

ĐẶC ĐIỂM PHÂN TỬ PROTEASE 3 LỚP SERINE LIÊN KẾT MANNAN Ở LỢN

Đỗ Võ Anh Khoa¹, Klaus Wimmers²

¹Dại học Cần Thơ

²Viện Sinh học Gia súc, FBN, Dummerstorf, Đức

TÓM TẮT

Gen mã hóa protease lớp serine liên kết mannan (MASP) đóng vai trò quan trọng trong cơ chế miễn dịch ban đầu của vật chủ, đặc biệt là cơ chế hoạt hóa phức hợp bổ thể theo con đường lectin và đường nhánh MASP gồm 3 gen phân biệt *MASPI*, *MASP2* và *MASP3*. Trong nghiên cứu này, giải mã trình tự chuỗi cDNA và phân tích cấu trúc phân tử MASP3 (*MASPI* isoform 2) ở lợn được thực hiện. cDNA của *MASP3* dài 2657 bp chứa 11 exon mã hóa 728 amino acid. Sự tương đồng cao ở cDNA (87-90%) và protein (92-94%) được tìm thấy giữa lợn, chó, người và bò. Đại phân tử MASP3 chứa 27 cysteine, tuy nhiên thành phần amino acid chiếm tỷ lệ cao trong cấu trúc protein bậc 1 của nó gồm có Ser (8,4%), Val (8,1%), Leu (8,0%), Gly (7,6%), và Glu (7,1%). Với khối lượng phân tử 81,55 kDa, MASP3 là một tập hợp của nhiều miền protein chức năng giàu cysteine như CUB1 (aa 18-138, 2 cysteine), EGF_CA (aa 139-182, 6 cysteine), CUB2 (aa 185-297, 3 cysteine), CCP1 (aa 301-362, 4 cysteine), CCP2 (aa 367-432, 4 cysteine) và Tryp_SPc (aa 449-711, 5 cysteine). Trong đó, miền Tryp_SPc lớn nhất (263 aa) và EGF_CA là miền nhỏ nhất (44 aa) nhưng là miền giàu cysteine nhất (6 cysteine). Ngoài ra, signal peptide (aa 1-19), 13 cầu nối disulfide, 5 vị trí N-glycosylation và methionine loop (aa 630-649) cũng đã được nhận diện. Đây là tiền đề cho những nghiên cứu xa hơn về vai trò của gen *MASP3* trong cơ chế đáp ứng miễn dịch ở lợn.

Từ khóa: gen *MASP3*, cDNA, protein *MASP3*, đặc điểm phân tử

DAT VẤN ĐỀ

Phức hợp bổ thể đóng vai trò quan trọng trong cơ chế miễn dịch ban đầu (innate immunity) của vật chủ chống lại mầm bệnh xâm nhiễm. Sự kích hoạt phức hợp bổ thể được thực hiện theo 3 con đường khác nhau: cổ điển (classical pathway), nhánh (alternative pathway) và lectin (lectin pathway). Mannose-binding lectin (MBL) và ficolins (L-ficolin và H-ficolin) là những phân tử khởi đầu cho sự hoạt hóa con đường lectin. Trong khi đó, mannose-binding lectin associated serine protease (MASP) là enzyme xúc tác chính của con đường này nhờ vào sự kết dính của MASP với thành phần carbohydrate vi khuẩn (Takahashi *et al.*, 2008; Schwaebler *et al.*, 2002). MASP hiện diện trong huyết tương như là những zymogen (Sørensen *et al.*, 2005). 1 Lớp MASP gồm 3 serine protease (*MASPI*, *MASP2* và *MASP3*) và một protein không có chức năng enzyme (non-enzymatic protein, Map19) (Schwaebler *et al.*, 2002). *MASPI* và *MASP3* là sản phẩm cắt nối thay thế (alternative splice products) và là những gen được mã hóa từ một gen đơn *MASPI/3* (Dahl *et al.*, 2001). Mỗi gen *MASPI* hoặc *MASP3* có 2 chuỗi, chuỗi A được nhận diện chung cho cả hai gen và chuỗi B mang đặc điểm riêng của từng gen và chứa miền serine protease (Schwaebler *et al.*, 2002). Trong khi đó, Map19 là sản phẩm bị cắt xén của

gen *MASP2*. Vì vậy, cấu trúc của Map19 chỉ có hai miền giống *MASP2* và 4 amino acid thêm vào (Schwaebler *et al.*, 2002).

MASP2 là protease đóng vai trò quan trọng trong sự chia tách C4 và C2 để sản sinh ra enzyme chính của phức hợp bổ thể C3-convertase, một sản phẩm kết hợp của C4bC2b (Schwaebler *et al.*, 2002). Trong khi đó, *MASPI* có khả năng kích hoạt trực tiếp phân tử C3 (Takahashi *et al.*, 2007; Sørensen *et al.*, 2005; Matsushita *et al.*, 1998). Sự thiếu hụt *MASP2* và *MASPI/3* làm giảm hoạt động C3 trên bề mặt mannan và zymosan. Điều này có nghĩa là cả hai phân tử đều có ảnh hưởng đến sự kích hoạt phức hợp bổ thể theo con đường lectin và sự thiếu hụt *MASPI/3* có liên quan đến sự nhạy cảm với influenza virus ở chuột (Takahashi *et al.*, 2007). Khi MBL hoặc ficolin kết dính với carbohydrate trên bề mặt vi khuẩn gây kích hoạt vùng zymogen của MASP (Takahashi *et al.*, 2007) và vì thế phức hợp bổ thể được kích hoạt theo con đường lectin (lectin pathway). Hai serine protease là *MASPI* và *MASP3* liên kết với MBL hoặc ficolins để kích hoạt phức hợp bổ thể theo con đường lectin. *MASPI* đóng góp trực tiếp đến sự hoạt hóa con đường lectin thông qua sự hoạt hóa *MASP2* (Takahashi *et*

