

ĐẶC ĐIỂM GEN MÃ HÓA MANNAN-BINDING LECTIN SERINE PROTEASE 1 Ở LỢN

Đỗ Văn Anh Khoa¹, Klaus Wimmers²

¹Đại học Cần Thơ, Việt Nam

²Viện Sinh học Gia súc FBN-Dummerstorf, Cộng hòa Liên bang Đức

TÓM TẮT

Phức hợp bổ thể được kích hoạt bằng ba con đường khác nhau - con đường cổ điển, nhánh và lectin. Trong sự hiện diện của carbohydrate trên bề mặt vi khuẩn, manose-binding lectin (MBL) kết nối với các thành viên của gia đình MBL-serine protease (MASP1, MASP2 và MASP3) để sản sinh proenzyme kích hoạt các thành phần C4, C2 và C3 của phức hợp bổ thể theo con đường lectin và đường nhánh. Tuy nhiên, các serine protease này chưa được nghiên cứu ở lợn. Vì vậy, nghiên cứu này sẽ tập trung giải mã trình tự chuỗi cDNA cũng như phân tích cấu trúc protein của phân tử MASP1. Kết quả đã đọc được toàn bộ chiều dài 3.009 nucleotide của chuỗi cDNA mã hóa 699 amino acid (a.a). Gen MASP1 có 16 exon và protein của nó chứa đựng các miền chức năng giàu cysteine như CUB (a.a 18-138, 121 a.a, 2 cysteine), EGF C'A (a.a 139-182, 44 a.a, 6 cysteine), CUB (a.a 185-297, 113 a.a, 4 cysteine) ở đầu N-terminus, hai miền CPP (a.a 301-362, 62 a.a, 4 cysteine và a.a 367-432, 66 a.a, 4 cysteine) nằm ở vùng trung tâm và cuối cùng là miền Tryp_SPe (a.a 448-691, 244 a.a, 7 cysteine) ở đầu C-terminus. Signal peptide, vị trí N-glycosylation, vị trí catalytic và các cysteine được bảo tồn cao giữa người và lợn. So với người, protein MASP1 ở lợn giống khoảng 88%. Ngoài ra còn có 3 điểm đa hình nucleotide đơn trên gen MASP1 đã được tìm thấy tại các locus 1 650A→G (Asp→Gly, exon 12), 1 748C→T (Leu→Phe, exon 13) và 1 807C→T (exon 13) ở quần thể giống lợn Landrace. Đây là nguồn thông tin có giá trị trong công tác chọn lọc giống lợn kháng bệnh trong tương lai.

Từ khóa: MASP1, cDNA, protein, đặc điểm phân tử

ĐẶT VẤN ĐỀ

Phức hợp bổ thể đóng vai trò quan trọng trong sự đáp trả miễn dịch ban đầu và miễn dịch thụ nhận (innate và adaptive immune response). Sự kích hoạt phức hợp bổ thể theo 3 con đường khác nhau (cổ điển, nhánh và lectin) sẽ hình thành phức hợp tấn công màng tế bào (membrane attack complex, MAC), một cơ chế kháng khuẩn và tiêu diệt tế bào mục tiêu của vật chủ, giúp tăng khả năng thực bào và cytotoxicity (Stover *et al.*, 2003, Wimmers *et al.*, 2003). Ở con đường lectin (lectin pathway), mannan-binding lectin (MBL), L-ficolin và H-ficolin là những oligomer có những nét đặc trưng khác nhau về cấu trúc polysaccharide (Holmskov *et al.*, 2000; Matsushita *et al.*, 2000, 2001). Ba protein này có cùng chức năng là hình thành các phức hợp kích hoạt miễn dịch bằng cách kết dính với các serine proteases MASP1, MASP2 và MASP3 (Matsushita *et al.*, 2000; 2002, Takayama *et al.*, 1994; Thiel *et al.*, 1997; Dahl *et al.*, 2001). Ở đó, MASP1 tham gia vào quá trình kích hoạt bổ thể bằng cách trực tiếp chia tách thành phần C3 (Matsushita *et al.*, 1995; 2000) và tham gia kích hoạt con đường nhánh (Takahashi *et al.*, 2010), trong khi MASP2 tham gia vào quá trình kích hoạt bổ thể bằng cách chia tách

