

鳊鱼(*Siniperca chuatsi*)幽门垂中两种胰蛋白酶的 cDNA 克隆及其生物信息学分析*

杜翠红 苏文金 卢宝驹 刘光明 邱晓燕 曹敏杰

(福建省高校水产科学技术与食品安全重点实验室 集美大学生物工程学院 厦门 361021)

摘要 采用 RT-PCR、3'-RACE 及 5'-RACE 技术从鳊鱼幽门垂组织中克隆了两种胰蛋白酶(分别命名为 TRS-A 与 TRS-B)cDNA 全长序列。将测序结果递交 GenBank 数据库, 分别获得登录号为 ACD70339 和 ACD70340。采用生物信息学的方法和工具预测了鳊鱼胰蛋白酶的理化参数及其高级结构, 并构建了系统进化树。结果表明: (1) TRS-A 和 TRS-B 的 cDNA 开放阅读框编码的蛋白序列中均含有一段长度为 15 个氨基酸残基的信号肽、5 个氨基酸残基组成的激活肽、222 个氨基酸残基组成的成熟蛋白区域; (2) TRS-A 成熟蛋白的相对分子量为 24.1kDa、理论等电点为 5.69, TRS-B 的相对分子量为 24.2kDa、理论等电点为 5.75; (3) 建立了 TRS-A 和 TRS-B 的三维结构模型, 预测了其活性位点及 12 个半胱氨酸残基形成二硫键的位置。

关键词 鳊鱼, 胰蛋白酶, 基因克隆, 生物信息学

中图分类号 Q943

胰蛋白酶(EC3.4.21.4)属于丝氨酸蛋白酶家族, 它们以酶原前体形式从胰腺腺泡细胞中产生, 去掉信号肽后变成胰蛋白酶原, 而后释放运输到肠道中, 被肠激酶激活去掉激活肽变成有活性的胰蛋白酶(Chen *et al*, 2003)。胰蛋白酶的主要生理功能有两个: 一是消化食物中的蛋白质; 二是激活胰腺分泌的其他酶原, 包括胰凝乳蛋白酶酶原、弹性蛋白酶酶原和羧肽酶酶原等, 同时对其自身的酶原也有激活作用, 使这些酶变成活性的形式, 从而行使其消化功能(王镜岩等, 2002)。因此, 胰蛋白酶是动物消化道内的一种重要的水解酶, 它不仅对胰腺分泌的多种蛋白酶原的激活起着重要的作用, 而且也是重要的工具酶, 广泛应用于轻工业、医药工业、食品加工业、畜牧业和现代生物技术等领域(Gudmundsdottir *et al*, 2005; 刘欣华等, 2002)。

研究发现, 鱼类胰蛋白酶的活力是哺乳动物同类酶的数倍到数十倍(Smalas *et al*, 1994), 国内外对

鱼类胰蛋白酶的研究大多集中在海水鱼(Spilliaert *et al*, 1999; Douglas *et al*, 1998; Rune *et al*, 1995)上, 而对于淡水鱼类的研究相对较少。鳊鱼(*Siniperca chuatsi*)自古以来一直是我国最重要的淡水名贵经济鱼类, 具有生长快、肉质鲜嫩、少刺等特点, 享有“淡水石斑鱼”的美誉(刘秀霞等, 2009)。鳊鱼为典型的肉食性鱼类, 自开食起终生以活鱼虾为食, 且十分顽固, 因此, 有必要对其消化道中的一些重要水解酶进行深入研究。近期, 本课题组对鳊鱼胰蛋白酶的分离纯化及其酶学性质已进行了系统研究(Lu *et al*, 2008), 但有关其基因结构及分子特征等研究尚未见报道。

本研究通过 RT-PCR、3' RACE 及 5' RACE 技术, 从鳊鱼幽门垂组织中克隆到两种编码胰蛋白酶的 cDNA 全长序列, 并利用生物信息学相关网站及软件, 对其核苷酸序列和蛋白序列进行了分析并构建了基于胰蛋白酶氨基酸序列的系统进化树; 采用同源建模法, 建立了鳊鱼胰蛋白酶的三维结构模型, 从而预

* 国家自然科学基金项目资助, 30871947 号, 20872049 号; 福建省自然科学基金项目资助, 2008J0067 号, 2010J01212 号; 集美大学中青年创新团队基金, 2006A002 号。杜翠红, 副教授, E-mail: cuihongdu@jmu.edu.cn

通讯作者: 曹敏杰, 教授, E-mail: mjcao@jmu.edu.cn

收稿日期: 2010-07-23, 收修改稿日期: 2010-09-15

测其活性位点,可为进一步研究鱼类胰蛋白酶的结构和功能关系及其定点突变奠定基础。

1 材料与方法

1.1 材料

鲜活鳊鱼(*Siniperca chuatsi*), 体重约 500g, 购于福建厦门集美菜市场, 即杀后迅速取其幽门垂组织用于总 RNA 提取。

1.2 主要试剂

Trizol Reagent 为 Invitrogen 公司产品; PCR 产物纯化试剂盒为北京百泰克生物技术公司产品; 细菌培养用胰蛋白胨、酵母提取物及细菌培养用琼脂为美国 BD 公司产品; 琼脂糖为 Bio-Rad 公司产品; 逆转录酶 ReverTra Ace 和 RNA 酶抑制剂 RNase Inhibitor 为日本 TOYOBO 公司产品; *Taq* 酶、*E. coli* DH 5 α 感受态细胞和 pBS-T 载体购自北京天根生物技术有限公司; DNA marker 购自 TaKaRa(大连)有限公司; 核酸染料 Gold ViewTM 为北京赛百盛基因技术有限公司产品; SMARTTM RACE 试剂盒为美国 CLOTECH 公司产品; PCR 引物由 Invitrogen(中国)公司合成。