et al., 2008) và MASPI cũng đóng góp cho sự hoạt hóa con đường nhánh (alternative complement pathway) (Takahashi *et al.*, 2010). Trong khi đó, sự kích hoạt phức hợp bổ thể theo con đường nhánh có sự góp mặt của mannose-binding lectin-associated serine protease 1/3 (MASP1/3) nhờ vào sự tách rời tiền yếu tố D (pro-Df) thành Df trưởng thành (mature Df) nhờ vào sự kết nối với FCN-B hoặc 1 protein khác chưa được biết hoặc giả là MASPI có thể có chức năng như là một protein hòa tan không kết nối (Banda *et al.*, 2011)

Do vai trò quan trọng của MASP trong cơ chế miễn dịch bẩm sinh theo con đường lectin của phức hợp bổ thể, nhiều nghiên cứu cũng đã thực hiện để phân tích đặc điểm và chức năng của MASP ở các loài động vật khác nhau (Dahl *et al.*, 2001, Lynch *et al.*, 2005; Stover *et al.*, 2003). Vì vậy, trong nghiên cứu này, đặc điểm phân tử của gen MASP3 ở lợn sẽ được giải mã và phân tích.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Nghiên cứu này sử dụng nguồn vật liệu (đồng vật, hóa chất), thiết bị và phương pháp thí nghiệm (phân lập mRNA, tổng hợp cDNA, giải mã trình tự, thiết kế mồi, PCR) như được mô tả (Đỗ Võ Anh Khoa, 2010)

Năm cặp mồi đặc hiệu (Bảng 1) được thiết kế đọc theo chuỗi cDNA của gen *MASP3* (GenBank GU810083) để khuếch đại các đoạn gen tương ứng có chiều dài từ 462 đến 884 bp. Chu trình nhiệt PCR cho từng cặp mồi: 94°C/5', 35 chu kỳ (94°C/30", X°C/30", 72°C/1'), 72°C/5', ở đó X là 56°C đối với MASP3_1, 58°C đối với MASP3_4 và 60°C đối với MASP3_2, MASP3_3. Riêng đối với MASP3_5, 35 chu kỳ gồm 94°C/50", 66°C/50", 72°C/1'20".

Ngoài ra, các phần mềm online được sử dụng để phân tích kết quả, bao gồm: (1) ORF Finder (Open Reading Frame Finder) (www.ncbi.nlm.nih.gov/gorf/gorf1.html); nhận diện cấu trúc bậc 1 protein. (2) ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2/); so sánh mức độ tương đồng của chuỗi cDNA và protein. (3) SMART (smart.embl-heidelberg.de/); phân tích các miền protein chức năng. (4) SignalP 4.0 Server (www.cbs.dtu.dk/services/SignalP/); nhận diện signal peptide (5) Composition/Molecular Weight Calculation Form ([pirwww/search/comp_mw.shtml](http://pir.georgetown.edu/pirwww/search/comp_mw.shtml)); tính toán phân tử khối (6) DiANNA 1.1 web server (clavius.bc.edu/~clotelab/DiANNA/); phân tích cầu nối disulfide. (7) NetNGlyc 1.0 Server (www.cbs.dtu.dk/services/NetNGlyc/); nhận diện vị trí N-glycosylation

Bảng 1. Các cặp mồi sử dụng để khuếch đại các đoạn gen đọc theo chuỗi cDNA

Cặp mồi	Mỗi xuôi fw 5'-3' (vị trí nucleotide trên GenBank GU810083)	Mỗi ngược rev 3'-5' (vị trí nucleotide trên GenBank GU810083)	Chiều dài, bp
MASP3_1	aalgagggtggctctctctc (151-171)	tcacaggacagttctctctc (592-612)	462
MASP3_2	gagaccacggacacagaaca (431-450)	caagaagtgttgaccacgtt (900-920)	490
MASP3_3	gagctccgaatgtctctacacc (775-796)	tcctccaggacttcattcaic (1411-1430)	655
MASP3_4	gocagcagccctactacaag (1341-1360)	actgcaggacatctgacaagg (1986-2006)	666
MASP3_5	caaocctgaatttgacatcc (1774-1793)	acaacagcgcaaggagigaag (2638-2657)	884

KẾT QUẢ THẢO LUẬN

Đặc điểm phân tử cDNA và protein

Trình tự nucleotide của chuỗi cDNA gen *MASP3* là 2657 bp (GenBank GU810083, Bảng 3), trong đó vùng 5'-UTR 151 bp (1-151) và 3'-UTR 319 bp (2339-2657), codon bắt đầu ATG (nt 152-154) và codon kết thúc TGA (nt 2336-2338) (Hình 1). Giống như ở chuột (Stover *et al.*, 2003), vùng 5'-UTR của *MASP3* ở lợn chứa chuỗi khởi đầu cho quá trình phiên mã C C A C C C (nt 79-85), giống như kiểu C/T C/T A -J N T A C T C T (trong đó N là codon bắt đầu) (Javahery *et*

