

ĐẶC ĐIỂM PHÂN TỬ MANNAN BINDING LECTIN SERINE PROTEASE 3 Ở LỢN

Đỗ Võ Anh Khoa¹, Klaus Wimmers²

TÓM TẮT

Gien mã hóa protein mannan-binding lectin serine protease (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. MASP3 gồm 3 gen phân biệt (MASP1, MASP2 và MASP3). Nghiên cứu được thực hiện nhằm giải mã trình tự chuỗi AcDN và phân tích đặc điểm phân tử MASP3 (MASP1 isoform 2) ở lợn. Chuỗi AcDN của MASP3 dài 2 657 bp chứa 11 exon đã mã hóa 728 amino axit (a.a). Sự tương đồng cao ở mức độ AcDN (87-90%) và protein (92-94%) được tìm thấy ở lợn, chó, người và bò. Đại phân tử MASP3 chứa 27 cysteine, tuy nhiên thành phần amino axit 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%), Glu (7,1%). Với phân tử khối khoảng 81,55 kDa, MASP3 là một tập hợp của nhiều miền protein chức năng giàu cysteine như CUB1 (a.a 18-138, 2 cysteine), EGF_CA (a.a 139-182, 6 cysteine), CUB2 (a.a 185-297, 3 cysteine), CCP1 (a.a 301-362, 4 cysteine), CCP2 (a.a 367-432, 4 cysteine) và Tryp_SPc (a.a 449-711, 5 cysteine). Trong đó, miền Tryp_SPc lớn nhất (263 a.a) và EGF_CA là miền nhỏ nhất (44 a.a) nhưng là miền giàu cysteine nhất (6 cysteine). Ngoài ra, signal peptide (a.a 1-19), 13 cấu nối disulfua, 5 vị trí N-glycosylation và vòng methionin loop (a.a. 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: *MASP3, AcDN, protein, đặc điểm phân tử, lợn.*

L BÁT VĂN ĐÉ

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) and 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 proteases (MASPs) là enzym xúc tác chính của con đường này nhờ vào sự kết dính của MASPs với thành phần hydratcacbon vi khuẩn (Takahashi et al., 2008; Schwaebler et al., 2002). MASPs hiện diện trong huyết tương như là những zymogen (Sørensen et al., 2005). Gia đình MASP gồm 3 chuỗi proteaza (MASP1, MASP2 và MASP3) và một protein không có chức năng enzym (non-enzymatic protein, Map19) (Schwaebler et al., 2002). MASP1 và MASP3 là sản phẩm cắt nối thay thế (alternative splice products) và là những gen được mã hóa từ một gen đơn MASP1/3 (Dahl et al., 2001). Mỗi gen MASP1 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. Chuỗi B

chứa miền chuỗi proteaza (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 axit amin còn lại thêm vào (Schwaebler et al., 2002).

MASP2 là proteaza đóng vai trò quan trọng trong sự chia tách C4 và C2 để sản sinh ra enzym chính của phức hợp bổ thể C3-convertaza, một sản phẩm kết hợp của C4bC2b (Schwaebler et al., 2002). Trong khi đó, MASP1 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 MASP1/3 sẽ 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 MASP1/3 có liên quan đến sự nhạy cảm với vi rút gây bệnh cúm ở chuột (Takahashi et al., 2007). Khi MBL hoặc ficolin kết dính với hydratcacbon trên bề mặt vi khuẩn sẽ gây kích hoạt vùng zymogen của MASPs (Takahashi et al., 2007) và vì thế phức hợp bổ thể được kích hoạt theo con đường lectin (lectin pathway). Hai chuỗi proteaza là MASP1 và MASP3 liên kết với MBL hoặc ficolin để kích hoạt phức hợp bổ thể theo con đường lectin. MASP1 đóng góp trực tiếp đến sự hoạt hóa con đường lectin thông qua sự hoạt hóa MASP2 (Takahashi et al., 2008) và MASP1 cũng đóng góp cho sự hoạt hóa con đường nhánh (alternative complement pathway) (Takahashi et al.,

¹ Bộ môn Chăn nuôi, Khoa Nông nghiệp và Sinh học Ứng dụng, Trường Đại học Cần Thơ

² Viện Nghiên cứu Sinh học Gia súc FBN, CHLB Đức

2010). Trong khi đó, sự kích hoạt phức hợp bổ thể theo con đường nhánh có sự góp mặt của Mannose-binding lectin-associated chuỗi proteaza 1/3 (MASP1/3) nhờ vào sự tách rời pro-factor 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 MASP1 có thể có chức năng như là một protein hòa tan không kết nối (Banda et al., 2011).

Do bài vai trò quan trọng của MASP trong cơ chế miễn dịch ban đầu theo con đường lectin của phức hợp bổ thể, nhiều nghiên cứu cũng đã được 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.