1.3 方法

1.3.1 总 RNA 提取 取鳊鱼幽门垂组织于液氮下迅速冷冻研磨至粉末状。取 0.1g 样品转入 1.5ml 离心管, 加入 1ml Trizol 试剂。组织捣碎机捣碎 1—2min, 室温放置 5min。4℃, 12000g 离心 10min 后取上清转移至新离心管。加入 0.2ml 氯仿的量加入氯仿, 颠倒震荡 15s, 室温放置 3min。4℃, 12000g 离心 10min 后, 取上清至新离心管。上清中加入等体积的异丙醇, 颠倒震荡 15s, 室温放置 10min, 4℃, 12000g 离心 10min 后, 取沉淀。将沉淀中加入 1ml 75%酒精(DEPC 水配制)洗涤。4℃, 7500g 离心 5min 后得沉淀, 加入 30 μ l 的 DEPC H₂O 溶解沉淀后, 用核酸测定仪检测总 RNA 的纯度(A_{260}/A_{280} 值)和浓度。然后, 立即进行 cDNA 合成。

1.3.2 引物设计 根据本研究室报道的鳊鱼胰蛋白酶的 N-末端氨基酸测序结果(Lu *et al.*, 2008), 在 NCBI 网站上进行 BLAST 搜索, 获得相似性较高的其它来源(主要是鱼类)的胰蛋白酶(或胰蛋白酶原)基因序列。用 GeneDoc 软件进行多序列比对, 根据比对结果, 找到保守性较高的区域, 作为下游引物设计区; 上游引物直接根据 N-末端的氨基酸序列结合鱼类密码子的偏好性设计简并引物。然后利用 Primer Premier 5 软件和 DNASTAR 软件对引物进行设计和评价,

最终确定上下游引物 TF-A、TF-B 和 TR(表 1)。其中, 利用引物 TF-A 和 TR 可扩增出胰蛋白酶 TRS-A 的 cDNA 中间片段(大小约为 550bp); 利用引物 TF-B 和 TR 可扩增出胰蛋白酶 TRS-B 的 cDNA 中间片段(大小约为 500bp)。

根据胰蛋白酶 TRS-A 和 TRS-B 的 cDNA 中间片段测序结果, 设计 3'-RACE 的基因特异性引物 GSPF-A 和 GSPF-B 及 5'-RACE 的基因特异性引物 GSPR-A 和 GSPR-B(表 1); 引物 5'-RACE CDS Primer A、SMART IITM A Oligonucleotide、5'-UPM、3'-RACE CDS Primer A 及 3'-UPM 由 SMARTTM RACE 试剂盒提供。其中, 利用引物 GSPF-A 和 3'-UPM 可扩增出胰蛋白酶 TRS-A 的 3'-末端序列(大小约为 530bp); 利用引物 GSPF-B 和 3'-UPM 可扩增出胰蛋白酶原 TRS-B 的 3'-末端序列(大小约为 580bp); 利用引物 5'-UPM 和 GSPR-A 可扩增出胰蛋白酶原 TRS-A 的 5'-末端序列(大小约为 350bp); 利用引物 5'-UPM 和 GSPR-B 可扩增出胰蛋白酶原 TRS-B 的 5'-末端序列(大小约为 300bp)。

1.3.3 鳊鱼胰蛋白酶的 cDNA 中间片段扩增 采用 RT-PCR 方法扩增鳊鱼胰蛋白酶 TRS-A 和 TRS-B 的 cDNA 中间片段。反转录体系(20 μ l): 总 RNA 1.5 μ l, 3'-RACE CDS Primer A 2 μ l, 5 \times RT PCR Buffer 4 μ l, dNTP Mixture (10mol/L) 2 μ l, ReverTra Ace (100U/ μ l) 1 μ l, RNase Inhibitor (40U/ μ l) 0.5 μ l, DEPC H₂O 9 μ l。反应条件为 42℃ 温育 30min。PCR 反应体系(50 μ l): 以上反转录产物(cDNA) 2.0 μ l, TF-A/TF-B (10pmol/ μ l) 2.0 μ l, TR(10pmol/ μ l) 2.0 μ l, dNTP Mixture (10mol/L) 4.0 μ l, 10 \times Ex *Taq* buffer 5.0 μ l, Ex *Taq* 酶 0.25 μ l, 无菌水 34.75 μ l。反应条件为: 94℃ 预变性 5min 后, 94℃ 变性 40s, 53℃ 退火 30s, 72℃ 延伸 90s, 循环 35 次, 再于 72℃ 延伸 7min。PCR 产物于 1.0%琼脂糖凝胶中进行电泳检测。

表 1 PCR 扩增鳊鱼胰蛋白酶 cDNA 所用引物
Tab.1 Primers used in PCR amplification for trypsin cDNAs of *S. chuatsi*

引物名称	序列	备注
TF-A	5'-ATGA(G/T)GCCTCTGGTCTTCTT-3'	上游引物
TF-B	5'-ATCGT(C/T)GGAGG(A/G)TATGAGTG(C/T)-3'	上游引物
TR	5'-AGGTATCCAGC(A/G)CAGAACAT-3'	下游引物
GSPF-A	5'-ACATCATGCTGATCAAGCTGA-3'	上游引物
GSPF-B	5'-TTCAACCGCTACACCCTTGAG-3'	上游引物
GSPR-A	5'-TCAGTTTGAGCAGCATGATGT-3'	下游引物
GSPR-B	5'-AGGGTGTAGCGGTTGAAGGA-3'	下游引物