al., 1994). So sánh cDNA với GenBank NW_001885220 (Ssc13_WGA1011_1:115459-225459 *Sus scrofa* chromosome 13 genomic contig) (Bảng 3) đã nhận diện được 11 exon của gen *MASP3* dựa theo nguyên tắc cấu trúc DNA (exon 1-GT-intron-AG-exon 2). Vì thế, vùng mã hóa protein MASP3 bắt đầu từ nt 152 và kết thúc tại nt 2338, mã hóa polypeptide 728 amino acid (GenBank ADF30843)

cDNA và *MASP3* ở lợn có mức độ tương đồng cao với *MASP3* ở các loài hữu nhũ như chó (90%), người (87%) và bò (87%), gần gấp đôi so với gà (76%) hoặc cá ngựa vằn (62%). Ngoài ra sự tương đồng cao

còn tìm thấy giữa người và chó (90%) hay giữa chó và bò (88%). Protein MASP3 lợn có sự tương đồng cao với bò (94%), chó (93%) và người (92%), thấp

hơn với gà (74%) và cá (54%) (Bảng 4). Các loài động vật hữu nhũ như người, bò và chó có mức độ tương đồng về protein MASP3 rất cao 91-93%

```

1  gcgtgacctgaatgga
17  gaatgccagccaattccagagacacacagggacctcagaacaaaag
62  ataaggcatcgctgacaccacacccgggacgagcttgccagacaag
107 tcaggccggggggaccaggcagccaggagcaccacaaggcaggaaa
152 ATGagggtggctccttctctctcatgctctgtgcttctccttgttg
    M R W L L L S H A L C F S L L
197 aaggcttcggcccatgtcgtggagcttaacāatatgtttggccag
    K A S A H V V E L N N M F G Q
242 atccaatctccttggttatccagactcctatccaagtgattcagag
    I Q S P G Y P D S Y P S D S E
287 gtaacgtggaacattactgtcccagagggatttcggatcaagctt
    V T W   T   P E G F R I K L
332 taacttcatgcacttcaacttggaatcctcctacctttgtgaatac
    Y F M H F N L E S S Y L C E Y
377 gactatgtgaaggtagaaactgaagaccagggtgctgcccacttc
    D Y V K V E T E D Q V L A T F
422 tgtggcaaggagaccacggacacagaacagacccctggccaggaa
    C G K E T T D T E Q T P G Q E
467 gtggctcttctctcgggctccttcattgtccatcactttccggtea
    V V F S P G S F M S I T F R S
512 gatttctccaatgaggaacgggtacacaggatttgatgccactac
    D F S N E E R Y T G F D A H Y
557 atggctgtggatgtggacgagtgcacagagcgggaagacgaggaa
    M A V D V D E C T E R E D E E
602 ctgtcctgtgaccactactgcāacaactacatcggcggctactac
    L S C D H Y C H N Y I G G Y Y
647 tgcctcctggcgtctcggcāacatcctccacacagacaacaggaca
    C S C R F G Y I L H T D
692 tggcgaagtggagtgcagtgacaacctcttcacccagaggactggg
    R V E C S D N L F T Q R T G
737 gtgatcaccagcāctgactaccccagcccttaccctaagagctcc
    V I T S P D Y P S P Y P K S S
782 gaatgctctctacaccattgagctagaagggtttctatggctcagc
    E C L Y T I E L E E G F M V S
827 ctgāagtttgaggatattttcgacatagaggaccatcctgaggta
    L Q F E D I F D I E D H P E V
872 tcttgcctctatgactacatcaagattaaagctgggtccaaaactt
    S C P Y D Y I K I K A G P K L
917 cttgggccccttctgtggagagaaagccccggaacccatcaatacc
    L G P F C G E K A P E P I N T
962 cagagccacagtātcagatccagttccgcagtgacaactcaggc
    Q S H S I Q I Q F R S D N S G
1007 gagaaccggggctggaggctgtttgtacagggcgacaggggaatgag
    E N R G W R L L Y R A T G N E
1052 tgccccaagctgcagccgcctgtccatgggaaaattgagcccttc