II. VẬT LIỆU VÀ PHƯƠNG PHÁP

Thí nghiệm sử dụng nguồn vật liệu (động vật, hóa chất, thiết bị và dụng cụ) và phương pháp thí nghiệm (phân lập AmRN, tổng hợp AcDN, giải mã trình tự chuỗi, thiết kế mồi, thành phần phản ứng PCR) như được mô tả bởi Đỗ Võ Anh Khoa (2010). Trong đó:

(i) Động vật thí nghiệm là các giống lợn Hampshire (n=1), Landrace Đức (n=34), Duroc (n=1), Pietrain (n=1), Berlin Miniature Pig (n=1) và Mường Khương (n=24).

(ii) Tổng số AmRN được phân lập từ mẫu gan của động vật thí nghiệm sử dụng phương pháp TRI-Reagent (Sigma-Aldrich, Taufkirchen, Đức) và sau đó được tổng hợp chuỗi AcDN nhờ vào oligo (dT) 13 VN primer (Promega, Mannheim, Đức), random hexamer primers (Promega, Mannheim, Đức) và Super ScriptTMIII Rnase H (Invitrogen, Karlsruhe, Đức).

Các cặp mồi được thiết kế sử dụng trong thí nghiệm và chu trình nhiệt thực hiện phản ứng PCR cho từng cặp mồi riêng biệt được trình bày trong bảng 1 và 2. Các cặp mồi đặc hiệu được thiết kế dựa trên ngân hàng dữ liệu GenBank BW979216, CJ004744, XM_001927110, XM_001927155 và dựa trên kết quả sàng lọc các chuỗi nucleotit tương đồng giữa các loài động vật (người, chuột,...).

Nghiên cứu đã sử dụng các phần mềm online để phân tích kết quả gồm:

Primer3 (<http://frodo.wi.mit.edu/>): thiết kế mồi.

ORF Finder (Open Reading Frame Finder) (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>): nhận diện cấu trúc bậc 1 protein.

ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>): so sánh mức độ tương đồng của chuỗi AcDN và protein.

SMART (<http://smart.embl-heidelberg.de/>): phân tích các miền protein chức năng.

SignalP 4.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>): nhận diện tín hiệu pepit.

Composition/Molecular Weight Calculation Form (http://pir.georgetown.edu/pirwww/search/comp_mw.shtml): tính toán phân tử khối.

DiANNA 1.1 web server (<http://clavius.bc.edu/~clotelab/DiANNA/>): phân tích cấu nối disulfua.

NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>): nhận diện vị trí N-glycosylation.

Một phản ứng PCR được thực hiện trong tổng số 20 μ l dung dịch, trong đó gồm có 50 ng AcDN, 0,2 mM mỗi muối hoặc mỗi ngược, 50 μ M dNTP, 0,5 U Taq polymeraza, 1xTaq buffer và 1,5 mM MgCl₂.

Bảng 1: Các cặp mồi sử dụng để khuếch đại các đoạn gen dọc theo chuỗi AcDN

Tên mồi	Mồi xuôi fw 5'3' (vị trí trên GenBank GU810083)	Mồi ngược rev 3'5' (vị trí trên GenBank GU810083)	Chiều dài, bp
MASP3_1	aatgaggtggctctctctc (nt. 151-171)	tcacaggacagttctctgct (nt. 592-612)	462
MASP3_2	gagaccacggacacagaaca (nt. 431-450)	caagaagttttggaccagctt (nt. 900-920)	490
MASP3_3	gagctccgaatgtctctacacc (nt. 775-796)	tccccaggattcattcacc (nt. 1.411-1.430)	655
MASP3_4	gccagcagccctactacaag (nt. 1.341-1.360)	actgcaggacatctgacaagg (nt. 1.986-2.006)	666
MASP3_5	ccacctgactttgacatcc (nt. 1.774-1.793)	acaacggcaaggagtgaag (nt. 2.638-2.657)	884

Bảng 2: Chu trình nhiệt PCR

Tên mẫu	Biến tính ban đầu (thời gian)	Biến tính (thời gian)	Gắn mồi (thời gian)	Tổng hợp (thời gian)	Kết thúc (thời gian)	Giữ mẫu (thời gian)	Tổng số chu kỳ
MASP3_1	94°C (5')	94°C (30'')	56°C (30'')	72°C (1')	72°C (5')	15°C (∞)	35
MASP3_2	94°C (5')	94°C (30'')	60°C (30'')	72°C (1')	72°C (5')	15°C (∞)	35
MASP3_3	94°C (5')	94°C (30'')	60°C (30'')	72°C (1')	72°C (5')	15°C (∞)	35
MASP3_4	94°C (5')	94°C (30'')	58°C (30'')	72°C (1')	72°C (5')	15°C (∞)	35
MASP3_5	94°C (5')	94°C (50'')	66°C (50'')	72°C (1'20'')	72°C (5')	15°C (∞)	50

Để giải mã trình tự gen, sản phẩm từ phản ứng PCR sau khi được tinh sạch sẽ được gửi đến Viện Công nghệ Sinh học (Viện Khoa học Công nghệ Việt Nam) để giải mã trình tự chuỗi AcDN.