1.3.4 采用 3'-RACE 扩增鳊鱼胰蛋白酶 cDNA 的 3'-端序列 第一链 cDNA 合成的反转录体系及反应条件与 1.3.3 相同。然后采用 Nested-PCR 扩增鳊鱼胰蛋白酶 cDNA 的 3'-端序列。其中, TRS-A 的 3'-RACE 以第一链 cDNA 为模板, 使用上游引物 TF-A 和下游引物 3'-UPM 进行第一轮 PCR 扩增($T_m=57$), 取第一次扩增产物作为 Nested-PCR 的模板, 使用引物 GSPF-A/3'-UPM 进行 Nest-PCR。具体 PCR 反应体系与反应条件与 1.3.3 相同; TRS-B 的 3'-RACE 以第一链 cDNA 为模板, 使用引物 TF-B/3'-UPM 进行第一轮 PCR 扩增($T_m=57$), 取第一次扩增产物作为 Nested-PCR 模板, 使用引物 GSPF-B/3'-UPM 进行第二轮扩增。具体 PCR 反应体系与条件与 1.3.3 相同。

1.3.5 采用 5'-RACE 扩增鳊鱼胰蛋白酶 cDNA 的 5'-端序列 鳊鱼胰蛋白酶 cDNA 的 5'-RACE 根据 CLOTECH 公司提供的 SMARTTM RACE 试剂盒使用说明稍作改动。反应体系 1 (10 μ l) 中含有总 RNA 1.5 μ l, 5'-RACE CDS Primer A (12 μ mol/L) 2 μ l, SMART IITM A Oligonucleotide (12 μ mol/L) 2 μ l, DEPC 水 4.5 μ l; 反应体系 2 (10 μ l) 中含有 5 \times RT PCR Buffer 4 μ l, dNTP Mixture (10mol/L) 2 μ l, ReverTra Ace (100U/ μ l) 1 μ l, RNase Inhibitor (40U/ μ l) 0.5 μ l, DEPC 水 2.5 μ l。首先将反应体系 1 在 70 $^{\circ}$ C 温育 30min 后, 冰上放置 2min; 然后将反应体系 1 与反应体系 2 混合后, 在 42 $^{\circ}$ C 温育 90min, 即可获得用于 5'-RACE 的第一链 cDNA 模板。采用 Nested-PCR 扩增鳊鱼胰蛋白酶 cDNA 的 5'-端序列。其中, TRS-A 的 5'-RACE 以第一链 cDNA 为模板, 使用上游引物 5'-UPM/TR 进行第一轮扩增($T_m=63$), 取第一次扩增产物作为 Nested-PCR 的模板, 使用引物 5'-UPM/GSPR-A 进行 Nest-PCR ($T_m=55$)。具体 PCR 反应体系和条件与 1.3.3 相同; TRS-B 的 5'-RACE 以第一链 cDNA 为模板, 使用上游引物 5'-UPM/TR 进行第一轮扩增($T_m=63$), 取第一次扩增产物作为 Nested-PCR 的模板, 使用引物 5'-UPM/GSPR-B 进行 Nest-PCR ($T_m=55$)。具体 PCR 反应体系和条件与 1.3.3 相同。

1.3.6 PCR 产物克隆及测序 将以上 RT-PCR 产物、3'-RACE 及 5'-RACE 扩增产物分别用 1.0% 琼脂糖凝胶电泳分离、切胶后, 用 PCR 产物纯化试剂盒对目的基因片段进行回收纯化。将纯化后的 DNA 片段与 T 载体 pBS-T 连接构建重组克隆质粒, 转化感受态大肠杆菌 DH5 α , 在含 4 μ l IPTG (浓度为 200mg/ml)、40 μ l X-gal (浓度为 20mg/ml) 及 100mg/L 氨苄青霉素

的 LB 平板上挑选白色菌落, 经菌落 PCR 鉴定后, 送 Invitrogen (中国) 公司测序。

1.3.7 生物信息学分析 应用 DNASTar 软件进行开放阅读框 (ORF) 分析并将其推导为相应的氨基酸序列; 通过 NCBI (<http://www.ncbi.nlm.nih.gov>) BLAST 进行核苷酸和蛋白质序列相似性搜索 (鲁双庆等, 2008); 信号肽分析利用 Signal P3.0 server 程序 (<http://www.cbs.dtu.dk/service/SignalP>); 氨基酸序列比对应应用 GeneDoc 软件及 Clustal W (<http://www.ebi.ac.uk/clustalw/index.html>) 进行; 应用软件 MEGA 4.1 构建基于胰蛋白酶氨基酸序列的系统进化树 (Larsen *et al*, 2008); 蛋白质的分子量、等电点及基本性质分析用 ProtParam (<http://www.expasy.ch/tools/protparam.html>) 预测 (孙之荣, 2006); 采用 GOR IV (http://npsa-pbil.ibcp.fr/cgi-bin/secpred_gor4.pl) 预测鳊鱼胰蛋白酶的二级结构; 利用 Swiss PdbViewer 软件, 选择 Project Mode, 以基于 SWISS-MODEL 服务器的同源建模方法 (Larsen *et al*, 2008), 构建鳊鱼胰蛋白酶的三级结构模型, 利用 PyMol 软件对模型进行相关分析及显示。

2 结果与分析

2.1 鳊鱼胰蛋白酶 cDNA 的克隆

2.1.1 鳊鱼胰蛋白酶 cDNA 中间片段扩增 从鳊鱼幽门垂组织中提取总 RNA 后, 测定其浓度为 3000ng/ μ l, $A_{260nm}/A_{280nm}=1.88$, 说明所提取的总 RNA 纯度和浓度均可作为 RT-PCR 的模板。首先以总 RNA 为模板、以 3'-RACE CDS Primer A 为引物进行反转录获得 cDNA 模板。然后以引物 TF-A/TR 和 TF-B/TR 分别扩增胰蛋白酶 TRS-A 和 TRS-B 的 cDNA 中间片段, 分别获得了约 550bp 和 500bp 左右的 DNA 片段 (图 1)。将以上扩增片段分别进行 T-克隆后测序。结果表明, TRS-A 和 TRS-B 的 cDNA 中间片段大小分别为 554bp 和 494bp, 与预期相符。