thành phần C4 (thành C4a và C4b) và C2 (thành C2a và C2b) để từ đó hình thành nên C3 convertase, một enzyme chính trong quá trình thành lập MAC (Matsushita *et al.*, 2000; Schwaible *et al.*, 2002). Nhiều nghiên cứu đã nhận diện đặc điểm phân tử và một số chức năng của MASP1 ở các loài động vật khác nhau: chuột (Stover *et al.*, 2003), gà (Lynch *et al.*, 2005), người (Endo *et al.*, 1998). Vì vậy, trong nghiên cứu này, đặc điểm phân tử cDNA và protein MASP1 ở lợn sẽ được phân tích.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Nguyên vật liệu (động vật, hóa chất, thiết bị và dụng cụ) và phương pháp thí nghiệm (phân lập mRNA, tổng hợp cDNA, giải mã trình tự chuỗi, thiết kế mồi, thành phần phản ứng PCR) được sử dụng cho nghiên cứu này đã được mô tả trước đây (Đỗ Văn Anh Khoa, 2010).

Thí nghiệm sử dụng 6 cặp mồi đặc hiệu được thiết kế dựa theo chuỗi cDNA của gen MASP1 để khuếch đại các đoạn gen tương ứng có chiều dài 462-904 bp. Vị trí của các cặp mồi trên GenBank được trình bày qua bảng 1. Chu trình nhiệt thực hiện phản ứng PCR cho từng cặp mồi được mô tả trong

hàng 2. Các phần mềm online được sử dụng để phân tích kết quả gồm: ORF Finder (Open Reading Frame Finder)(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) nhận diện cấu trúc bậc 1 protein. Primer3 (<http://frodo.wi.mit.edu/>); thiết kế mô. ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>); so sánh mức độ tương đồng của chuỗi cDNA 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 signal peptide. 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 disulfide. NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>); nhận diện vị trí N-glycosylation. NetOGlyc 3.1 Server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) nhận diện vị trí O-glycosylation.

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

Tên mồi	Mỗi xuôi fw 5'-3' (vị trí trên GenBank GU810082)	Mỗi ngược rev 3'-5' (vị trí trên GenBank GU810082)	Chiều dài, bp
MASP1_1	aatgagggtggctctctctc (nt 151-171)	tcacaggacagttctctctc (nt 592-612)	462
MASP1_2	gagaccacggacacagaaca (nt 431-450)	caagaagtlttggaccagctt (nt 900-920)	490
MASP1_3	gagctccgaatgctctacacc (nt 775-796)	tcctccaggactcattcctc (nt 1411-1430)	655
MASP1_4	gccagcagccctactacaag (nt 1341-1360)	ctctccaggagaccctctg (nt 2225-2244)	904
MASP1_5	gtgcagtgactctggaggt (nt 2076-2095)	agacacactggcctctctc (nt 2635-2654)	579
MASP1_6	tgagcccgagattctctc (nt 2492-2511)	tgcagagtcacaggaagctg (nt 2990-3009)	518

Bảng 2. Chu trình nhiệt PCR

Tên mồi	Biến tính ban đầu (thời gian)	Biến tính (thời gian)	Gán mồi (thời gian)	Tổng hợp (thời gian)	Kết thúc (thời gian)	Giữ mẫu (thời gian)	Tổng số chu kỳ
MASP1_1			56°C (30")				
MASP1_2			60°C (30")				
MASP1_3	94°C (5')	94°C (30")	60°C (30")	72°C (1')	72°C (5')	15°C (∞)	35
MASP1_4			58°C (30")				
MASP1_5			56°C (30")				
MASP1_6			56°C (30")				

