  折叠域, 分布在 4~5, 39~40, 53~56, 1283~1284, 1320~1322, 1334~1336, 1350~1356, 1359~1360, 1390~1396, 1400~1414, 1437~1438, 1440~1442 区域, 其余区域为无规则卷曲域。此外, 该蛋白无跨膜区域。

COL1A1 蛋白结构功能域分析 利用 PROSITE 软件预测结果如图 4-A 和 4-C 所示, COL1A1 蛋白三重螺旋重复域有 18 段, 低复杂度域有 22 段, 功能域主要集中在 1~200, 有 1 个为胶原蛋白纤维 C-末端区域 COL1F(Fibrillar collagens C-terminal domain)(图 4-A)和 1 个 N-末端 VWC(von Willebrand factor)功能域, 对其功能域进一步分析(图 4-C), 有两个绑定钙的区域(EGF CA 和 CLECT), 分别在 33~69, 1249~1355 位置, 有 1 个

绑定锌的区域(ZnFF ZZ), 在 62~104 位置, 钙和锌的区域有重叠区域。此外, 还有 1 个纤维蛋白原相关功能域 FBG 等共 17 段小功能域。

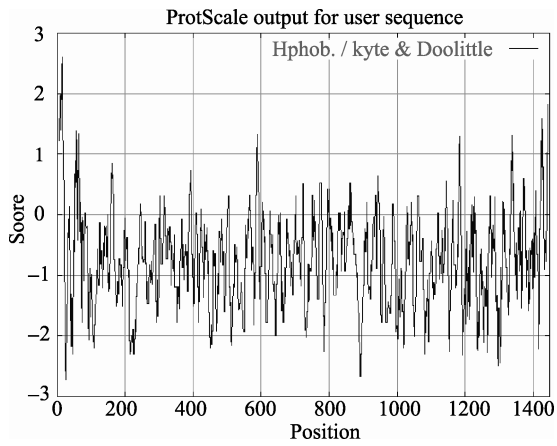
草鱼 COL1A1 基因的组织分布及差异性表达 半定量 PCR 结果显示, COL1A1 基因在草鱼的肌肉、肝胰脏、肠道、鳃、脾脏、肾脏、皮肤、鳍 8 个组织中均有表达分布(图 5)。以 COL1A1 基因与  $\beta$ -actin 电泳带吸光度的比值绘制了 COL1A1 在草鱼各组织的表达相对丰度(图 5), 结果显示: COL1A1 基因在皮肤、鳃、肾脏、鳍条中的表达量较高( $P>0.05$ ), 高于其他 4 个组织( $P<0.05$ )。

### 3 讨论

3.1 草鱼肌肉 I 型胶原蛋白 COL1A1 进化树比较 从系统进化树结果来看, COL1A1 基因氨基酸

**表1 COL1A1 蛋白序列的主要理化参数分析**  
**Tab.1 Main physico-chemical parameters of COL1A1 proteins analysis**

氨基酸残基数	number of amino acids residues	1448
分子质量	molecular weight	137162.6
理论等电点	theoretical pI	5.44
氨基酸组成	amino acid composition	
Ala (A)	155	10.7%
.....		
Val (V)	39	2.7%
负电荷氨基酸残基总数	total number of negatively charged residues (Asp + Glu)	143
正电荷氨基酸残基总数	total number of positively charged residues (Arg + Lys)	127
原子组成	atomic composition	
Carbon	C	5863
Hydrogen	H	9144
Nitrogen	N	1804
Oxygen	O	1928
Sulfur	S	44
分子式	formula	C5863H9144N1804O1928S44
原子总数	total number of atoms	18783
消光系数	extinction coefficients	
Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.		
Ext. coefficient	49025	
Abs 0.1% (=1 g/l)	0.357,	assuming all pairs of Cys residues form cystines
Ext. coefficient	47900	
Abs 0.1% (=1 g/l)	0.349,	assuming all Cys residues are reduced
半衰期	estimated half-life	
The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).		
不稳定系数	instability index	
The instability index (II) is computed to be 24.28		
脂肪系数	aliphatic index	36.56
总平均疏水性	grand average of hydropathicity (GRAVY):	-0.739