2.1.2 鳊鱼胰蛋白酶 cDNA 的 3'末端扩增 以 3'-RACE 第一链 cDNA 为模板, 以 RACE 引物 GSPF-A/3'-UPM 和 GSPF-B/3'-UPM 扩增 TRS-A 与 TRS-B 的 3'-末端序列, 分别获得了位于 500—600bp 和 500—750bp 之间的 DNA 片段 (图 2)。将以上扩增片段分别进行 T-克隆后测序。结果表明, TRS-A 和 TRS-B 的 3'-末端序列大小分别为 536bp 和 577bp, 与预期相符。

2.1.3 鳊鱼胰蛋白酶 cDNA 的 5'末端扩增 以

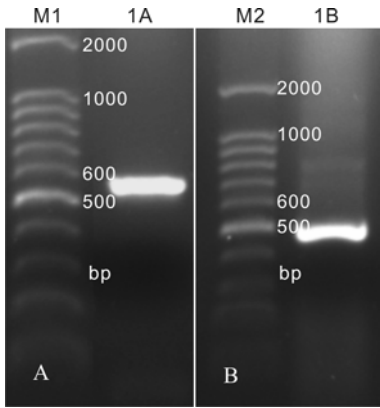


图1 TRS-A(A)与 TRS-B(B)的 RT-PCR 产物琼脂糖凝胶(1.0%)电泳检测图

Fig.1 Agarose gel (1.0%) electrophoresis analysis of RT-PCR products of TRS-A (A) and TRS-B (B)
注: M1: 100bp DNA Ladder; 1A: RT-PCR product of TRS-A; M2: 100bp DNA Ladder; 1B: RT-PCR product of TRS-B

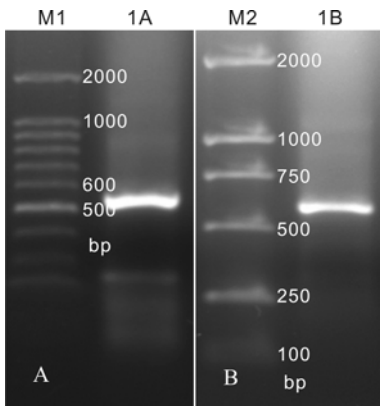


图2 TRS-A(A)与 TRS-B(B)的 3'-RACE 产物琼脂糖凝胶(1.0%)电泳检测图

Fig.2 Agarose gel (1.0%) electrophoresis analysis of 3'-RACE products of TRS-A (A) and TRS-B (B)
注: M1: 100bp DNA Ladder; 1A: 3'-RACE product of TRS-A; M2: DNA Marker DL2000; 1B: 3'-RACE product of TRS-B

和 ACD70340。其中, TRS-A 的 cDNA 序列大小为 869bp, 包括一个 23bp 的 5'非编码区(Untranslated Region, UTR)、729bp 开放阅读框(编码 242 个氨基酸残基)和 117bp 的 3'非编码区(在 polyA 尾上游 15 核苷酸处存在加尾信号肽 AATAAA); TRS-B 的 cDNA 序列大小为 885bp, 包括一个 23bp 的 5'-UTR, 729bp 开

5'-RACE 第一链 cDNA 为模板, 以 RACE 引物 5'-UPM/GSPR-A 和 5'-UPM/GSPR-B 进行 5'-RACE 扩增 TRS-A 与 TRS-B 的 5'-末端序列, 分别获得了位于 300—400bp 和 250—500bp 之间的 DNA 片段(图 3)。将以上扩增片段分别进行 T-克隆后测序。结果表明, TRS-A 和 TRS-B 的 5'-末端序列大小分别为 354bp 和 304bp, 与预期相符。

2.2 鳓鱼胰蛋白酶 cDNA 序列及其氨基酸序列分析

将以上 RT-PCR、3'-RACE 和 5'-RACE 扩增的三个片段进行拼接后分别获得鳓鱼胰蛋白酶 TRS-A 和 TRS-B 基因的全长 cDNA 序列。将其递交 GenBank 数据库, 分别获得登录号为 EU688996 和 EU688997, 推导的氨基酸序列登录号为 ACD 70339

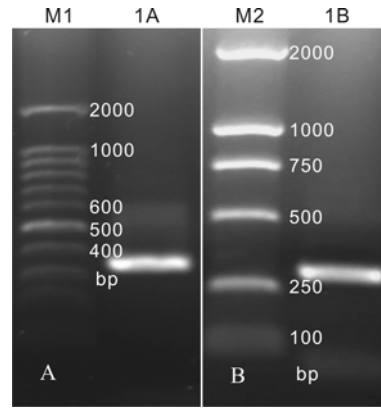


图3 TRS-A(A)与 TRS-B(B)的 5'-RACE 产物琼脂糖凝胶(1.0%)电泳检测图

Fig.3 Agarose gel (1.0%) electrophoresis analysis of 5'-RACE products of TRS-A (A) and TRS-B (B)
注: M1: 100bp DNA Ladder; 1A: 5'-RACE product of TRS-A; M2: DNA Marker DL2000; 1B: 5'-RACE product of TRS-B

AFA, 说明本文获得的两个 729bp 开放阅读框所编码的氨基酸序列均为胰蛋白酶原前体(Pretrypsinogen), 去掉信号肽后即可得到鳓鱼胰蛋白酶原(含有 227 个氨基酸残基)。将其与通过蛋白质纯化后得到的鳓鱼胰蛋白酶的 N-末端氨基酸测序结果(Lu *et al*, 2008)比对发现, TRS-A 的 N-末端的 21—30 氨基酸序列与成熟蛋白 N-末端测序结果完全一致, TRS-B 的 N-末端的 21—27 氨基酸序列与成熟蛋白 N-末端测序的前 7 个氨基酸序列一致。说明获得的核苷酸序列确实为鳓鱼胰蛋白酶 cDNA 序列。进一步结合相关文献(黎军胜等, 2006)报道, 可推测其激活肽序列由 5 个氨基酸残基组成(均为 TEDDK), 由此推测鳓鱼胰蛋白酶 TRS-A 和 TRS-B 蛋白序列均含有 222 个氨基酸残基。