KẾT QUẢ THẢO LUẬN

Đặc điểm chuỗi cDNA và protein

Sử dụng các cặp mồi đặc hiệu để khuếch đại 6 đoạn gen có chiều dài khác nhau (từ 462-904 bp) và dán xen nhau dưới sự hỗ trợ của kỹ thuật PCR và kỹ thuật giải mã được trình tự chuỗi (sequencing) đã nhận diện được chiều dài chuỗi cDNA của gen MASP1 dài 3009 bp (GenBank GU810082). Vùng 5'-UTR dài 151 bp (nt. 1-151) và 3'-UTR dài 758 bp (nt. 2.252-3.009) Đầu hiệu polyadenylation được nhận diện trong chuỗi AATAAA (nt 2621-2626). Vị trí codon khởi đầu (start codon) ATG từ nt. 152-154 và codon kết thúc (stop codon) TGA từ nt. 2.249-2.251. Như vậy, vùng mã hóa (coding region, CDS)

nằm trong khoảng nt. 152-2.251 tương ứng với 699 amino acid (GenBank ADF30842) theo khung đọc mở (Open Reading Frame Finder ORF). Kết quả so sánh chuỗi cDNA (GenBank GU810082) với chuỗi DNA (GenBank NW_001885220, Ssc13_WGA1011_1:115459-225459 *Sus scrofa* chromosome 13 genomic contig) đã nhận diện được 16 exon của gen MASP1 dựa theo nguyên tắc cấu trúc DNA (exon 1-GT.intron.AG-exon 2). Số lượng exon của gen MASP1 ở lợn và người là giống nhau.

Khối lượng phân tử protein MASP1 khoảng 79,116 kDa. Chuỗi polypeptide của MASP1 được cấu trúc bởi các amino acid, trong đó Gly, Leu và Ser chiếm > 7%; Glu, Val, Pro và Thr chiếm > 6%; Asp, Ile và Lys chiếm ≥ 5%; Cys, Phe, Gln, Arg và

Tyr chiếm $\geq 4\%$ và mỗi thành phần amino acid khác chiếm tỷ lệ $< 4\%$. Đại phân tử MASPI có 29 cysteine, từ đây có thể hình thành nên 14 cầu nối disulfide giữa các Cys¹¹⁻⁴³², Cys⁷³⁻²¹², Cys⁹¹⁻¹⁶⁶, Cys¹⁴³⁻²⁶⁰, Cys¹⁵³⁻¹⁵⁷, Cys¹⁶⁸⁻³²⁹, Cys¹⁸¹⁻¹⁸⁵, Cys²⁴²⁻³⁰¹, Cys³⁴⁹⁻⁴⁷⁵, Cys³⁶²⁻⁴⁹¹, Cys³⁶⁷⁻⁴³⁶, Cys³⁹⁷⁻⁵⁷², Cys⁶¹¹⁻⁶³¹ và Cys⁶⁴²⁻⁶⁷². Ở mức độ protein, phân tử MASPI của người và lợn có mức tương đồng khoảng 88%

Đọc theo chuỗi polypeptide của MASPI, vị trí tách rời peptide tín hiệu (signal peptide) được nhận diện giữa a.a 19 và 20, tiếp theo là các miền protein chức năng như CUB (a.a 18-138, 121 a.a, 2 cysteine), EGF_CA (a.a 139-182, 44 a.a, 6 cysteine), CUB (a.a 185-297, 113 a.a, 4 cysteine) ở đầu N-, hai miền CPP (a.a 301-362, 62 a.a, 4 cysteine và a.a 367-432, 66 a.a, 4 cysteine) nằm ở