```

C P K L Q P P V H G K I E P L
 1097 caagccaagtactccttcaaagaccaggtgctcatcagctgtgac
 Q A K Y S F K D Q V L I S C D
 1142 actggctacaaagtgtggaaggataatgtggagatggatgtattc
 T G Y K V L K D N V E M D V F
 1187 cagatcgagtgtctgaaagatgggacgtggagtaacaagattccc
 Q I E C L K D G T W S N K I P
 1232 acctgcaaaatttgcggactgtggcactccggcaaaagctggaaaat
 T C K I A D C G T P A K L E N
 1277 gggctgggtcaccttttctaccaggaacaaccttaccacatacaaa
 G L V T F S T R N N L T T Y K
 1322 tctgaaatcatatactcctgccagcagccctactacaagatgctg
 S E I I Y S C Q Q P Y Y K M L
 1367 cacagtaccacaggtgtaataacgtgtttctgcccaaggagctctgg
 H S T T G V Y T C S A Q G V W
 1412 atgaatgaagtcctgggaaaagccagcccatgcctaccagtg
 M N E V L G K S Q P T C L P V
 1457 tgtgggtcagccctcccgtccctgccgaacctggtcaagcggatc
 C G Q P S R S L P N L V K R ↑ I
 1502 atcgaggcccggaacgctgagcccgccctctttccatggcaggcc
 I G G R N A E P G L F P W Q A
 1547 ctgatcgtggtggaggacacgtcgagggtgcccaacgacaagtgg
 L I V V E D T S R V P N D K W
 1592 tttgggagcggggccttgctctctgagtcctggatcctcacggca
 F G S G A L L S E S W I L T A
 1637 gccacgtgtgtgcgtccacagcgacagacaacaggtgacgcca
 A [H] V L R S Q R R D N T V T P
 1682 gtctccaaggagcatgtcacctgtacctgggcctgcatgacgtg
 V S K E H V T V Y L G L H D V
 1727 cgcgacaaatccggggctgtcaacagctcagctgcccgagtggtg
 R D K S G A V [M] [I] [S] A R V V
 1772 ctccacctgaactttgacatccagaactacaaccatgacattgt
 L H P D F D I Q N Y N H [D] I A
 1817 ctagtgcagctgcaggagccggtgcccctggggccgcacgtcatg
 L V Q L Q E P V P L G P H V M
 1862 cccgctgtgctgcccggcctgagccgaaggcccgccggccac
 P V C L P R P E P E G P A P H
 1907 atgctggggctggtagctggctggggcatctccaatcctaattgtg
 M L G L V A G W G I S N P [M] [I]
 1952 acagtggatgagatcatcagcagcgccagcgacacctgtcagat
 [S] D E I I S S G T R T L S D
 1997 gtcctgcagtatgtcaagttaccgctgggtgccgcacgcccagtg
 V L Q Y V K L P V V P H A E C
 2042 aagaccagctacagagtcgccggtcaggcaactacagcgtcacag
 K T S Y E S R S G [M] [I] [S] T E
 2087 aacatgtttctgtgccggctactatgaggcgccgaaggacacgtgc
 N M F C A G Y Y E G G K D T C
 2132 ctaggagacagcggcgggcctttgtcatccttgacgacctgagc
 L G D [S] G G A F V I L D D L S

```

2177 caacgctgggtggctcaaggcctgggtgtcctgggggggcccctgaa
    Q R W V A Q G L V S W G G P E
2222 gaatgtggcagtaagcaggtgtatgggtctacaccaaagtctcc
    E C G S K Q V Y G V Y T K V S
2267 aacTacgtggactgggtgtgggagcagatgggtccccctccaggc
    N Y V D W V W E Q M G S P P G
2312 ctgggggagctccaggtggagcggTGA 2338
    L G E L Q V E R *
2339 gccctgacttccttggcctccagcccccagccctgcaaggct
2384 gcatcatacactcctgaggatggcatactccatgttactcatgag
2429 accaaatgggatggaacacgctgcttccaccctcccaagacaacc
2474 cctgggaggcagcagagtggtatgggaaaaaggatctattcagga
2519 gacctgggtccctggccctgggctgggtcccttccccctcctggggc
2564 ctcactctcccaggagcacagcaactcttggccaactcccactcc
2609 tctcctggagtgcctaccccaggggcctccttcaactccttgcct
2654 ttgt 2657
    
```

Hình 1. Đặc điểm cDNA và protein MASP3. Start (ATG) và stop (TAG) codon được in hoa đậm. Chuỗi khởi đầu cho quá trình chuyển mã được in đậm. Các cysteine (C) được gạch dưới. Các vị trí dễ xúc tác được đóng khung. Vị trí glycosylation (*) được in viền (outline). Methionine loop được nhân đôi trong chuỗi được gạch dưới liên tục. Vị trí chia tách MASP3 thành hai chuỗi A và B được đánh dấu mũi tên (↑).

Bảng 2. Vị trí của các exon được xác định trên các gen GenBank

Exon (bp)	Vị trí nt trên GenBank GU810083	Vị trí nt trên GenBank NW_001885220	Exon (bp)	Vị trí nt trên GenBank GU810083	Vị trí nt trên GenBank NW_001885220
1 (156)	1-156	6106-6261	7 (119)	1044-1162	40356-40474
2 (232)	157-388	13508-13739	8 (79)	1163-1241	41458-41536
3 (178)	339-566	32888-33065	9 (138)	1242-1379	47782-47919
4 (132)	567-698	34689-34820	10 (75)	1380-1454	48724-48798
5 (197)	699-895	36977-37173	11 (1201)	1455-1655	52929-54171
6 (148)	896-1043	38896-39043			

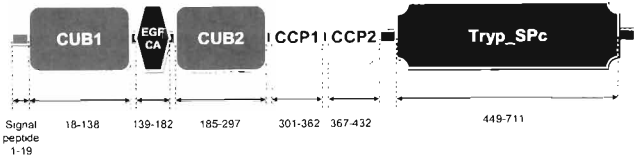
Dại phân tử MASP3 là một tập hợp của nhiều miền protein chức năng giàu cysteine như CUB1 (aa 18-138, 2 cysteine), EGF C (aa 139-182, 6 cysteine), CUB2 (aa 185-297, 3 cysteine), CCP1 (aa 301-362, 4 cysteine), CCP2 (aa 367-432, 4 cysteine) và Tryp_SpC (aa 449-711, 5 cysteine). Trong đó, miền Tryp_SpC lớn nhất (263 aa) và EGF C là miền nhỏ nhất (44 aa) nhưng giàu cysteine nhất (6 cysteine). Một số nghiên cứu chỉ ra rằng các serine protease của những phức hợp kích hoạt ở con đường cổ điển như C1r, C1s, và con đường lectin có đầu N với 5 miền cấu trúc (CUB1, EGF-like, CUB2, CCP1 và CCP2) và đầu C có một miền serine protease đặc trưng (Hình 2). Do sự chuyển đổi từ một zymogen sang thể kích hoạt, đầu N và C được chia tách thành

hai chuỗi khác nhau: chuỗi A có cầu nối disulfide và chuỗi B nhỏ hơn mang tính đặc trưng của miền serine protease (Stover *et al.*, 2003; Bork, Beckmann, 1993; Matsushita *et al.*, 1998).