KẾT QUẢ THẢO LUẬN

1. Đặc điểm phân tử AcDN và protein

Kết quả giải mã trình tự nucleotit đã nhận diện được chiều dài chuỗi AcDN của gen MASP3 là 2.657 bp (GenBank GU810083). Vùng 5'-UTR (nt. 1-151) và 3'-UTR (nt. 2.339-2.657) có chiều dài lần lượt là 151 bp và 319 bp. Codon bắt đầu (start codon) là một methionin ATG được xác định từ nt. 152-154 và codon kết thúc (stop codon) TGA từ nt. 2.336-2.338. Giống như ở chuột (Stover et al., 2003), vùng 5'-UTR của MASP3 ở lợn có chứa chuỗi khởi đầu cho quá trình chuyển mã (transcription initiation site) CCACACC (nt. 79-85), giống như kiểu C/T C/T A.1 N T/A C/T C/T (trong đó N là codon bắt đầu) (Javahery et al., 1994). Kết quả so sánh chuỗi AcDN

với GenBank NW_001885220 (Ssc13_WGA1011_1: 115459-225459 *Sus scrofa* chromosome 13 genomic contig) đã nhận diện được 11 exon của gen MASP3 dựa theo nguyên tắc cấu trúc ADN (exon 1-GT.intron.AG-exon 2) (bảng 3). Vì thế, vùng mã hóa protein MASP3 bắt đầu từ nt. 152 và kết thúc tại nt. 2.338. Vùng này đã mã hóa chuỗi polypeptit dài 728 amino axit (GenBank ADF30843). Ở cấp độ AcDN, MASP3 của lợn có mức độ tương đồng cao với MASP3 của các loài hữu nhũ như chó (90%), người (87%), bò (87%). Mức độ tương đồng giảm dần khi so với gà (76%) hoặc cá zebrafish (62%). Ngoài ra sự tương đồng cao còn tìm thấy ở người và chó (90%) hay giữa chó và bò (88%). Ở cấp độ protein MASP3, có sự tương đồng cao được nhận diện giữa lợn với bò (94%), chó (93%) hoặc người (92%). Mức độ tương đồng thấp hơn được tìm thấy giữa lợn với gà (74%) hoặc cá (54%). Giữa các loài động vật hữu nhũ như người, bò và chó có mức độ tương đồng về protein MASP3 rất cao 91-93% (bảng 4).

Bảng 3: Vị trí của các exon được xác định trên các ngân hàng gen

Exon (chiều dài, bp)	Vị trí trên ngân hàng gen GU810083	Vị trí trên ngân hàng gen NW_001885220	Exon (chiều dài, bp)	Vị trí trên ngân hàng gen GU810083	Vị trí trên ngân hàng gen NW_001885220
1 (156)	nt. 1-156	nt. 6.106-6.261	7 (119)	nt. 1.044-1.162	nt. 40.356-40.474
2 (232)	nt. 157-388	nt. 13.508-13.739	8 (79)	nt. 1.163-1.241	nt. 41.458-41.536
3 (178)	nt. 339-566	nt. 32.888-33.065	9 (138)	nt. 1.242-1.379	nt. 47.782-47.919
4 (132)	nt. 567-698	nt. 34.689-34.820	10 (75)	nt. 1.380-1.454	nt. 48.724-48.798
5 (197)	nt. 699-895	nt. 36.977-37.173	11 (1.201)	nt. 1.455-1.655	nt. 52.929-54.171
6 (148)	nt. 896-1.043	nt. 38.896-39.043			

Đại phân tử MASP3 là một tập hợp của nhiều miền protein chức năng giàu cysteine như CUB1 (a.a 18-138, 2 cysteine), EGF_CA (a.a 139-182, 6 cysteine), CUB2 (a.a 185-297, 3 cysteine), CCP1 (a.a 301-362, 4 cysteine), CCP2 (a.a 367-432, 4 cysteine) và Tryp_SPc (a.a 449-711, 5 cysteine). Trong đó, miền Tryp_SPc lớn nhất (263 a.a) và EGF_CA là miền nhỏ nhất (44 a.a) nhưng giàu cysteine nhất (6 cysteine). Một số

ngiên cứu chỉ ra rằng các chuỗi proteaza của những phức hợp kích hoạt ở con đường cổ điển, i.e. C1r, C1s, và con đường lectin có đầu N-terminus với 5 miền cấu trúc (CUB1, EGF-like, CUB2, CCP1 và CCP2) và đầu C-terminus có một miền chuỗi proteaza đặc trưng. Do bởi sự chuyển đổi từ một zymogen sang thể kích hoạt, đầu N- và C-terminus được chia tách thành hai chuỗi khác nhau: chuỗi A có cấu nối disulfua và chuỗi B