2.3 鳓鱼胰蛋白酶原氨基酸序列相似性比对及其系统进化关系分析

将鳓鱼胰蛋白酶 TRS-A 和 TRS-B 氨基酸序列与其它来源的胰蛋白酶序列进行相似性比对。结果(表 2)表明, 鳓鱼胰蛋白酶 TRS-A 与 TRS-B 的氨基酸序列相似性为 85%; TRS-A 与其它鱼类胰蛋白酶的相似性为 81%—90%、与其它脊椎动物(如人、鼠、牛、狗、鸡等)的相似性为 60%—64%; TRS-B 与其它鱼类胰蛋白酶的相似性为 76%—81%、与其它脊椎动物(如人、鼠、牛、狗、鸡等)的相似性为 60%—64%。

使用软件 MEGA 4.1 对不同物种基于胰蛋白酶原

放阅读框(编码 242 个氨基酸残基)和 133bp 的 3'非编码区(在 polyA 尾上游 15 核苷酸处存在加尾信号肽 AAT- AAA)。将其氨基酸序列在网站 ([http:// www.cbs.dtu.dk/se-rvices/Signal P/](http://www.cbs.dtu.dk/se-rvices/SignalP/)) 上进行信号肽预测, 结果表明, 其 N 末端均含有一段信号肽序列(1—15 氨基酸残基), 其中 TRS-A 为 MRS-LVFVLLIG AAFA, TRS-B 为 MRPLVFFLLIGA

表 2 鳊鱼胰蛋白酶原与 GenBank 中其它来源的胰蛋白酶原氨基酸序列的相似性比对
Tab.2 The amino acid sequence alignments of trypsinogens from *S. chuatsi* and some other organisms

物种	GenBank 登录号	与 TRS-A 相似度%	与 TRS-B 相似度%
<i>Siniperca chuatsi</i> TRS-A	ACD70339	100	85
<i>Siniperca chuatsi</i> TRS-B	ACD70340	85	100
<i>Salmo salar</i> trypsin IA	CAA49676	90	81
<i>Paralichthys olivaceus</i> trypsinogen 1	BAA82362	87	78
<i>Paranotothenia magellanica</i> trypsin	CAA57701	86	78
<i>Sparus aurata</i> trypsinogen II	ABE68639	86	78
<i>Pseudopleuronectes americanus</i> trypsinogen 2	AAC32752	85	76
<i>Engraulis japonicus</i> trypsinogen	BAB40329	83	76
<i>Engraulis japonicus</i> trypsinogen II	BAB40330	81	76
<i>Mus musculus</i> trypsinogen 7	AAB69044	64	63
<i>Rattus norvegicus</i> trypsin II	P00763	64	64
<i>Canis lupus familiaris</i> Anionic trypsin	P06872	64	63
<i>Homo sapiens</i> trypsinogen	AAA61232	61	61
<i>Gallus gallus</i> trypsinogen	AAA79913	61	60
<i>Bos taurus</i> anionic trypsin	NP_777115	60	60

氨基酸序列构建系统进化树(图 4)。结果表明同属内种间关系最近, 同目内科间亲缘关系次之, 不同目间的亲缘关系最远。胰蛋白酶首先从细菌(Bacteria)中进化而来, 然后向鱼类分化, 最后再进化出陆生脊椎动物及哺乳动物。鳊鱼胰蛋白酶 TRS-A 和大西洋鲑鱼(*Atlantic salmon*)亲缘关系最为接近。与其它鱼类胰蛋白酶相比, 鳊鱼胰蛋白酶 TRS-B 在进化上表现的更为高级, 它和陆生脊椎动物胰蛋白酶亲缘关系更为紧密。由于鳊鱼是一种淡水鱼类, 该结果也提示淡水鱼在进化上比海水鱼更为高级。

2.4 鳊鱼胰蛋白酶的理化参数及其结构预测

利用 ProtParam 预测了鳊鱼成熟胰蛋白酶(222 个氨基酸残基)的理化参数: TRS-A 相对分子质量为 24.1kDa, 等电点为 5.96, 理论半衰期大于 20h, 不稳定参数为 40.51, 属于不稳定蛋白; TRS-B 相对分子质量为 24.2kDa, 等电点为 5.75, 理论半衰期大于 20h, 不稳定参数为 41.64, 属于不稳定蛋白。采用 GOR IV 预测鳊鱼胰蛋白酶(222 个氨基酸残基)的二级结构中 包含 螺旋(Alpha helix)、延伸带(Extended strand)和随机卷曲(Random coil)三种形式, 其中, TRS-A 分别占 4.05%、32.43%和 63.51%; TRS-B 分别占 9.01%、34.23%和 56.76%。

为了进一步获得鳊鱼胰蛋白酶的结构与功能的信息, 从该蛋白的氨基酸序列(222 个氨基酸残基)出发, 在蛋白数据库(PDB)中进行 BLAST 搜索, 结果发现, 该胰蛋白酶的氨基酸序列与来自大西洋鲑鱼的胰蛋白酶(PDB 编码为 1HG8)分别具有 90%(TRS-A)