CUB (ở các thành phần bộ thể C1r/C1s, Uegf, Bmp1) đại diện cho các protein liên quan đến sự điều hòa phát triển, được tìm thấy ở hầu hết các protein liên kết mạng của tế bào plasma (plasma membrane-associated proteins) và protein ngoại bào (extracellular proteins). Tuy nhiên, CUB không được tìm thấy ở prokaryote, thực vật và nấm men (Bork và Beckmann, 1993; Bork, 1991; Varela *et al.*, 1997). Các protein CUB có liên quan đến việc thay đổi chức năng, kể cả sự kích hoạt bộ thể, định hình sự phát triển, sửa chữa mô, hướng dẫn sợi trục thần kinh axon, sự hình thành

mạch (angiogenesis), tín hiệu dẫn tế bào (cell signaling), sự thụ tinh, sự cầm máu (haemostasis), sự

viêm nhiễm, xâm thực bào (endocytosis) (Perry *et al.*, 2007; Abdul Ajees *et al.*, 2006).



Hình 2. Đặc điểm cấu trúc bậc 1 protein MASP3 với các miền chức năng giàu cysteine

Bảng 3. Mức độ tương đồng của cDNA và protein giữa một số loài

Loài	bp ¹		Loài	bp ¹	%	Loài	aa ²		Loài	aa ²	%
Lợn	2657	vs	Người	4184	87	Lợn	728	vs	Người	728	92
Lợn	2657	vs	Chó	2286	90	Lợn	728	vs	Chó	728	93
Lợn	2657	vs	Bò	3915	87	Lợn	728	vs	Bò	728	94
Lợn	2657	vs	Gà	3636	76	Lợn	728	vs	Gà	730	74
Lợn	2657	vs	Cá	2307	62	Lợn	728	vs	Cá	740	54
Người	4184	vs	Chó	2286	90	Người	728	vs	Chó	728	91
Người	4184	vs	Bò	3915	79	Người	728	vs	Bò	728	91
Người	4184	vs	Gà	3636	68	Người	728	vs	Gà	730	74
Người	4184	vs	Cá	2307	64	Người	728	vs	Cá	740	54
Chó	2286	vs	Bò	3915	88	Chó	728	vs	Bò	728	93
Chó	2286	vs	Gà	3636	75	Chó	728	vs	Gà	730	73
Chó	2286	vs	Cá	2307	62	Chó	728	vs	Cá	740	55
Bò	3915	vs	Gà	3636	67	Bò	728	vs	Gà	730	73
Bò	3915	vs	Cá	2307	61	Bò	728	vs	Cá	740	54
Gà	3636	vs	Cá	2307	64	Gà	730	vs	Cá	740	55

¹ cDNA lợn (GenBank GU810083), người (GenBank NM_139125), chó (GenBank XM_845180), bò (GenBank NM_001076968), gà (GenBank NM_213586), cá Zebrafish (GenBank XM_001341900)

² Protein lợn (ADP30843), người (NP_624302), chó (XP_850273), bò (NP_001070436), gà (NP_998751), cá Zebrafish (XP_001341936)

Trong khi đó, EGF_CA (calcium-binding EGF-like domain) đại diện cho protein liên kết màng và protein ngoại bào ở động vật. Các protein này có vị trí kết nối với calcium tại đầu N của các vùng giống EGF giúp tăng cường chức năng sinh học của protein. Sự kết nối calcium cũng là cơ chế tương tác

giữa hầu hết các protein (Selander-Sunnerhagen *et al.*, 1992; Davis 1990; Blomquist *et al.*, 1984; Doolittle *et al.*, 1984; Appella *et al.*, 1988)

Riêng CCP (complement control protein module) còn được gọi là aka short consensus repeat (SCR) hoặc SUSHI repeat đã được nhận diện ở một vài protein của

phức hợp bó thể ở động vật hữu nhũ (Nonman *et al.*, 1991). Protein Tryp_SPC (Trypsin-like serine protease) được tổng hợp dưới dạng zymogen không kích hoạt, được tách rời suốt quá trình phân giải protein để sản sinh ra những thể kích hoạt. Các enzyme phân giải protein sử dụng serine trong quá trình xúc tác của chúng thường được tìm thấy trong cấu trúc virus, vi khuẩn và eukaryote (Rawlings và Barrett, 1994). Miền Tryp_SPC có 3 vị trí dễ xúc tác, đó là His⁴⁰⁷, Asp⁵⁵¹ và Ser⁶⁶⁴. Hiện nay đã có 7371 miền Tryp_SPC được nhận diện ở 7147 protein trong ngân hàng dữ liệu SMART's genomes database (<http://smart.embl-heidelberg.de/smart>). Theo Stover *et al.* (2003), methionine loop có cấu trúc CKTSYERSGNYSVTENMFC được nhận diện ở chuột, người, chuột cống. Vòng này cũng được tìm thấy

ở lợn aa 630-649.