nhỏ hơn mang tính đặc trưng của miền chuỗi proteaza (Stover et al., 2003; Bork và Beckmann, 1993; Matsushita et al., 1998). CUB (ở các thành phần bổ 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 màng của tế bào huyết tương (plasma membrane-associated proteins) và protein ngoại bào (extracellular proteins). Tuy nhiên, CUB không được tìm thấy ở prokaryotes, thực vật và nấm men (yeast) (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 bổ thể, định hình sự phát triển, sửa chữa mô, hướng dẫn sợi trục thần kinh axon, sự hình thành mạch (angiogenesis), đánh dấu tế bào (cell signaling), sự thụ tinh, sự cầm máu (haemostasis), sự viêm nhiễm, endocytosis,... (Perry et al., 2007; Abdul Ajees et al., 2006). Trong khi đó, EGF_CA (calcium-binding EGF-like domain) là đại diện cho các protein kết nối màng (membrane-bound protein) và các protein ngoại bào (extracellular protein) ở động vật. Các protein này có vị trí kết nối với canxi tại đầu N-terminus của EGF-like domains giúp tăng cường chức năng sinh học của

protein. Sự kết nối canxi 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 (suất protein kiểm soát bổ thể) còn được gọi là lặp lại sự liên ứng (SCRs) hoặc SUSHI repeats đã được nhận diện ở một vài protein của phức hợp bổ thể ở động vật hữu nhũ (Norman et al., 1991). Protein Tryp_SPC (Trypsin-like serine protease) được tổng hợp như là zymogens không kích hoạt - được tách rời suốt quá trình proteolysis để sản sinh ra những thể kích hoạt. Các proteolytic enzyme sử dụng chuỗi trong quá trình xúc tác của chúng thường được tìm thấy trong cấu trúc vi rút, vi khuẩn và eukaryotes (Rawlings và Barrett, 1994). Miền Tryp_SPC có 3 vị trí dễ xúc tác, đó là His⁴⁹⁷, Asp⁵⁵³ và Ser⁶⁶⁴. Hiện nay đã có 7.371 miền Tryp_SPC được nhận diện ở 1.147 proteins trong ngân hàng dữ liệu SMART's genomes database (<http://smart.embl-heidelberg.de/smart>). Theo Stover et al. (2003), vòng methionin có cấu trúc CKTSYESRSGNYSVTENMFPC được nhận diện ở chuột, người, loài gặm nhấm (rat). Vòng này cũng được tìm thấy ở lợn a.a. 630-649.

Bảng 4: Mức độ tương đồng chuỗi AcDN và protein giữa một số loài

Loài	bp ¹		Loài	bp ¹	%	Loài	a.a ²		Loài	a.a 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_AcDN: lợn (ngân hàng gen GU810083), người (ngân hàng gen NM_139125), chó (ngân hàng gen XM_845180), bò (ngân hàng gen NM_001076968), gà (ngân hàng gen NM_213586), cá Zebrafish (ngân hàng gen XM_001341900)

²_Protein: lợn (ADF30843), người (NP_624302), chó (XP_850273), bò (NP_001070436), gà (NP_998751), cá Zebrafish (XP_001341936)

Tin hiệu pepit được nhận diện từ a.a 1-19. Với phân tử khối khoảng 81,55 kDa Daltons, protein MASP3 chứa 27 cysteine. Giữa các cysteine có thể

hình 13 cấu nối disulfua như Cys¹¹⁻¹¹⁴, Cys⁷³⁻³²⁹, Cys³⁹⁷, Cys¹⁴³⁻¹⁵⁷, Cys¹⁵³⁻⁴³⁶, Cys¹⁶⁶⁻²¹², Cys¹⁶⁸⁻²⁶⁰, Cys¹⁸¹⁻¹⁸⁵, Cys³⁰¹⁻⁴³², Cys³⁴⁹⁻³⁶⁷, Cys³⁶²⁻⁵⁷³, Cys⁶³⁰⁻⁶⁴⁹ và Cys⁶⁶⁰⁻⁶⁸². Ty

le các thành phần axit amin được xác định cao nhất là Ser (8,4%), Val (8,1%), Leu (8,0%), kế đến là Gly (7,6%), Glu (7,1%), tiếp theo là Pro (6,5%), Thr (5,9%), Asp (5,5%) và cuối cùng là các axit amin còn lại (mỗi

loại chiếm <5%) (Bảng 5). Ngoài ra, trên chuỗi polypeptit có 5 atparagin (N) có dấu hiệu N-glycosylation tại các vị trí N⁴⁹ITV, N¹⁷⁸RTC, N⁵²³SSA, N⁵⁹⁹VTV và N⁶⁴⁰YSV.