vùng trung tâm và cuối cùng là miền Tryp_SPC (a.a 448-691, 244 a.a, 7 cysteine) ở đầu C. Miền protein lớn nhất ở MASPI là Tryp_SPC có nhiều cysteine nhất (7 cysteine) và miền protein nhỏ nhất EGF_CA có số lượng cysteine khá nhiều (6 cysteine). Các vị trí dễ xúc tác được tìm thấy trong miền Tryp_SPC là His⁴⁹¹, Asp⁵⁷² và Ser⁶⁴⁶. Các vị trí cần thiết cho trung tâm hoạt động serine protease trên lợn cũng tương tự như ở người và gần giống với *Xenopus laevis* His⁴⁸⁹, Asp⁵⁷¹ và Ser⁶⁴⁵ (Endo *et al.*, 1998; Kraut, 1977). Vị trí chưa tách nằm giữa Agr⁴⁴⁸ và Ile⁴⁴⁹ thì cần thiết cho sự sản sinh hai chuỗi polypeptide kích hoạt riêng biệt của MASPI ở người (Kraut, 1977) cũng được nhận diện cùng vị trí với lợn Histidine loop CCGSLLGSSWIVTAAHC (từ Cys⁴⁷⁴-Cys⁴⁹¹) được nhận diện đồng chuỗi giữa lợn và người (Arlaud, Gagnon, 1981).

Bảng 3. Vị trí của các exon được xác định trên các GenBank

Exon (chiều dài, bp)	Vị trí trên GenBank GU810082	Vị trí trên GenBank NW_001885220	Exon (chiều dài, bp)	Vị trí trên GenBank GU810082	Vị trí trên GenBank NW_001885220
1 (156)	nt. 1-156	nt 6 106-6 261	9 (138)	nt 1 242-1 379	nt 47 782-47 919
2 (232)	nt. 157-388	nt 13 508-13 739	10 (75)	nt. 1 380-1 454	nt 48 724-48 798
3 (178)	nt. 339-566	nt 32 888-33 065	11 (138)	nt 1 455-1 592	nt 59 230-59 367
4 (132)	nt 567-698	nt 34 689-34 820	12 (114)	nt 1 593-1 706	nt 63 060-63 173
5 (197)	nt 699-895	nt 36 977-37 173	13 (186)	nt 1 707-1 892	nt 64 037-64 222
6 (148)	nt 896-1 043	nt 38 896-39 043	14 (68)	nt 1 893-1 960	nt 66 205-66 272
7 (119)	nt 1 044-1 162	nt 40 356-40 474	15 (100)	nt 1 961-2 060	nt 69 103-69 202
8 (79)	nt 1 163-1 241	nt 41 458-41 536	16 (949)	nt 2 061-3 009	nt 69 644-70 592

Bảng 4. Tỷ lệ các thành phần amino acid trong chuỗi protein

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

Người	MRWLLLYALCFSLSKASAHTVELNNMFGQIQSPGYPDSYPSDSEVTWFITVPGDFRIKL	60
Lợn	MRWLLLSHALCFSLKASAHVVELNNMFGQIQSPGYPDSYPSDSEVTWFITVPEGFRIKL	60

Người	YFMHFNLESSYLCEYDYVKVETEDQVLATFCGRETDTDEQTGGQEVVLSPGSFMSITFRS	120
Lợn	YFMHFNLESSYLCEYDYVKVETEDQVLATFCGKETTDTDEQTGGQEVVLSPGSFMSITFRS	120

Người	DFSNEERFTGFDAHYMAVDVDECKEREDEELSCDHYCHNYIGGYYCSCRFYILHTDWRRT	180
Lợn	DFSNEERYTGFDAHYMAVDVDECKEREDEELSCDHYCHNYIGGYYCSCRFYILHTDWRRT	180

Người	CRVECSNDLFTQRTGVTSPDFPNPYKSSSECLYTIIELEEGFMVNLQFEDI FDI EDHPEV	240
Lợn	CRVECSNDLFTQRTGVTSPDYSPYKSSSECLYTIIELEEGFMVSLQFEDI FDI EDHPEV	240