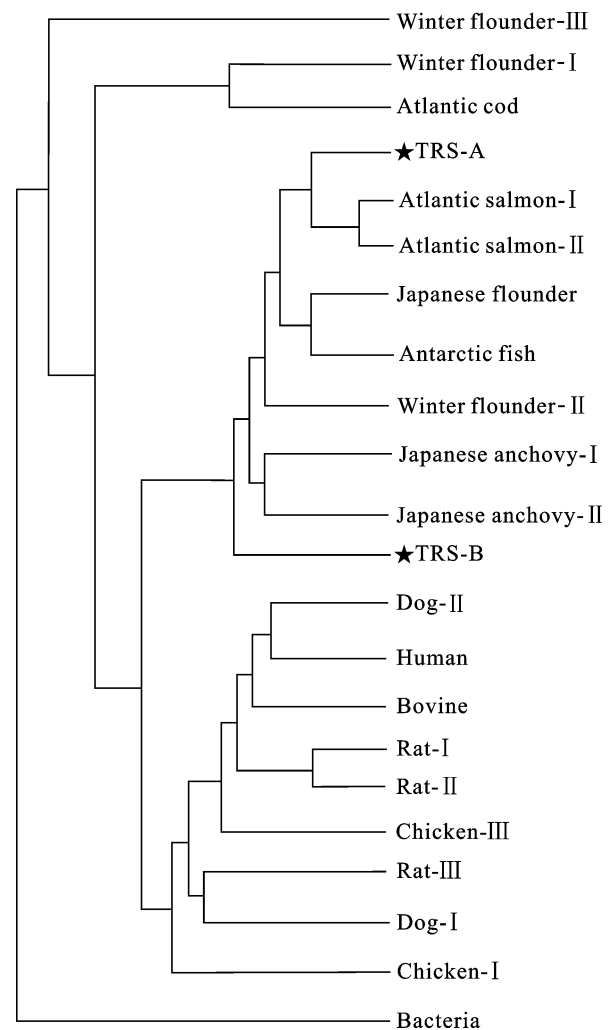


图 4 基于胰蛋白酶氨基酸序列的系统进化树(NJ 法)
Fig.4 Rooted NJ tree based on the amino acid sequences of the trypsins from different organisms

和 81%(TRS-B)的序列一致性。利用 Swiss PdbViewer 软件和 SWISS-MODEL 服务器开展模型构建工作。采用 1HG8 作为模板序列, 构建了鳊鱼胰蛋白酶 TRS-A 和 TRS-B 的三级结构模型。结果表明, 所预测的鳊鱼胰蛋白酶三级结构(图 5A、图 5B)与模板蛋白(图 5C)的非常相似, 因此, 可根据模板蛋白的催化位点(Leiros *et al*, 2001), 推测鳊鱼胰蛋白酶 TRS-A 和 TRS-B 的催化活性中心均由 His40、Asp84 和 Ser176 三个氨基酸残基组成; 而维持其空间结构的 6 个二硫键分别由 Cys7 和 Cys136、Cys25 和 Cys41、Cys109 和 Cys209、Cys116 和 Cys182、Cys147 和 Cys161、Cys172 和 Cys196 形成(图 5A、图 5B)。

3 讨论

脊椎动物胰蛋白酶(原)是一个多基因家族。其中, 哺乳动物胰蛋白酶是胰蛋白酶酶学研究中最为充分的, 所研究的物种包括人(Emi *et al*, 1986)、牛(Huerou *et al*, 1990)、狗(Pinsky *et al*, 1985)等。鱼类如大西洋鳕鱼(Atlantic Cod)(Spilliaert *et al*, 1999)、比目鱼(winter flounder)(Douglas *et al*, 1998)和大西洋鲑鱼(Atlantic salmon)(Rune *et al*, 1995)等胰蛋白酶的基因也被克隆。一些胰蛋白酶(原)的基因序列已登录到 GenBank 数据库, 这为其它鱼类的胰蛋白酶(原)的基因克隆提供了重要的信息。

在前期的工作中, 本课题组从鳊鱼幽门垂组织中分离纯化得到两种胰蛋白酶, 分别为阳离子型和阴离子型(Lu *et al*, 2008)。其中, 阴离子型胰蛋白酶成功测得了 N-末端 10 个氨基酸残基序列, 这为本研究提供了重要参考。RACE(Rapid Amplification of cDNA Ends)是近年来广泛采用的扩增完整 cDNA 的方法, 与常规通过建立 cDNA 文库进行筛选的方法相比, 它

具有高效、经济的优点。本研究以鳊鱼胰蛋白酶的 N-末端氨基酸测序结果及胰蛋白酶保守位点为依据设计引物, 利用 RT-PCR、3'-RACE 和 5'-RACE 的方法得到了两条完整的鳊鱼胰蛋白酶 TRS-A 与 TRS-B 的 cDNA 序列, 它们的开放阅读框大小均为 729bp 用来编码 242 个氨基酸残基, 其 N-末端含有一段由 15 个氨基酸残基组成的信号肽和 5 个氨基酸残基组成的激活肽, 因此, 编码成熟蛋白区均由 222 个氨基酸残基组成。通过相似性比对发现所克隆的鳊鱼胰蛋白酶 TRS-A 与 TRS-B 的 cDNA 在开放阅读框上的相似性为 88%, 所推导的氨基酸序列相似性为 85%, 这表明 TRS-A 与 TRS-B 在长期的选择性进化过程中可能发生了不均一的核酸组成转变。

将鳊鱼胰蛋白酶原氨基酸序列与其它物种的胰蛋白酶原氨基酸序列进行相似性比对并构建系统进化树, 结果显示, 胰蛋白酶首先从细菌(Bacteria)中进化出来, 然后向鱼类分化, 最后再进化出陆生脊椎动物及哺乳动物; 鳊鱼胰蛋白酶 TRS-A 与大西洋鲑鱼(Atlantic salmon) trypsin IA 最先聚类, 亲缘关系最近; 鳊鱼胰蛋白酶 TRS-B 在进化上表现的更为高级, 它与陆生脊椎动物胰蛋白酶亲缘关系更为紧密。