Bảng 5: Tỷ lệ các thành phần axit amino trong chuỗi protein

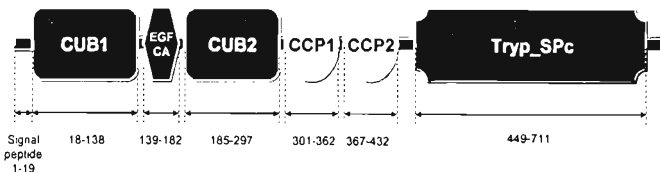
Kí hiệu a.a	Tên a.a	Tỉ lệ, %	Kí hiệu a.a	Tên a.a	Tỉ lệ, %	Kí hiệu a.a	Tên a.a	Tỉ lệ, %
A	Ala	3,8	I	Ile	4,4	S	Ser	8,4
B	Asx	-	K	Lys	4,3	T	Thr	5,9
C	Cys	3,7	L	Leu	8,0	V	Val	8,1
D	Asp	5,5	M	Met	1,8	W	Trp	1,8
E	Glu	7,1	N	Asn	4,1	Y	Tyr	4,9
F	Phe	3,6	P	Pro	6,5	Z	Glx	
G	Gly	7,6	Q	Gln	4,1			
H	His	2,6	R	Arg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

1 gcgtgacctgaatgga
17 gaatgccagccaalttcagagacacacagggacctcagaacaag
62 ataaaggcatcgtcgtacacacaccgggagcagctcgtcagacaag
107 tcaggcgggggaccaggcagccaggagcaccacaggraggaaa
152 ATGAGgttgctctctctctctcgtctcgtctctctctgttg
MRWL LSHALCSLL
197 aaggcttcggccatgctgtggagcttaacaatatgtttggccag
KASAHVVVELNNMFGQ
242 atccaatctcttggttatccagactctctatccaagtgttcagag
IQSPGYPDSPSDSE
287 gtaactgtgaacattactgtcccaggagggttcgtcatcaagctt
VTWHITVPEGRFKL
332 tacttcagacttcaactgtgaatctctcactcttgtgaatac
YFMHFNLESSYLQ EY
377 gactatgtgaaggtagaaaactgaagaccaggtgctggccaccttc
DYVKVETEDQVLATF
422 tgtggcaaggagaccacggacacagaaacagccctgcaggaa
CGKETTDTEQTPEG E
467 gtggtctctctcgggctcttcctatgtccatcatttcgggtca
VVFSPGFSFMSITFRS
512 gattttccaatgaggaaacgggtacacaggatttgatcccactac
DFSNEERYTGFD AHY
557 atggctgtggatgtggacaggtgcacagagcgggaagacaggaa
MAVDVDECTEREDE E
602 ctgtctgtgacctactgccacaactacatcggcgactactac
LSCDHYCHNYIG GYY
647 tgcctctgccgtctcggctacatctccacacagacaacaggaca
CSQRFGYILHTDNRT
692 tggcaggtgagggtgcagtgacaacctcttaccagaggactggt
Q RVEQSDNLFTRTG
737 gtgacaccagccctgactacccccagcccttaccctaaagactcc
VITSPDYSPYPKSS
782 gaatgctctctaccattgagctagaagagggttcatggtcagc

E C L Y T I E L E E G F M V S
827 ctgcagtttgaggatattttcgcacatagaggaccalcctgaggta
L Q F E D I F D I E D H P E V
872 tectgccctatgactacatcaagattaaagctgttcccaaaactt
S C P Y D Y I K I K A G P K L
917 ctggggccctctgtggagagaaagccccggaaacctcaatacc
L G P F C G E K A P E P I N T
962 cagagccacagatalcagatccagcttcgcagtgacaactcaggc
Q S H S I Q I Q F R S D N S G
1007 gaaagccggggctggaggtgtgtacaggccgacaggggaatgag
E N R G W R L L Y R A T G N E
1052 tgcaccaagctgcagccgctgtccatgggaaaatgagcccttg
C P K L Q P P V H G K I E P L
1097 caagccaagtaactcttcaaaagaccaggtgtcatcagctgtgac
Q A K Y S F K D Q V L I S C D
1142 actcgctacaaaigtctgaagcataatgtggagatggatgtattc
T G Y K V L K D N V E M D V F
1187 cagatcgagtgctgaaagatgggactgtggagttaacaagattccc
Q I E C L K D G T W S N K I P
1232 acctgcaaaattgcggactgtggcacttcggcaagctggaaaat
T C K I A D C G T P A K L E N
1277 ggctctggctaccttttctaccaggacaaccttaccacatacaa
G L V T F S T R N N L T T Y K
1322 tctgaatatatactcttgcacagcagcctactacaagatgttg
S E I I Y S C Q P P Y Y K M L
1367 cacagttaccaggtgtatatactgttctcccaaggagtgctgg
H S T T G V Y T C S A Q G W W
1412 atgaatgaagtcctggggaaaagccagcccacctgctaccagtg
M N E V L G K S Q P T C L P V
1457 tgtgtgcagccctcccgtccctgcgaacctgttcaagcggatc
C G Q P S R S L P N L V K R T I
1502 atcggagccgggaacgtgagcccgccctttccatgctggcaggcc
I G G R N A E P G L F P W Q A