Peptide tín hiệu được nhận diện từ aa 1-19. Với khối lượng phân tử 81,55 kDa, MASP3 chứa 27 cysteine. Giữa các cysteine hình thành 13 disulfide như Cys¹¹⁻⁴¹⁴, Cys²³⁻³²⁹, Cys⁹¹⁻³⁴⁷, Cys¹⁴¹⁻¹⁵⁷, Cys¹⁵³⁻⁴³⁶, Cys¹⁶⁶⁻²¹², Cys¹⁶⁸⁻²⁶⁰, Cys¹⁸¹⁻¹⁸⁵, Cys³⁰¹⁻⁴³², Cys³⁴⁹⁻³⁶⁷, Cys³⁶²⁻³⁷³, Cys⁴¹⁶⁻⁶⁴⁹ và Cys⁶⁶⁰⁻⁶⁹². Tỷ lệ amino acid được xác định cao nhất là Ser (8,4%), Val (8,1%), Leu (8,0%), kế đến là Gly (7,6%), Glu (7,1%), Pro (6,5%), Thr (5,9%), Asp (5,5%) và cuối cùng là các amino acid còn lại (mỗi loại chiếm <5%) (Bảng 5). Ngoài ra, trên polypeptide có 5 asparagine (N) có dấu hiệu N-glycosyl hóa tại các vị trí N²⁴¹ITV, N^{17x}RTC, N⁵³SSA, N⁵⁹⁹VTV và N⁶⁴¹YSV.

Bảng 4. Tỷ lệ amino acid trong chuỗi protein

Amino acid	Tỷ lệ (%)	Amino acid	Tỷ lệ, %	Amino acid	Tỷ lệ, %
A (Ala)	3,8	I (Ile)	4,4	S (Ser)	8,4
B (Asx)		K (Lys)	4,3	T (Thr)	5,9
C (Cys)	3,7	L (Leu)	8,0	V (Val)	8,1
D (Asp)	5,5	M (Met)	1,8	W (Trp)	1,8
E (Glu)	7,1	N (Asn)	4,1	Y (Tyr)	4,9
F (Phe)	3,6	P (Pro)	6,5	Z (Glx)	
G (Gly)	7,6	Q (Gln)	4,1		
H (His)	2,6	R (Arg)	3,8		

Vị trí của gen trên bản đồ và đa hình nucleotide

cDNA mã hóa MASP3 gồm 11 exon, được định vị từ 134.301.569-134.349.594 bp (Sscrofa10.2, Map viewer, NCBI) hoặc từ 94.375.011-94.422.595 bp (Sscrofa9, Ensembl Sus scrofa version 65.9, www.ensembl.org/Sus_scrofa/Info/Index) trên nhiễm sắc thể 13. Trong khi đó, gen *MASPI* ở người được định vị trên vùng nhiễm sắc thể 3q27-q28 và ở chuột trên nhiễm sắc thể 16 B2-B3 (Takada *et al.*, 1995). Mapping cũng đã chỉ ra sự hiện diện của một gen đơn *MASPI/3* ở chuột (Takada *et al.*, 1995).

Trình tự cDNA ở các giống lợn khác nhau Pietrain (n = 30) và Landrace (n = 30) không có thêm đa hình nucleotide trên gen *MASPI* tại vùng khác nhau giữa hai gen *MASPI* và *MASP3*.

Sự tương đồng giữa các serine protease của phức hợp bó thể

Giữa MASPI và MASP3 có sự tương đồng 81% ở mức độ cDNA (tương ứng GenBank GU810082 và GU810083) và 75% ở mức độ protein (tương ứng GenBank ADF30842 và ADF30843). Phân tích cũng

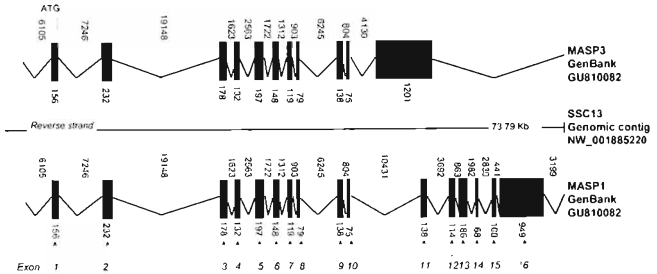
cho thấy, có sự giống nhau hoàn toàn 1461 nucleotide đầu tiên (10 exon) trong chuỗi cDNA (nt 1-1461) giữa hai gen *MASPI* và *MASP3* (Hình 3). Sự khác biệt giữa hai gen ở đoạn cuối (chỉ giống nhau khoảng 58%) bắt đầu từ nt 1642- 2657 (1196 bp) của gen *MASP3* và từ nt 1642-3009 (1548 bp) của gen *MASPI*. Mặc dù chuỗi nucleotide dài hơn, nhưng protein MASPI (699 aa) ngắn hơn protein MASP3 (728 aa).

Sự khác nhau về cấu trúc protein cũng tương tự như cDNA, nghĩa là cả hai protein có 439 aa ở đầu N (chứa hai miền CUB và EGF_CA) và đoạn trung tâm (chứa hai miền CCP) giống nhau hoàn toàn. Sự khác nhau được tìm thấy ở đầu C chứa miền Tryp_SPC. Đoạn khác nhau bắt đầu từ aa 440-728 (289 aa) ở MASP3 và từ aa 440-699 (260 aa) ở MASPI. Sự giống nhau ở đầu C là 39% giữa hai protein.