Người	PCPYDYIKIKVGPVKVLGPFCEKAKEPIISTQSHSVLILFHSDNSGENRGWRLLSYRAAGNE	300
Lợn	SCPYDYIKIKAGPKLLGPFCEKAKEPIINTQSHSIQIQFRSDNSGENRGWRLLSYRATGNE	300

Người	CPELQPPVHGKIEPSQAKYFFKDQVLVSCDGTGYKVLKDNVEMDTFQIECLDKGTWSNKIP	360
Lợn	CPKLGPPVHGKIEPLQAKYSEFKDQVLVSCDGTGYKVLKDNVEMDVFQIECLDKGTWSNKIP	360

Người	TCKIVDCRAPGELEHGLITFFSTRNNLTYYKSEIKYSCQEPYYKMLNNTGIYTC SAQGVW	420
Lợn	TCKIADCGTPAKLENGLVTFSTRNNLTYYKSEI IYSCQPPYYKMLHSTTGVYTC SAQGVW	420

	↓	
Người	MNKVLGRSLPTCLPVCGLPKFSRKLARI FNGRPAQKGTTPWIAMLSHLNGQPFCGSSLL	480
Lợn	MNEVLGKSQPTCLPVCGLTPRFSRNLARI FNGHPAQKGTTPWIAMLSHPNGQPFCCGSSLL	480

Người	GSSWIVTAAHCLHQSLDPEDPTLRDSDLLSPDFKILIGKHWLRSDENEQHLGVKHTTL	540
Lợn	GSKWIVTAAHCLHQSLGSEDPTVHVLDDLSPSDFRIIMGKHWRRRSDENEQSFTAKQIFL	540

Người	HPQYDPNTFENDVALVELLESPLVNAFVMPICLPEGPQEQEGAMVIVSWGKQFLQRFPET	600
Lợn	HPQYPTTFENDVALVELLESPLVNDVMPICLPOGPPQEQEGAMVIVSWGKQFLQRFPET	600

Người	LMEIEIPIVDHSTQKAYAPLKKKVTRDMICAGEKEGGKDACAGDSGGPMVTLNRRGQW	660
Lợn	LMEIEIPIVDHRIQEQAYAPLKKKVTRDMICAGEKEGGKDACAGDSGGPMVTLNRRGQW	660

Người	YLVGTVSWGDDCGKKDRYGVYSYIHNKDWIQRVTGVRN	699
Lợn	YLVGTVSWGVDCCGKKDRYGVYSYIFYNKDWIQRVTGMRN	699

Hình 1. So sánh chuỗi protein giữa người (GenBank NP_001870) và lợn (GenBank ADF30842). Các điểm đột biến được in đậm. Các vị trí asparagine (N) mang dấu hiệu N-glycosylation được in viền (outline). Các cysteine được gạch dưới. Threonine (T) là O-glycosylation được in nghiêng đậm. Các amino acid (tâm xúc tác (catalytic site) được in trong khung. Vị trí chia tách giữa arginine (R) và isoleucine (I) để sản sinh ra hai chuỗi protein (heavy and light chains) được đánh dấu bằng mũi tên (↓). Chuỗi của vòng histidine loop được gạch dưới liên tục. Các catalytic site được in nghiêng và gạch dưới. Khẽ hở đặc trưng của chất nền (substrate specificity crevice) được in nghiêng đậm và gạch dưới đối. Vị trí amino acid liên quan đến serine hoạt hóa được in nghiêng đậm và gạch dưới (Endo *et al.*, 1998).