1547 clgacgtgtggtgaggacacgtcagggtgcccacgacaagtgg
L I V V E D T S R V P N D K W
1592 ttggggagcggggccttgctcctgagtcctggatcctcacggca
F G S G A L L S E S W I L T A
1637 gccacgtgctgctgctcacagcgcagagacacacgggtgagccca
A H V L R S Q R R D N T V T P
1682 gtccacaggtgacatgctaccgtctacgtcgggcctgcatgacgtg
V S K E H V T V Y L G L H D V
1727 cgcgacaaatccgggctgcaacagctcagctcccagtggtg
R D K S G A V A C C I A R V V
1772 ctccaccctgactttgacatccagaactacaacatgacatgtt
L H P D F D I Q N Y N H D I A
1817 ctatgctgagctgcaggagccggtgccccggggcgccagctatg
L V Q L Q E P V P L G P H V M
1862
cccgtctgctgctgagggcctgagccgcaagggcggcgccccac
P V C L P R P E P E G P A P H
1907 atgctggggctgtagctggctggggcactccaalctaatgtg
M L G L V A G W G I S N P I T T
1952 acagtggatgagatcagcagcgcggcagcggcagcttgtagat
C T D E I I S S G T R T L S D
1997 gtctgcagtagtcaagttaccgtgggtgcgcacgcagtgctg
V L Q Y V K L P V V P H A E C
2042 aagaccagctacaggtcccggcaggcaactacagcgtcacagag
K T S Y E S R S G I T T T E
2087 aacalgtctgtgcccgtactatgaggcgccgaaggacacgtgc
N M E C A G Y Y E G G K D T C
2132 ctaggagacagcggcggggccttgctccttgtagcagctgagc
L G D S G G A F V I L D D L S
2177 caacgtggtgctgacaggtgaggtggtgctcctggggggccctgaa
Q R W V A Q G L V S W G G P E
2222 gaatgtggcagtaagcaggtgtatgggggtctacacaaagtccc
E C G S K Q V Y G V Y T K V S
2267 aactacgtggactgggtgtggggagcagatgggctcccctccagc
N Y V D W V W E Q M G S P P G
2312 ctggggggagctccaggtggagcggTGA 2338
L G E L Q V E R *
2339 gccctgacttcttggtctccagccaccagccctgcaagct
2384 gcatcacactcctgaggatggcactacatgttactatgag
2429 accaaatgggaggaacacgtctccaccctcccaagacaacc
2474 ctggggaggcagcagatggtatgggaanaaggatattcagga
2519 gactcgggtccctggccctgggtgggtcctccctcctggggc
2564 ctacactccagagacacgaaccttggccaactcccaacc
2609 tctcctggagtccacccaggggctccttactccttgcct
2654 ttgt 2657

Hình 2: Đặc điểm chuỗi AcDN và protein của 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 nhận diện trong chuỗi được gạch dưới liên tục. Vị trí chia tách MASP3 thành hai chuỗi A và B được đánh dấu mũi tên (↑).



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

2. Vị trí của gen trên bản đồ và đa hình nucleotit
Sự hệ trình kết quả AcDN trên ngân hàng dữ liệu cho thấy 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, http://www.ensembl.org/Sus_scrofa/Info/Index) trên nhiễm sắc thể 13. Trong khi đó, gen MASP1/3 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). Kết quả lập bản đồ gen cũng đã chỉ ra sự hiện diện

của một gen đơn MASP1/3 ở chuột (Takada et al. 1995).

Kết quả đọc trình tự chuỗi AcDN trên các giống heo khác nhau Pietrain (n=30) và Landrace (n=30) không tìm thấy được điểm đột biến trên gen MASP3 tại vùng khác nhau giữa hai gen MASP1 và MASP3.