根据鳊鱼胰蛋白酶成熟蛋白氨基酸序列, 利用相关生物信息学软件预测其理论分子量 TRS-A 为 24.1kDa、等电点为 5.96; TRS-B 为 24.2kDa、等电点为 5.75。由此说明本文克隆的两个鳊鱼胰蛋白酶 TRS-A 与 TRS-B 均为酸性蛋白(即阴离子型胰蛋白酶)。将以上克隆的核苷酸序列翻译成相应的氨基酸序列后, 与从鳊鱼中纯化得到的胰蛋白酶 N-末端测序结果进行比对发现, 胰蛋白酶 TRS-A 的 N-末端的 10 个氨基酸残基与纯化蛋白的 N-末端测序结果完全一致; 而 TRS-B 的 N-末端序列只有 7 个氨基酸残基

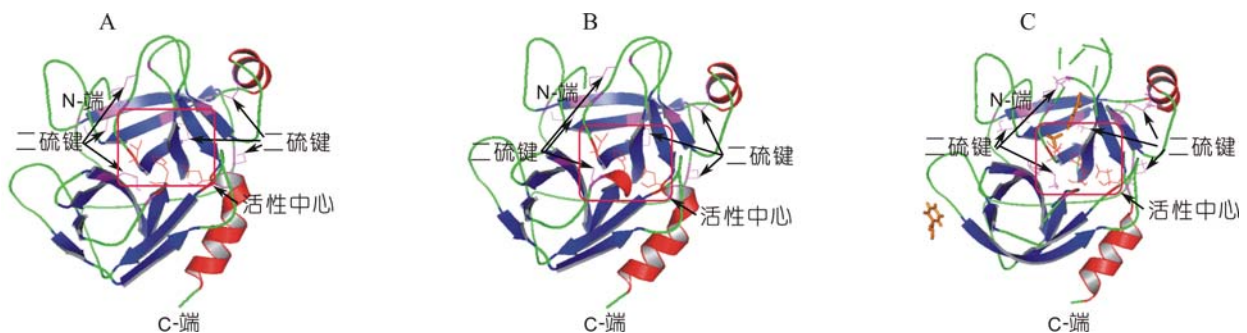


图 5 利用 1HG8 蛋白三级结构(C)为模板构建鳊鱼胰蛋白酶 TRS-A(A)和 TRS-B(B)三级结构模型

Fig.5 The predicted three-dimensional structures of TRS-A (A) and TRS-B (B) from *S. chuatsi* using SWISS-MODEL and 1HG8 as template (C)

与纯化蛋白的 N-末端测序结果一致。从而推测 TRS-A 可能为调控从鳊鱼中纯化的阴离子型胰蛋白酶蛋白合成的基因。但纯化的蛋白分子量为 21.5kDa 略小于其理论分子量 24.1kDa。分析原因可能是在 TRS-A 的活性中心 GDSGGP 之后的 C-末端的潜在酶切位点 199 上被胰蛋白酶自身水解所致。因为经过该位点水解后 TRS-A 缺失了其 C-末端肽段 NHPGVYAKVCL-FNEWLESTMASY, 该片段的理论分子量为 2.65kDa, 这样剩下 TRS-A 的活性主体分子量为 21.45kDa, 这与本研究室前期纯化得到的 21.5kDa 左右的分子量一致(Lu *et al.*, 2008)。说明鳊鱼胰蛋白酶在纯化的过程中发生自身降解, 从而导致纯化得到的胰蛋白酶分子量偏低。鳊鱼体内阳离子型胰蛋白酶基因序列已有文献报道(陈亮等, 2009), 其激活肽序列由 9 个氨基酸组成, 等电点为 7.94, 所推测的氨基酸序列与本研究获得的两个胰蛋白酶 TRS-A 与 TRS-B 的一致性分别为 67%和 66%, 而 TRS-A 与 TRS-B 的激活肽序列均由 5 个氨基酸组成, 它们的等电点分别为 5.96 和 5.75。由此可以说明, 本文所获得的两个鳊鱼胰蛋白酶与文献报道的是两种不同类型的胰蛋白酶, 从而进一步说明了胰蛋白酶在生物体内存在多样性(Rune *et al.*, 1995; Lu *et al.*, 2008)。

目前, 三级结构的预测方法主要有同源建模、折叠类型识别和从头预测(孙之荣, 2006)。预测过程中一般先将目标序列(建模序列)与结构数据库中的数据匹配, 寻找与已知结构有明显同源性的序列, 如建模序列和模板序列匹配后相似性在 30%以上, 则可以用同源蛋白建模的方法预测其三维结构。如序列相似性小于 30%, 建立理想的同源模板比较困难, 则用折叠模式识别方法来寻找远源的同源性或类似的折叠模式; 如果找不到同源或类似结构, 对于小分子蛋白则可尝试二级结构堆积计算或简化模型的从头预测方法。同源建模是目前最为成功及实用的蛋白质结构预测方法, 可在一定程度上用于解释实验数据、解释功能作用、进行突变体设计和药物设计等。SWISS-MODEL 是国际上同源建模最常用的工具(Arnold *et al.*, 2006)。本研究采用同源建模方法, 以源于大西洋鲑鱼的胰蛋白酶 1HG8 的三级结构为模板(与 TRS-A 和 TRS-B 序列一致性分别为 90%和 81%), 构建了鳊鱼胰蛋白酶成熟蛋白的三级结构模型。大西洋鲑鱼的胰蛋白酶已被纯化, 通过 X-衍射得到了其空间结构, 其催化活性中心由 His55、Asp99 和 Ser191 氨基酸残基组成; 而维持其空间结构的 6 个二硫键分别由

Cys22 和 Cys156、Cys42 和 Cys58、Cys127 和 Cys232、Cys136 和 Cys201、Cys168 和 Cys182、Cys191 和 Cys220 形成(Leiros *et al.*, 2001)(图 5C), 由此可以推测鳊鱼胰蛋白酶原的活性中心所对应的氨基酸残基组成及 12 个必需半胱氨酸残基的位置, 从而为今后研究其结构与功能关系及定点突变提供了有用的信息。