Ngoài ra, dấu hiệu N-glycosylation được tìm thấy ở 3 vị trí asparagine khác nhau Asp⁴⁹, Asp¹⁷⁸ và Asp⁵⁴⁵ tương ứng với các chuỗi amino acid NITV,

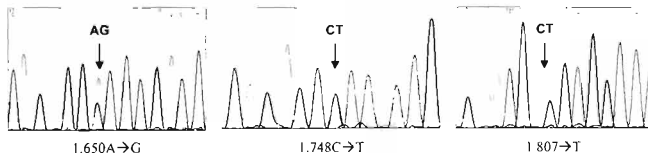
NRTC và NPPT ở lợn, trong khi ở người chỉ có 2 vị trí đầu tiên (Endo *et al.*, 1998). Thông thường, N-glycosylation sites được nhận diện ở kiểu chuỗi Asp-

Xaa-Ser/Thr (trong đó Xaa không phải là Pro) (Gupta *et al.*, 2004). Trong khi đó, threonine tại vị trí 438 (với G-score = 0.212 và T-score = 0.535) được biết như là một vị trí O-glycosyl hóa. O-GalNAc-glycosylation là một trong những dạng chính glycosylation ở tế bào động vật hữu nhũ (Julenius *et al.*, 2005). Nó được biết rằng, glycosylation là sự chỉnh sửa sau chuyển mã (post-translational modification), có ảnh hưởng đến cấu trúc (folding), định vị (localization), vận chuyển protein (trafficking), tính hòa tan (solubility), đặc tính kháng nguyên (antigenicity) và hoạt tính sinh học (biological activity)... của protein, cũng như sự tương tác giữa các tế bào (cell-cell interactions) (Seitz, 2000; Gupta *et al.*, 2004). Một trong những dạng glycosylation có nhiều trong tế bào động vật hữu nhũ là khi một N-acetylgalactosamine (GalNAc) được kết nối α -1 với nhóm hydroxyl của serine hoặc threonine. Đây cũng được gọi là kiểu mucin-type.

Kiểu này có thành phần carbohydrate cao >50% và cũng là thành phần chính của dịch nhầy (mucus) - chất gel bảo vệ bề mặt biểu mô trong quá trình khử nước (dehydration), tổn thương cơ học (mechanical injury), proteases và các mầm bệnh (Carraway, Hull, 1991; Srouso, Dekker, 1992). Vì vậy, sự dư thừa glycosylation sites thì cực kỳ có giá trị trong nhận diện đặc điểm phân tử của những protein mới, i.e. hỗ trợ giải thích kết quả đo quang phổ xác định phân tử khối của protein. Việc xác định mucin-type O-glycosylation có thể được xem như là phần không thể thiếu được khi dự đoán chức năng của một protein lạ (Jensen *et al.*, 2002, 2003) bởi vì O-glycosylation có ảnh hưởng đến cấu trúc của protein. Sự tiên đoán này có ý nghĩa quan trọng trong việc giữ lại hay hủy bỏ vị trí O-glycosylation khi thiết kế những yếu tố ức chế cạnh tranh (competitive inhibitors) của glycosyltransferases (Hansen *et al.*, 1998).



Hình 2. Các miền chức năng của protein MASP1



Hình 3. Mẫu đại diện cho các kiểu gen di hợp tử tại các điểm đa hình

Đa hình đơn nucleotide

Kết quả đọc trình tự toàn bộ chiều dài cDNA của gen MASP1 đã phát hiện 3 điểm đa hình nucleotide tại các vị trí 1.650A→G (exon 12), 1.748C→T (exon 13) và 1.807C→T (exon 13), tương ứng với các codon 500GAC→GGC, 533CTC→TTC và 552GAC→GAT mã hóa các amino acid Asp→Gly, Leu→Phe và Asp. Kết quả đọc trình tự 33 mẫu cDNA của giống lợn Landrace cho thấy tần số các

kiểu gen và alen tại (i) đột biến điểm 1.650A→G AA = 72,73% (n = 24), AG = 21,21% (n = 7) và GG = 6,06% (n = 2); alen A = 0,83% và G = 0,17 %, (ii) đột biến điểm 1.748C→T: CC = 0,00 % (n = 0), CT = 6,06% (n = 2) và TT = 93,94% (n = 31); alen C = 0,03% và T = 0,97%, và (iii) đột biến điểm 1.807C→T: CC = 3,03% (n = 1), CT = 12,12% (n = 4) và TT = 84,85% (n = 28), alen C = 0,09% và T = 0,91%. Ở người, Sirmaci *et al.* (2010) đã chỉ ra rằng hai đột biến điểm trên gen MASP1 c.2059G>A