3. Sự tương đồng giữa các chuỗi proteaza của phức hợp bổ thể

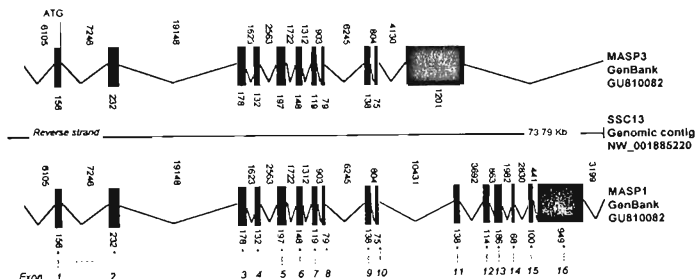
Giữa MASP1 và MASP3 có sự tương đồng 81% ở mức độ AcDN (tương ứng GenBank GU810082 và GU810083) và 75% ở mức độ protein (tương ứng ngân

hàng gen ADF30842 và ngân hàng gen ADF30843). Kết quả phân tích cũng cho thấy, có sự giống nhau hoàn toàn 1.461 nucleotit đầu tiên (10 exon) trong chuỗi AcDN (nt. 1-1.461) giữa hai gen MASP1 và MASP3. Sự khác biệt giữa hai gen ở đoạn cuối (chỉ giống nhau khoảng 58%) bắt đầu từ nt. 1.642- 2.657 (1.196 bp) của gen MASP3 và từ nt. 1.642-3.009 (1.548 bp) của gen MASP1. Mặc dầu chuỗi nucleotit được giải trình nhiều trình tự hơn, nhưng gen MASP1 mã hóa protein ít hơn gen MASP3 (699 a.a vs 728 a.a).

Sự khác nhau về cấu trúc protein cũng tương tự như AcDN, nghĩa là cả hai protein có 439 a.a ở đầu N-terminus (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 đoạn C-terminus chứa miền Tryp_SPc. Đoạn khác nhau bắt đầu từ a.a 440-728 (289 a.a) ở MASP3 và từ a.a 440-699 (260 a.a)

ở MASP1. Sự giống nhau trong đoạn C-terminus là 39% giữa hai protein.

Mặc dù có sự khác nhau về chuỗi nucleotit và axit amin 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. Ngoại trừ miền Tryp_SPc của MASP3 (263 a.a) lớn hơn MASP1 (244 a.a), 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 vòng methionin (CKTSYESRSGNYSVTENMF), trong khi MASP1 có vòng histidinin (CGGSLLGSSWVTAHC). Vì hai protein khá tương đồng về cấu trúc nên vai trò và chức năng cũng có thể sẽ 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 AcDN 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 chuỗi proteaza khác trong phức hợp bổ thể, ở mức độ AcDN, MASP3 và MASP1 có độ tương đồng với thành phần bổ thể C1r (ngân hàng gen AY349421) lần lượt ở mức 68% và 71%. Mức độ tương đồng này thấp hơn được tìm thấy giữa MASP3 và MASP1 với C1s (ngân hàng gen 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 (ngân hàng gen ADF30843) với C1r (ngân hàng gen AAR20889) là 68% hoặc với C1s (ngân hàng gen AAR20894) là 62%, trong khi giữa MASP1 (ngân hàng gen ADF30842) 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 proteaza của chúng có đôi chút khác biệt.

IV. KẾT LUẬN

Toàn bộ chiều dài gen MASP3 đã được giải mã trình tự chuỗi AcDN và cấu trúc protein bậc một của nó cũng đã được nhận diện. Vị trí các exon, miền protein chức năng giàu cysteine, N-glycosylation,... đã được xác định. Tuy nhiên, nghiên cứu không tìm thấy được đa hình gen trên vùng khác biệt giữa gen MASP1 và MASP3. Kết quả nghiên cứu đạt được

những kết quả bước đầu về đặc điểm phân tử và phân tích một số vùng miền chức năng của gen. Đây là tiền đề cơ bản cho những nghiên cứu tiếp theo về vai trò của gen trong cơ chế kích hoạt miễn 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 tạ

Công trình được hoàn thành dựa sự hỗ trợ kinh phí của Bộ Nghiên cứu Giáo dục CHLB Đức (BMBF) trong khuôn khổ dự án VNB 03/B01