**图3 COL1A1 蛋白质氨基酸序列的亲疏水性**  
**Fig.3 Predicted hydrophobicity and hydrophilicity of deduced amino acid sequence of COL1A1 protein**

序列系统发育树有 3 种聚类类群：鱼类、两栖类及哺乳类，草鱼与金鱼、斑马鱼等鱼类首先聚类，并且与斑马鱼、金鱼处在一支，最后再依次与两栖类、哺乳类聚类，说明 COL1A1 序列具有较高的保守性。草鱼 COL1A1 基因系统进化树分析结果与传统的鱼类分类地位相吻合。对鱼类 COL1A1 基因研究结果发现在不同分类地位的鱼类中 COL1A1 的氨基酸同源性很高<sup>[10-12]</sup>，提示 COL1A1 在进化过程中较为保守，对于鱼类的 COL1A1 基因研究有助于鱼类损伤组织修复中的分子机制作用阐述。

**3.2 草鱼肌肉 I 型胶原蛋白 COL1A1 结构**

草鱼肌肉的 COL1A1 基因 cDNA 全序列为 5 772 bp，对其 COL1A1 链的初级结构分析发现，其含有与高等脊椎动物纤维状胶原蛋白具有相同的信号肽、C-前肽、N-前肽结构域和三重螺旋结构域<sup>[16-18]</sup>。C-前肽对胶原蛋白三重螺旋的形成有着重要的作用，不仅促进形成成熟的前胶原蛋白并使其转化成胶原蛋白，且是胶原蛋白形成胶原纤维的基础<sup>[10]</sup>。在草鱼 COL1A1 基因的 C-前肽区域中发现 8 个二硫键残基(-cys-)，其数量与虹鳟的 COL1A1<sup>[10]</sup>、斑马鱼<sup>[19]</sup>的 COL1A2 相同，该残基在 3 个前肽 a 链绑定成三聚胶原蛋白分子过程中是必不可少的<sup>[20]</sup>。N-前肽调控胶原蛋白纤维的直径和形态，进而影响其结构和功能<sup>[18]</sup>。三重螺旋结构对 mRNA 转录起着重要的调控作用<sup>[21]</sup>。在草鱼 COL1A1 基因的三重螺旋结构中有 GMKGGHRGM-KGLR 两个交错偶联点，分别在 248~253 和 1091~1096 位置，这些交错偶联域是胶原蛋白纤维理化性质特征和网状结构形态稳定的的必要条件<sup>[20]</sup>。在虹鳟<sup>[12]</sup>COL1A1、斑马鱼<sup>[19]</sup> COL1A2 中也发现这两个偶联位点，表明草鱼和虹鳟、斑马鱼有相似的交错偶联信息。该三重螺旋结构有 3 个 RGD 细胞粘连位点，这种位点在虹鳟和其他脊椎动物中均有发现<sup>[12]</sup>。同时草鱼 COL1A1 中还发现 1 个氨基化位点(EGKR)和 1 个 ATP/GTP 绑定位点。在草鱼 COL1A1 肽链的三重螺旋区域有 31 个 G-P-P 重复肽，已有的研究证明虹鳟有 23 个 G-P-P 重复肽，含有较高的这些重复肽是脊椎动物胶原蛋白的特征<sup>[10]</sup>。目前，这些重复肽的功能还不清楚，Saito 等<sup>[10]</sup>在研究虹鳟皮肤和肌肉的热稳定性中发现这些重复肽可能对胶原蛋白的稳定性和修复发挥着重要作用。