(p.G687R) và c.870G>A (p.W290X) là nguyên nhân dẫn đến những dị tật ở mặt, tai... (malformation syndrome) trong giai đoạn bào thai.

KẾT LUẬN

Nghiên cứu đã giải mã trình tự nucleotide chuỗi cDNA và phân tích đặc điểm protein cấu trúc bậc một của phân tử MASP1. Gen mã hóa protein MASP1 ở lợn và người có mức độ tương đồng khá cao về đặc điểm phân tử, các vùng miền protein chức năng, kể cả các vị trí chức năng đặc biệt khác. Đa hình gen MASP1 đã được phát hiện ở lợn. Đây là tiền đề cơ bản cho những nghiên cứu tiếp theo về vai trò và chức năng của gen trong hệ miễn dịch của vật chủ, một trong những mục tiêu của các chương trình chọn lọc giống kháng bệnh trong tương lai.

Lời cảm ơn: Nghiên cứu nhân được sự hỗ trợ kinh phí của BMBF/DLR trong khuôn khổ dự án VN03/B01

TÀI LIỆU THAM KHẢO

Arlaud GJ, Gagnon J (1981) C1r and C1s subcomponents of human complement, two serine proteases lacking the "histidine-loop" disulphide bridge. *Biosci Rep* 1: 779-784.

Carraway KL, Hull SR (1991) Cell surface mucin-type glycoproteins and mucin-like domains. *Glycobiology* 1: 131-138.

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 mannan-binding lectin complement activation pathway. *Immunity* 15: 127-135.

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.

Gupta EJ, Brunak S (2004) Prediction of N-glycosylation sites in human proteins. In preparation. <http://www.cbs.dtu.dk/services/NetNGlyc/abstract.php>.

Hansen JE, Lund O, Gooley AA, Williams KL, Brunak S (1998) NetOGlyc. Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycocon J* 15: 115-130.

Holmskov UL (2000) Collectins and collectin receptors in innate immunity. *APMIS* 108(Suppl): 1-59.

Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Stærfeldt HH, Rapacki K, Workman C, Andersen CA, Knudsen S, Krogh A, Valencia A, Brunak S (2002) Prediction of human protein function from post-translational modifications and localization features. *J Mol Biol* 319: 1257-1265.

Jensen LJ, Gupta R, Stærfeldt HH, Brunak S (2003) Prediction of human protein function according to Gene Ontology categories. *Bioinformatics* 19: 635-642.

Julenius KA, Gupta MR, Brunak S (2005) Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15:153-164.

Kraut J (1977) Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem* 46:331-358.

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 mannan-binding lectin-associated serine protease-1 in birds. *J Immunol* 174: 4998-5006.

Matsushita M, Endo Y, Fujita T (2000) Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease. *J Immunol* 164: 2281-2284.

Matsushita M, Fujita T (2001) Ficolins and the lectin complement pathway. *Immunol Rev* 180: 78-85.

Matsushita M, Fujita T (1995) Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. *Immunobiol* 194: 443-448.

Matsushita M, Kuraya M, Hamasaki N, Tsujimura M, Shiraki H, Fujita T (2002) Activation of the lectin complement pathway by H-ficolin (Hakata antigen). *J Immunol* 168: 3502-3506.

Matsushita M, Thiel S, Jensenius JC, Terai I, Fujita T (2000) Proteolytic activities of two types of mannose-binding lectin-associated serine protease. *J Immunol* 165: 2637-2642.