TÀI LIỆU THAM KHẢO

1. Abdul Ajces A., Gunasekaran K., Volanakis J. E., Narayana S. V., Kotwal G. J., Murthy H. M., 2006. The structure of complement C3b provides insights into complement activation and regulation. *Nature* 444(7116): 221-225.
2. Appella E., Weber I. T., Blas F., 1988. Structure and function of epidermal growth factor-like regions in proteins. *FEBS Lett* 231(1): 1-4.
3. Blomquist M. C., Hunt L. T., Barker W. C., 1984. Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. *Proc. Natl. Acad. Sci. USA* 81(23): 7363-7367.
4. Banda N. K., Takahashi M., Takahashi K., Stahl G. L., Hyatt S., Glogowska M., Wiles T. A., Endo Y., Fujita T., Holers V. M., Arend W. P., 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.
5. 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.
6. Bork P. and Beckmann G., 1993. The CUB domain. A widespread module in developmentally regulated proteins. *J. Mol. Biol.* 231(2): 539-545.
7. Dahl M. R., Thiel S., Matsushita M., Fujita T., Willis A. C., Christensen T., Vorup-Jensen T., Jensenius J. C., 2001. MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway. *Immunity* 15: 127-135.
8. Davis C. G., 1990. The many faces of epidermal growth factor repeats. *New Biol.* 2(5): 410-419.
9. Doolittle R. F., Feng D. F., Johnson M. S., 1984. Computer-based characterization of epidermal growth factor precursor. *Nature* 307(5951): 558-560.
10. 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.
11. 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.
12. Javahery R., Khachi A., Lo K., Zenzie-Gregory B., Smale S. T., 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* 14: 116-127.
13. Lynch N. J., Khan S. H., Stover C. M., Sandrini S. M., Marston D., Presanis J. S., Schwaebler W. J., 2005. Composition of the lectin pathway of complement in Gallus gallus: Absence of mannan-binding lectin-associated serine protease-1 in birds. *J. Immunol.* 174: 4998-5006.
14. Matsushita M., Endo Y., Fujita T., 1998. MASP1 (MBL-associated serine protease 1). *Immunobiol* 199(2): 340-347.
15. Norman D. G., Barlow P. N., Baron M., Day A. J., Sim R. B., Campbell I. D., 1991. Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* 219(4): 717-725.
16. Perry S. E., Robinson P., Melcher A., Quirke P., Bühring H. J., Cook G. P., Blair G. E., 2007. Expression of the CUB domain containing protein 1 (CDCP1) gene in colorectal tumour cells. *FEBS Lett* 581(6): 1137-1142.
17. Rawlings N. D. and Barrett A. J., 1994. Families of serine peptidases. *Methods Enzymol* 244: 19-61.
18. Schwaebler W., Dahl M. R., Thiel S., Stover C., Jensenius J. C., 2002. The mannan-binding lectin-associated serine proteases (MASPs) and MASP-3: four components of the lectin pathway activation complex encoded by two genes. *Immunobiol* 205(4 5): 455-466.
19. 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.

20. Sørensen R., Thiel S., Jensenius J. C., 2005. Mannan-binding-lectin-associated serine proteases, characteristics and disease associations. *Springer Semin Immunopathol* 27(3): 299-319.

21. Stover C. M., Lynch N. J., Dahl M. R., Hanson S., Takahashi M., Frankenberger M., Ziegler-Heitbrock L., Eperon I., Thiel S., Schwaebler W. J., 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.

22. Takada F., Seki N., Matsuda Y. I., Takayama Y., Kawakami M., 1995. Localization of the genes for the 100-kDa complement activating components of R-reactive factor (CRARF and Crarf) to human 3q27-q28 and mouse 16B2-B3. *Genomics* 25: 757-759.

23. Takahashi M., Ishida Y., Iwaki D., Kanno K., Suzuki T., Endo Y., Homma 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.

24. Takahashi M., Iwaki D., Kanno K., Ishida Y., Xiong J., Matsushita M., Endo Y., Miura S., Ishii N., Suganuma 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.

25. 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.

26. Varela P. F., Romero A., Sanz L., Romão M. J., Töpfer-Petersen E., Calvete J. J., 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, Klaus Wimmers

Summary

Gene encoding 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 genes (MASP1, MASP2 and MASP3). Therefore, the current study aimed at sequencing cDNA and analyzing molecular structure of the MASP3 (MASP1 isoform 2). With 2,657 bp and 11 exons, in which the coding region is from nucleotide nt. 152-2388, MASP3 encoded 728 amino acids. There was high homology among mammalian species (pig, human, cattle and dog) at both cDNA (87-90%) and protein (92-94%) levels. Macromolecule MASP3 had 27 cysteine residues. However, amino acid with high percent in its protein 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 kilo Daltons, MASP3 was an assembly of rich-cysteine protein domains such as CUB1 (a.a 18-138, 2 cysteine), EGF_CA (a.a 139-182, 6 cysteine), CUB2 (a.a 185-297, 3 cysteine), CCP1 (a.a 301-362, 4 cysteine), CCP2 (a.a 367-432, 4 cysteine) in N-terminus and Tryp_SPC (a.a 449-711, 5 cysteine) in C-terminus. The Tryp_SPC is the largest domain whereas EGF_CA is the smallest. Besides those, a signal peptide (a.a 1-19), 13 disulfide bonds, 5 N-glycosylation sites, and methionine loop (a.a. 630-649) were also identified. These are as major premises for further studies for roles of the MASP3 in immune response mechanism for disease resistance in pigs.

Key words: *MASP3, cDNA, protein, molecular characteristics.*

Người phản biện: TS. Lê Quang Trung

Ngày nhận bài: 14/9/2012

Ngày thông qua phản biện: 25/12/2012

Ngày duyệt đăng: 3/01/2013