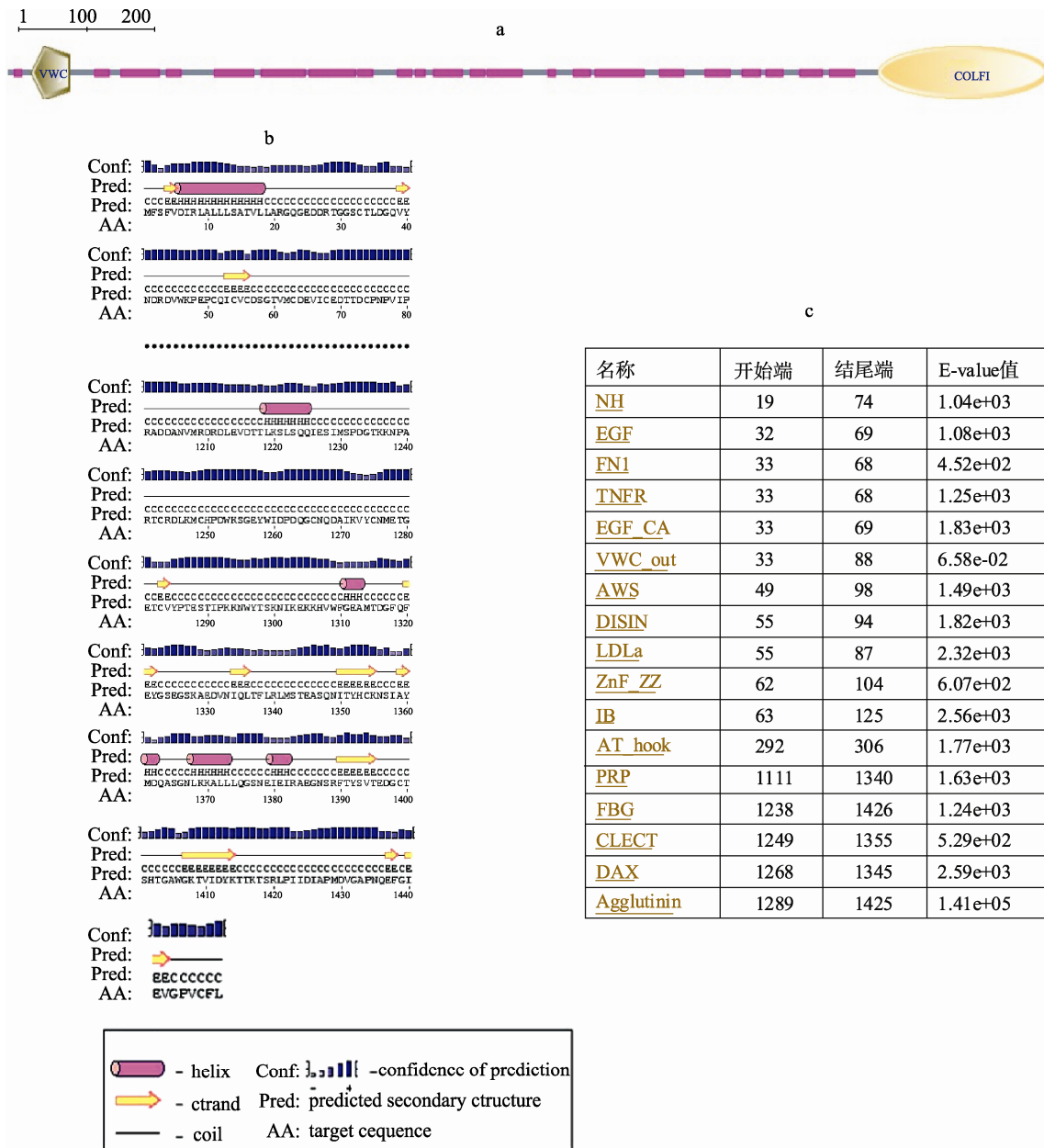


图 4 草鱼 COL1A1 蛋白序列二级结构特征和结构、功能域预测分析

Fig. 4 Secondary structure characterization and predicted domains, functional sites of COL1A1 proteins

NH: Neurohypophysial hormones; EGF:Epidermal growth factor-like domain; FN1: Fibronectin type 1 domain; TNFR:Tumor necrosis factor receptor; EGF\_CA:Calcium-binding EGF-like domain; VWC\_out: von Willebrand factor out; AWS:associated with SET domains; DISIN:Homologues of snake disintegrins; LDLa:Low-density lipoprotein receptor domain class A; ZnF\_ZZ:Zinc-binding domain, present in Dystrophin, CREB-binding protein; IB:Insulin growth factor-binding protein homologues; AT\_hook: DNA binding domain with preference for A/T rich regions; PRP:Major prion protein; FBG:Fibrinogen-related domains (FReDs); CLECT:C-type lectin (CTL) or carbohydrate-recognition domain (CRD); DAX: Domain present in Dishevelled and axin; Agglutinin: Agglutinin.