Schwaebler W, Dahl MR, Thiel S, Stover C, Jensenius JC (2002) The mannan-binding lectin-associated serine proteases (MASPs) and Map19: four components of the lectin pathway activation complex encoded by two genes. *Immunobiol* 205(4-5): 455-466.

Sirmaci A, Walsh T, Akay H, Spiliopoulos M, Sakalar YB, Hasaneldiöglü-Bayrak A, Duman D, Farooq A, King MC, Tekin M (2010) MASP1 mutations in patients with facial, umbilical, coecygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am J Hum Genet* 87(5): 679-686.

Stover CM, Lynch NJ, Dahl MR, Hanson S, Takahashi M,

Frankenberger M, Ziegler-Heitbrock L, Eperon I, Thiel S, Schwaeble 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

Strous GJ, Dekker J (1992) Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol* 27: 57-92.

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

Takayama Y, Takada F, Takahashi A, Kawakami M (1994) A 100-kDa protein in the C4-activating component of Ra-reactive factor is a new serine protease having

module organization similar to C1r and C1s. *J Immunol* 152: 2308-2316.

Thiel S, Vorup-Jensen T, Stover CM, Schwaeble M, Laursen SP, Poulsen K, Willis AC, Eggleton P, Hansen S, Holmskov U, Reid KBM, Jensenius JC (1997) A second serine protease associated with mannan-binding lectin that activates complement. *Nature* 386: 506-510

Wimmers K, Mekchay S, Schellander K, Ponsuksili S (2003) Molecular characterization of the pig C3 gene and its association with complement activity. *Immunogenet* 54(10): 714-724

Seitz O (2000) Glycopeptide synthesis and the effects of glycosylation on protein structure and activity. *Chembiochem* 1: 214-246.

STRUCTURAL CHARACTERISTICS OF THE GENE ENCODING PORCINE MANNAN-BINDING LECTIN SERINE PROTEASE 1 IN PIG

Do Vo Anh Khoa^{1,*}, Klaus Wimmers²

¹Can Tho University, Vietnam

²Leibniz Institute for Farm Animal Biology, Germany

SUMMARY

Complement system is activated by three different ways - classical, alternative and lectin pathways. In presence of carbohydrate on the surface of microbes, mannose-binding lectin (MBL) binds to members of the MBL-serine protease family (MASP1, MASP2 and MASP3) to generate proenzymes activating C4, C2, and C3 of complement system via the lectin and alternative pathways. However, these serine proteases have not been identified in pig yet. Therefore, this study was mainly focused on sequencing of cDNA and analyzing of structural protein of MASP1 molecule in pig. It was indicated that 3 009 nucleotide of MASP1 cDNA encoding 699 deduced amino acids was sequenced. The MASP1 gene had 16 exons and its encode polypeptide protein contained rich-cysteine domains such as CUB1 (a.a 18-138, 121 a.a, 2 cysteine), EGF_CA (a.a 139-182, 44 a.a, 6 cysteine), CUB2 (a.a 185-297, 113 a.a, 4 cysteine) in N-terminus, two CPP domains (CCP1: a.a 301-362, 62 a.a, 4 cysteine and CCP2: a.a 367-432, 66 a.a, 4 cysteine) at the center of protein and Tryp_SPC domain (a.a 448-691, 244 a.a, 7 cysteine) in C-terminus. The putative signal peptide, N-glycosylation sites, catalytic sites and cysteines were also highly conserved between human and pig. Compared with human, porcine MASP1 showed 88% identity at protein level. Additionally, three single nucleotide polymorphisms at loci 1 650A→G (Asp→Gly, exon 12), 1 748C→T (Leu→Phe, exon 13) và 1 807C→T (exon 13) were also found in a population of Landrace breed. This is valuable information for breeding program towards disease resistance in pigs.

Keywords: MASP1 gene, cDNA, protein, molecular characteristics