参 考 文 献

- 王镜岩, 朱圣庚, 徐长法, 2002. 生物化学(第三版上册). 北京: 高等教育出版社, 423
- 刘秀霞, 梁旭方, 王琳等, 2009. 鳊鱼(*Siniperca chuatsi*) - 肌动蛋白基因 cDNA 全序列与 5'侧翼区的克隆与分析. 海洋与湖沼, 40(1): 102—108
- 刘欣华, 刘弈志, 夏朝霞, 2002. 胰蛋白酶预防后发性白内障的实验研究. 眼外伤职业眼病杂志, 24(4): 369—370
- 孙之荣, 2006. 生物信息学与功能基因组学. 北京: 化学工业出版社, 223—296
- 陈亮, 梁旭方, 王琳等, 2009. 鳊鱼胰蛋白酶和淀粉酶与胃蛋白酶原基因的克隆与序列分析. 中国生物化学与分子生物学报, 25(12): 1115—1123
- 鲁双庆, 刘峰, 刘臻, 2008. 三种鳊鱼(*Siniperca*)生长激素基因的克隆及序列比较. 海洋与湖沼, 39(4): 354—361
- 黎军胜, 李建林, 吴婷, 2006. 罗非鱼胰蛋白酶 I mRNA 克隆与分析. 水生生物学报, 30(2): 200—213
- Arnold K, Bordoli L, Kopp J *et al.*, 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics, 22(2): 195—201
- Chen J M, Kukor Z, Marechal C L *et al.*, 2003. Evolution of trypsinogen activation peptides. Mol Biol Evol, 20(11): 1767—1777
- Douglas S E, Gallant J W, 1998. Isolation of cDNAs for trypsinogen from the winter flounder, *Pleuronectes americanus*. J Mar Biotechnol, 6(4): 214—219
- Emi M, Nakamura Y, Ogawa M *et al.*, 1986. Cloning, characterization and nucleotide sequences of two cDNAs encoding human pancreatic trypsinogens. Gene, 41(2—3): 305—310
- Gudmundsdottir A, Palsdottir H M, 2005. Atlantic cod trypsins: from basic research to practical application. Mar Biotech, 7(2): 77—88
- Huerou I L, Wicker C, Guilloteau P *et al.*, 1990. Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsinogen: Structural identity within the trypsin family. Eur J Biochem, 193(3): 767—773
- Larsen J B, Larsen A, Bratbak G *et al.*, 2008. Phylogenetic analysis of members of the phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. Appl Environ Microbiol, 74(10): 3048—3057
- Leiros H K, McSweeney S M, Smalas A O, 2001. Atomic resolution structures of trypsin provide insight into structural ra-

- diation damage. *Acta Crystallogr D*, 57(4): 488—497
- Lu B J, Zhou L G, Cai Q F *et al*, 2008. Purification and characterization of trypsins from the pyloric caeca of mandarin fish (*Siniperca chuatsi*). *Food Chem*, 110: 352—360
- Pinsky S D, LaForge K S, Scheele G, 1985. Differential regulation of trypsinogen mRNA translation: full-length mRNA sequences encoding two oppositely charged trypsinogen isoenzymes in the dog pancreas. *Mol Cell Biol*, 5(10): 2669—2676
- Rune M, James B L, Arne O S *et al*, 1995. Molecular cloning and characterization of anionic and cationic variants of trypsin from Atlantic salmon. *Eur J Biochem*, 232: 677—685
- Smalas A O, Heimstad E S, Hordvikn A *et al*, 1994. Cold adaptation of enzymes: structural comparison between salmon and bovine trypsins. *Proteins*, 20(2): 149—166
- Spilliaert R, Gudmundsdottir A, 1999. Atlantic cod trypsin Y-member of a novel trypsin group. *Mar Biotechnol*, 1(6): 598—607

CLONING AND BIOINFORMATIC ANALYSIS OF TWO TRYPSIN GENES FROM MANDARIN FISH *SINIPERCA CHUATSI*

DU Cui-Hong, SU Wen-Jin, LU Bao-Ju, LIU Guang-Ming, QIU Xiao-Yan, CAO Min-Jie
(College of Biological Engineering, the Key Laboratory of Science and Technology for Aquaculture and Food Safety, Jimei University, Xiamen, 361021)

Abstract The cDNAs of two isoforms of trypsins, designated as TRS-A and TRS-B, were isolated from the pyloric caeca of mandarin fish *Siniperca chuatsi* through RT-PCR, 3'-RACE and 5'-RACE. Sequence analysis indicated that both trypsin cDNAs had an open reading frame of 729bp encoding 242 amino acid residues. The two sequences had been deposited into GenBank under accession number of ACD70339 and ACD70340. Alignment of two amino acid sequences for fish trypsinogens and construction of the corresponding phylogenetic trees were performed based on amino acid sequences of the prepro-forms (pretrypsinogens). The results suggested that TRS-A was more closely related to trypsin IA from Atlantic salmon (*Salmo salar*) while TRS-B was more closely related to trypsins from vertebrates lives on land. Some physical and chemical parameters and three-dimensional structures of the two trypsins were predicted by bioinformatics. The results indicated that: (1) both the amino acid sequences of the prepro-forms of TRS-A and TRS-B were composed of a signal peptide (15 residues), a propeptide (5 residues) and a trypsin moiety (222 residues); (2) TRS-A had a molecular mass of 24.1kDa and its theoretical pI was 5.96. TRS-B had a molecular mass of 24.2kDa and its theoretical pI was 5.75; (3) the three-dimensional structures of TRS-A and TRS-B were constructed by homology modeling to predict their active sites and disulfide bonds formed by twelve cysteine residues. This study provides an insight into the relationship between structure and function of fish trypsins and established a foundation for further research in fish trypsins using site-directed mutagenesis.

Key words *Siniperca chuatsi*, Trypsin, Gene cloning, Bioinformatics