草鱼 COL1A1 肽链的最后 10 个氨基酸序列具有较高的保守性, 在哺乳动物中均有发现, 且其对 3 个 a 链折叠成三重螺旋结构适当排列和标记具有重要的作用<sup>[22]</sup>。胶原蛋白的生物合成是从基因转录开始到形成大的复合纤维状多聚体的过程, 同时其初级结构组成特点是发挥生物功能的前提基础, 特别是基于纤维状胶原蛋白中的 I 型胶原蛋白, 它的

三重螺旋信息和功能特征的基本机制可以应用到肌肉损伤的胶原蛋白修复研究<sup>[20]</sup>。

3.3 草鱼肌肉 I 型胶原蛋白生物信息学特征比较

草鱼 COL1A1 蛋白的总平均疏水性指数为 -0.739, 亲水性区域(Scare<0)均匀分布在整個肽链中, 而疏水性区域(Scare>0)较少, 没有跨膜结构域。已有研究表明, 跨膜区常出现在高疏水区区域<sup>[23]</sup>,

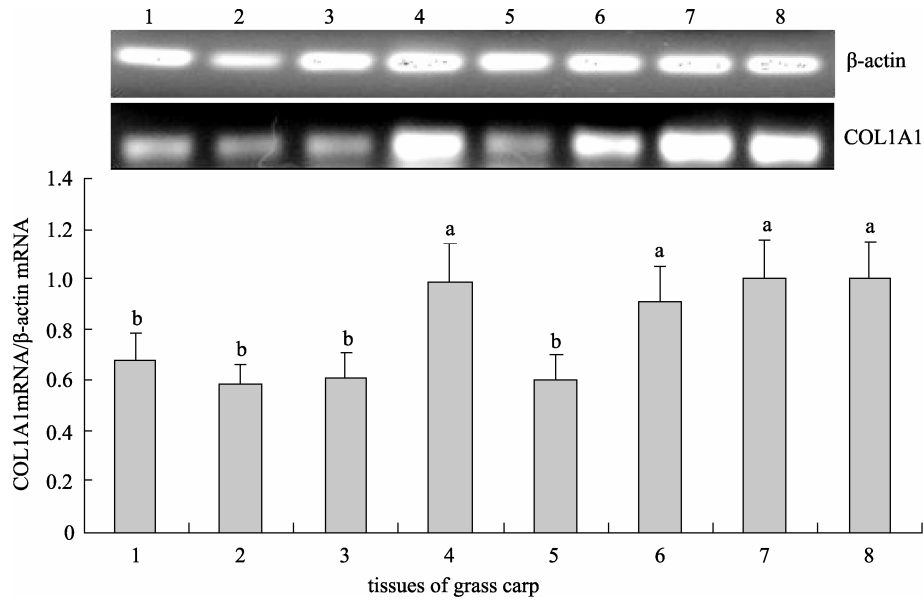


图 5 COL1A1 mRNA 在草鱼不同组织中的差异性表达

1.肌肉; 2.肝胰脏; 3.肠道; 4.鳃; 5.脾脏; 6.肾脏; 7.皮肤; 8.鳍条。

Fig. 5 The significance tests and expression of COL1A1 in different tissues of grass carp by RT-PCR

1. muscle; 2. hepatopancreas; 3. intestines; 4. gill; 5. spleen; 6. kidney; 7. skin; 8. fin.