Mặc dù có sự khác nhau về chuỗi nucleotide và amino acid nhưng cấu trúc protein của hai gen đều chứa các miền chức năng giàu cysteine tương đồng nhau như CUB, EGF_CA, CCP và Tryp_SPC. Ngoài

trừ miền Tryp_SpC của MASP3 (263 aa) lớn hơn MASP1 (244 aa), thì các miền chức năng còn lại đều bằng nhau giữa hai protein. Tuy nhiên, dù nhỏ hơn nhưng miền Tryp_SpC của MASP1 (7 cysteine) lại là miền giàu cysteine hơn MASP3 (5 cysteine). Trong miền Tryp_SpC, MASP3 có methionine loop

(CKTSYSESRSGNYSVTENMFC), trong khi MASP1 có histidine loop (CGGSLLGSSWIVTAAHC). Vì hai protein khá tương đồng về cấu trúc nên vai trò và chức năng cũng có thể giống nhau trong cơ chế hỗ trợ quá trình kích hoạt miền dịch tự nhiên thông qua con đường lectin của phức hợp bổ thể.



Hình 3. Sự tương đồng chuỗi cDNA MASP3 và MASP1. Hàng số bên trên của mỗi gen là chiều dài (bp) của những đoạn intron và hàng số bên dưới là chiều dài (bp) của những đoạn exon

Trong khi đó, so với các serine protease khác trong phức hợp bổ thể, ở mức độ cDNA, MASP3 và MASP1 có độ tương đồng với thành phần bổ thể C1r (GenBank AY349421) lần lượt ở mức 68% và 71%. Mức độ tương đồng này thấp hơn được tìm thấy giữa MASP3 và MASP1 với C1s (GenBank AY349426) lần lượt là 62% và 64%. Trong khi đó, giữa C1r và C1 có mức độ tương đồng là 66%. Ở mức độ protein, sự tương đồng được tìm thấy giữa MASP3 (GenBank ADF30843) với C1r (GenBank AAR20889) là 68% hoặc với C1s (GenBank AAR20894) là 62%, trong khi giữa MASP1 (GenBank ADF30843) với C1r là 71% hoặc với C1s là 64%. Riêng C1r và C1s thì giống nhau 66%. Endo et al. (1998) và Schwaebler et al. (2002) ngụ ý rằng, MASP, C1r và C1s xuất thân từ một gen tổ tiên chung, mặc dù cấu trúc miền protease của chúng có đôi chút khác biệt.

KẾT LUẬN

Toàn bộ gen *MASP3* đã được giải mã trình tự, chuỗi cDNA và cấu trúc protein bậc một của nó cũng được nhận diện. Vì trí các exon, miền protein chức

năng giàu cysteine, N-glycosyl hóa đã được xác định. Tuy nhiên, không có đa hình gen trên vùng khác biệt giữa gen *MASP1* và *MASP3*. Bước đầu đã nhận diện được đặc điểm phân tử và phân tích một số vùng miền chức năng của gen, là tiền đề cơ bản để nghiên cứu tiếp về vai trò của gen trong cơ chế kích hoạt miền dịch theo con đường lectin của phức hợp bổ thể - đại diện cho cơ chế đáp trả miễn dịch ban đầu của miễn dịch dịch thể ở vật chủ chống lại mầm bệnh xâm nhiễm.

Lời cảm ơn : Nhóm tác giả chân thành cảm ơn sự hỗ trợ kinh phí của Bộ Nghiên cứu Giáo dục Đức (BMBF) trong khuôn khổ dự án VN03/ B01)

TÀI LIỆU THAM KHẢO

- Abdul Ajees A, Gunasekaran K, Volanakis JF, Narayana SV, Kotwal GJ, Murthy HM (2006) The structure of complement C3b provides insights into complement activation and regulation. *Nature* 444(7116): 221-225.
- Appella E, Weber IT, Blasi F (1988) Structure and function of epidermal growth factor-like regions in proteins. *FEBS Lett* 231(1): 1-4.