此结果与亲疏水性分析相符合, COL1A1 蛋白属于高亲水性蛋白, 其较高的亲水能力促进了其蓄水和与其他物质的结合能力, 这与 Lapiere 等<sup>[24]</sup>研究 I 型胶原蛋白具有增强皮肤弹性和蓄水能力的功能论述相符。

COL1A1 蛋白二级结构预测为  $\alpha$  螺旋  $\beta$  折叠三重螺旋结构蛋白, 三维螺旋结构不但使其在 TGF- $\beta$ /Smad 信号传导通路中发挥重要作用<sup>[25-27]</sup>, 也使其胶原纤维形成一种复杂的网状结构, 从而加强了胶原纤维的硬度和机械强度, 其构象发生改变能引起肌肉的口感变化, 这在鱼类肌肉品质改良和修复研究中具有潜在价值<sup>[28]</sup>。COL1A1 蛋白还有无规则卷曲域, 这些区域保证了胶原蛋白纤维分子结构的有序排列和韧性。草鱼 COL1A1 蛋白具有 17 个功能域, 主要分布在 C-末端附近, 其中最重要的是具有 2 个含有 CXXC 和 (H)XCXXC 结构组成特征的绑定  $Ca^{2+}$  的结构域<sup>[29]</sup>和 1 个绑定  $Zn^{2+}$  的结构域, Prockop 等<sup>[30]</sup>研究表明 C-前肽的分离需要前胶原蛋白 C-蛋白酶, 这种酶属于  $Zn^{2+}$  绑定金属蛋白酶家族 ( $Zn^{2+}$ -dependent metalloproteinases), 这些特征与虹鳟<sup>[10]</sup>和人<sup>[29]</sup>肌肉组织成纤维细胞 I 型胶原蛋白的结构类似, 具有激活 C-前肽的分离, 促使 I 型胶原蛋白形成稳定结构的功能<sup>[30]</sup>。

### 3.4 草鱼肌肉 I 型胶原蛋白 COL1A1 的组织分布

草鱼 COL1A1 基因 mRNA 在 8 个组织中均有

表达且具有差异性, 其中, 在皮肤、鳃、肾脏、鳍条中的表达量较高 ( $P > 0.05$ ), 在其他 4 个组织中表达较低 ( $P < 0.05$ )。Kondo 等<sup>[31]</sup>对金鱼的研究结果表明, I 型胶原蛋白 COL1A1 在所检测的各组织中均有表达, 其中鳃、皮肤、鱼鳞中的表达含量较高。I 型胶原蛋白 COL1A1 在草鱼组织中有广泛的分布, 同时又具有差异性。已有的研究表明 COL1A1 的 mRNA 的表达量与温度、组织 HSP70 的 mRNA 水平及组织细胞生长有关<sup>[31-33]</sup>。这种组成性表达模式差异的生物修复方面的意义有待进一步研究。

鱼类损伤组织修复通常有多种途径来完成, 虽然 I 型胶原蛋白不是鱼类损伤组织修复过程的关键蛋白, 但它是一种多功能蛋白, 利用生物信息学方法分析得到的结果将有助于全面了解 COL1A1 的功能及其稳定表达信息。

### 参考文献:

- [1] Kimura S, Zhu X P, Matsui R, et al. Characterization of fish muscle type I collagen [J]. Journal of Food Science, 1988, 53(5): 1315-1318.
- [2] 蒋挺大, 张春萍. 胶原蛋白[M]. 北京: 化学工业出版社, 2001.
- [3] 陈国梁, 贺翠莲. 胶原蛋白的研究进展[J]. 延安大学学报: 自然科学版, 2000, 19(2): 78-81.
- [4] Prockop D J, Kivirikko K L. Collagens: molecular biology diseases and potentials for therapy [J]. Annual Review of Biochemistry, 1995, 64: 403-434.



- [5] Van der Rest M, Garrone R, Herbage D. Collagen: a family of proteins with many facets[M]/Kleinman H K (Ed.). *Advances in Molecular and Cell Biology* (Published by Elsevier B.V.), 1993, 6: 1–67.
- [6] Perris R, Perissinotto D. Role of the extracellular matrix during neural crest cell migration [J]. *Mechanisms of Development*, 2000, 95(1–2): 3–21.
- [7] Deutzmann R, Fowler S, Zhang X, *et al.* Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in Hydra[J]. *Development*, 2000, 127: 4669–4680.
- [8] Fowler S J, Jose S, Zhang X, *et al.* Characterisation of Hydra type IV collagen: Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose [J]. *The Journal of Biological Chemistry*, 2000, 275 (50): 39589–39599.
- [9] Lina B, Juan J B, Vincent T T, *et al.* Platelet thrombus formation on collagen type I. A model of deep vessel injury influence of blood rheology, von willebrand factor, and blood coagulation[J]. *Circulation*, 1988, 78: 1431–1442.
- [10] Saito M, Takenouchi Y, Kunisaki, *et al.* Complete primary structure of rainbow trout type I collagen consisting of  $\alpha 1(I)\alpha 2(I)\alpha 3(I)$  heterotrimers[J]. *European Journal of Biochemistry*, 2001, 268: 2817–2827.
- [11] Kimura S, Ohno Y. Fish type I collagen: tissue-specific existence of two molecular forms,  $(\alpha 1)2\alpha 2$  and  $\alpha 1\alpha 2\alpha 3$ , in Alaska pollack [J]. *Comparative Biochemistry and Physiology, Part B: Biochemistry & Molecular Biology* 1987, 88(2): 409–413.
- [12] D'Alessio M, Ramirez F, Suzuki H R, *et al.* Structure and developmental expression of a sea urchin fibrillar collagen gene[J]. *Proceedings of the National Acad Sciences*, 1989, 86 (23): 9303–9307.
- [13] 李维平, 王光荣. 两种大容器运输大规格草鱼种的对比试验[J]. *水利渔业*, 1998, 2: 46–47.
- [14] 余秀芳, 汤哲文, 周金富, 等. 精养水库草鱼轮捕试验[J]. *科学养鱼*, 2008, 11: 70–70.
- [15] 梁养贤. 模拟长途运输中草鱼血液和水质变化研究[D]. 长沙: 湖南农业大学, 2008.
- [16] Fuller F, Boedtker H. Sequence determination and analysis of the 3' region of chicken pro- $\alpha 1(I)$  and pro- $\alpha 2(I)$  collagen messenger ribonucleic acids including the carboxyterminal propeptide sequences[J]. *Biochemistry*, 1981, 20(4): 996–1006.
- [17] Finer M H, Boedtker H, Doty P. Construction and characterization of cDNA clones encoding the 5' end of the chicken pro  $\alpha 1(I)$  collagen Mrna[J]. *Gene*, 1987, 56(1): 71–78.
- [18] Asahina K, Utoh R, Obara M, *et al.* Cell-type specific and thyroid hormone-dependent expression of genes of  $\alpha 1(I)$  and  $\alpha 2(I)$  collagen in intestine during amphibian metamorphosis[J]. *Matrix Biology*, 1999, 18(1): 283–289.
- [19] Ghislaine M D, Zofia H, Carole C, *et al.* Structure and spatio temporal expression of the full length DNA complementary to RNA coding for  $\alpha 2$  type I collagen of zebrafish[J]. *Gene*, 2002, 294: 55–65.
- [20] Gelse K, Poschl E, Aigner T. Collagens—structure, function, and biosynthesis [J]. *Advanced Drug Delivery Reviews*, 2003, 55(12): 1531–1546.
- [21] Joseph P R O O, Thomas C I, Andrew M, *et al.* Microfibrillar structure of type I collagen *in situ*[J]. *Proteins Natural Academic Science*, 2006, 103(24): 9001–9005
- [22] Joseph P R O O, Andrew M, Thomas C I, *et al.* The *in situ* Supermolecular Structure of Type I Collagen[J]. *Structure*, 2001, 9: 1061–1069.
- [23] Finn R D, Tate J, Mistry J, *et al.* The Pfam protein family database[J]. *Nucleic Acids Research*, 2008, 36(suppl 1): D281–288.
- [24] Lapiere C M. The aging dermis : the main collagen for the appearance of “old”skin[J]. *British Journal of Dermatology*, 1990, 122 (S1): 35–38.
- [25] Gelse K, Poschl E, Aigner T. Collagens-structure, function, and biosynthesis[J]. *Advanced Drug Delivery Reviews*, 2003, 55(12): 1531–1539.
- [26] Vogel W F. Collagen-receptor signaling in health and disease [J]. *European Journal of Dermatology*, 2001, 11 (6): 506–509.
- [27] Yamaguchi Y, Mann D M, Ruoslathi E. Negative regulation of transforming growth factor by the proteoglycan decorin[J]. *Nature*, 1990, 346: 281–284.
- [28] Masashi A, Haruhiko T, Morihiko S, *et al.* Threedimensional structure of collagen fibrillar network of pericellular connective tissue in association with firmness of fish muscle[J]. *Nippon Suisan Gakkaishi*, 1992, 58(7): 1361–1364.
- [29] Cordula B, Annett S, Ernst P R. Transcriptional activation of the type I collagen genes COL1A1 and COL1A2 in fibroblasts by interleukin-4: Analysis of the functional collagen promoter sequences [J]. *Journal of Cellular Physiology*, 2004, 198 (2): 248–258.
- [30] Prockop D J, Sieron A L, Li S W. Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling[J]. *Matrix Biology*, 1998, 16(7): 399–408.
- [31] Kondo H, Watabe S. Temperature-dependent enhancement of cell proliferation and mRNA expression for type I collagen and HSP70 in primary cultured goldfish cells [J]. *Comparative Biochemistry and Physiology, Part A: Molecular & Integrative Physiology* 2004, 6 (138): 221–228.
- [32] Kindy M S, Chang C J, Sonenshein G E. Serum deprivation of vascular smooth muscle cells enhances collagen gene expression[J]. *The Journal of Biology Chemistry*, 1988, 263: 11426–11430.
- [33] Liao G, Chan L M. Regulation of extracellular matrix RNA levels in cultured smooth muscle cells. Relationship to cellular quiescence[J]. *The Journal of Biology Chemistry*, 1989, 264: 10315–10320.