- Blomquist MC, Hunt LT, Barker WC (1984) Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. *Proc Natl Acad Sci USA* 81(23): 7363-7367
- Banda NK, Takahashi M, Takahashi K, Stahl GL, Hyatt S, Glogowska M, Wiles TA, Endo Y, Fujita T, Holers VM, Arend WP (2011) Mechanisms of mannose-binding lectin-associated serine proteases-1/3 activation of the alternative pathway of complement. *Mol Immunol* 49(1-2): 281-289
- Bork P (1991) Complement components C1r/C1s, bone morphogenic protein 1 and Xenopus laevis developmentally regulated protein UVS 2 share common repeats. *FEBS Lett* 282(1): 9-12.
- Bork P, Beckmann G (1993) The CUB domain, A widespread module in developmentally regulated proteins. *J Mol Biol* 231(2): 539-545
- Dahl MR, Thiel S, Matsushita M, Fujita T, Willis AC, Christensen T, Vorup-Jensen T, Jensenius JC (2001) MASP-3 and its association with distinct complexes of the mannose-binding lectin complement activation pathway. *Immunity* 15: 127-135.
- Davis CG (1990) The many faces of epidermal growth factor repeats. *New Biol* 2(5): 410-419
- Doolittle RF, Feng DF, Johnson MS (1984) Computer-based characterization of epidermal growth factor precursor. *Nature* 307(5951): 558-560
- Đỗ Võ Anh Khoa (2010) Phân tích đặc điểm phân tử và vai trò của gen C9 trong hệ miễn dịch ở lợn. *Tạp chí Khoa học và Phát triển* 8(3): 448-457
- Endo Y, Takahashi M, Nakao M, Saiga H, Sekine H, Matsushita M, Nonaka M, Fujita T (1998) Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates. *J Immunol* 161: 4924-4930
- Javahery R, Khachi A, Lu K, Kenzie-Gregory B, Smale ST (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* 14: 116-127.
- Lynch NJ, Khan SH, Stover CM, Sandrini SM, Marston D, Presanis JS, Schwaebler WJ (2005) Composition of the lectin pathway of complement in *Gallus gallus*. Absence of mannose-binding lectin-associated serine protease-1 in birds. *J Immunol* 174: 4998-5006
- Matsushita M, Endo Y, Fujita T (1998) MASP1 (MBL-associated serine protease 1). *Immunobiology* 199(2): 340-347.
- Norman DG, Barlow PN, Baron M, Day AJ, Sun RB, Campbell ID (1991) Three-dimensional structure of a complement control protein module in solution. *J Mol Biol* 219(4): 717-725
- Perry SE, Robinson P, Melcher A, Quirke P, Böhning III, Cook GP, Blair GE (2007) Expression of the CUB domain containing protein 1 (CDCP1) gene in colorectal tumour cells. *FEBS Lett* 581(6): 1137-1142
- Rawlings ND, Barrett AJ (1994) Families of serine peptidases. *Methods Enzymol* 244: 19-61.
- Schwaebler W, Dahl MR, Thiel S, Stover C, Jensenius JC (2002) The mannose-binding lectin-associated serine proteases (MASPs) and MASP-3: four components of the lectin pathway activation complex encoded by two genes. *Immunobiology* 205(4-5): 455-466
- Selander-Sunnerhagen M, Ullner M, Persson E, Teleman O, Stenflo J, Drakenberg T (1992) How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH₂-terminal EGF-like domain in coagulation factor X. *J Biol Chem* 267(27): 19642-19649
- Sørensen R, Thiel S, Jensenius JC (2005) Mannan-binding-lectin-associated serine proteases, characteristics and disease associations. *Springer Semin Immunopathol* 27(3): 299-319.
- Stover CM, Lynch NJ, Dahl MR, Hanson S, Takahashi M, Frankenberger M, Ziegler-Heitbrock L, Eperon I, Thiel S, Schwaebler WJ (2003) Murine serine proteases MASP-1 and MASP-3, components of the lectin pathway activation complex of complement, are encoded by a single structural gene. *Genes Immun* 4: 374-384
- Takada F, Seki N, Matsuda YI, Takayama Y, Kawakami M (1995) Localization of the genes for the 100-kDa complement activating components of R-reactive factor (CRARF and Crar) to human 3q27-q28 and mouse 16B2-B3. *Genomics* 25: 757-759
- Takahashi M, Ishida Y, Iwaki D, Kanno K, Suzuki T, Endo Y, Honma Y, Fujita T (2010) Essential role of mannose-binding lectin-associated serine protease-1 in activation of the complement factor D. *J Exp Med* 207(1): 29-37.
- Takahashi M, Iwaki D, Kanno K, Ishida Y, Xiong J, Matsushita M, Endo Y, Mura S, Ishi N, Sugamura K, Fujita T (2008) Mannose-binding lectin (MBL)-associated serine protease (MASP)-1 contributes to activation of the lectin complement pathway. *J Immunol* 180(9): 6132-6138
- Takahashi M, Mori S, Shigeta S, Fujita T (2007) Role of MBL-associated serine protease (MASP) on activation of the lectin complement pathway. *Adv Exp Med Biol* 598: 93-104
- Varela PF, Romero A, Sanz L, Romão MJ, Topleer-Petersen E, Calvete JJ (1997) The 2.4 Å resolution crystal structure of boar seminal plasma PSP-I/PSP-II, a zona pellucida-binding glycoprotein heterodimer of the spermadhesin family built by a CUB domain architecture. *J Mol Biol* 274(4): 635-649

MOLECULAR CHARACTERIZATION OF THE PORCINE MANNAN BINDING LECTIN SERINE PROTEASE 3

Do Vo Anh Khoa^{1,*}, Klaus Wimmers²

University of Cantho, Vietnam

²Leibniz Institute for Farm Animal Biology FBV, Dummerstorf, Germany

SUMMARY

Mannan binding lectin serine protease (MASP) plays an essential important role in innate immune response of host, especially in activation of complement system via the lectin and alternative pathways. MASPs family consisted of three distinct proteins (MASP1, MASP2 and MASP3). Therefore, the current study aimed at sequencing the whole gene and analyzing cDNA and molecular structure of MASP3 (MASP1 isoform 2). With 2657 bp and 11 exons, in which the coding region (nt 152-2388) encoded the protein MASP3 with 728 amino acids. cDNA and protein MASP3 among mammalian species (pig, human, cattle and dog) showed a high homology of 87-90% and of 92-94%, respectively. MASP3 had 27 cysteine residues. Amino acids showing a high percentage in the polypeptide included Ser (8.4%), Val (8.1%), Leu (8.0%), Gly (7.6%) and Glu (7.1%). With a molecular mass of approximately 81.55 kDa, MASP3 was an assembly of rich-cysteine protein domains such as CUB1 (aa 18-138, 2 cysteine), EGF_CA (aa 139-182, 6 cysteine), CUB2 (aa 185-297, 3 cysteine), CCP1 (aa 301-362, 4 cysteine), CCP2 (aa 367-432, 4 cysteine) in N-terminus and Tryp_SPc (aa 449-711, 5 cysteine) in C-terminus. Tryp_SPc was the largest domain whereas EGF_CA was the smallest. Besides, a signal peptide (aa 1-19), 13 disulfide bonds, 5 N-glycosylation sites, and methionine loop (aa 630-649) were also identified. All this molecular structure of the protein MASP3 was basic to study the role of MASP3 in immune response mechanism for disease resistance in pigs.

Keywords: *Gene MASP3, cDNA, protein MASP3, molecular characterization*