## Cloning, mRNA expression and bioinformatic analysis of full length type I collagen alpha 1 cDNA from grass carp (*Ctenopharyngodon idellus*)

LIU Bang-hui<sup>1,2</sup>, YU Er-meng<sup>1</sup>, WANG Guang-jun<sup>1</sup>, YU De-guang<sup>1</sup>, XIE Jun<sup>1\*</sup>,  
WANG Hai-ying<sup>1</sup>, GONG Wang-bao<sup>1</sup>

(1. Key Laboratory of Tropical & Subtropical Fishery Resource Application & Cultivation, Ministry of Agriculture, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, China;

2. College of Fisheries and Life Sciences, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** Type I collagen, as a member of the collagen family, is the most abundant collagen and major components of the extracellular matrices of all metazoan life, and plays crucial roles in differentiation, formation of collagen fibers and tissue remodeling after injury, etc. Type I collagen alpha 1 (COL1A1) cDNA of grass carp (*Ctenopharyngodon idella*) was isolated through the RT-PCR and RACE approaches. The cDNA was 5 772-bp in length, including a 4 347-bp CDS (coding sequence) and encoded a polypeptide of 1 449 aa. The homology of COL1A1 amino acid with relative species (zebrafish, goldfish, etc.) was as high as 93% with zebra fish and goldfish. The protein peptide molecular weight was 137.2 ku and theoretical pI was 5.44 using ProtParam software on line. The protein peptides of COL1A1 possessed 6  $\alpha$ -helixes, 12  $\beta$ -sheets, others of ruleless coil regions, and 18 regions of triple helical repeats, 22 low complexity regions, 17 function domains. There were two calcium-binding sites and one zinc-binding site in the COL1A1 protein peptide. COL1A1 mRNA was determined in all the tested 8 tissues (muscle, intestine, hepatopancreas, gill, skin, fin, kidney and spleen) of grass carp by semi-quantitative RT-PCR, and the mRNAs expression in gill, kidney, skin and fin significantly higher than other tissues ( $P < 0.05$ ). The structure and bioinformatics characteristics of the COL1A1 from grass carp may help to further understand the function of COL1A1 gene in the repair process of damaged tissue in the grass carp.

**Key words:** *Ctenopharyngodon idella*; type I collagen; COL1A1; mRNA expression; bioinformatics

**Corresponding author:** XIE Jun. E-mail: xiejunhy@yahoo.com.cn