JOURNAL OF HUMAN GENETICS (2003) 48(2):82-85.

Haplotype analysis of the human collectin placenta 1 (hCL-P1) gene

Hiroyuki, Ohmori ; Yoshio, Makita ; Makiko, Funamizu ; Shin-ichi, Chiba ; Katsuki, Ohtani ; Nobutaka, Wakamiya ; Akira, Hata

Haplotype analysis of human collectin placenta 1 (hCL-P1) gene

Hiroyuki Ohmori, Yoshio Makita, Makiko Funamizu, Shin-ichi Chiba, Katsuki Ohtani, Nobutaka Wakamiya, Akira Hata

H. Ohmori, Y. Makita, M. Funamizu, S. Chiba, A. Hata*

Department of Public Health, Asahikawa Medical College, Midorigaoka-higashi 2-1-1-1, Asahikawa 078-8510, Japan

Tel.: +81-166-68-2410; Fax: +81-166-68-2419

e-mail: ahata@asahikawa-med.ac.jp

*corresponding author

K. Ohtani, N. Wakamiya

Department of Microbiology, Asahikawa Medical College

Abstract

Collectins are a family of C-type lectins that have collagen-like sequences and carbohydrate recognition domains (CRD). They are involved in host defense through their ability to bind to carbohydrate antigens of microorganisms. The scavenger receptors type A and macrophage receptor with collagenous structure are classical type scavenger receptors that have internal collagen-like domains. We previously described a new scavenger receptor that is membrane-type collectin from placenta (collectin placenta 1(CL-P1)). CL-P1 is a type II membrane protein, has a coiled coil region, a collagen-like domain, and a CRD. We found that CL-P1 can bind and phagocytes both bacteria and yeast. Furthermore, it reacts with oxidized low-density lipoprotein (OxLDL) but not with acetylated LDL (AcLDL). These results indicate that CL-P1 might play important roles in host defenses and/or atherosclerosis formation (Ohtani et al 2001). One rational strategy to study the role of CL-P1 in these pathological conditions would be a haplotype association study using human samples. As a first step for this strategy, we analyzed haplotype structure of the CL-P1 gene. Sequencing the CL-P1 gene region of ten Japanese volunteers identified five single-nucleotide polymorphisms (SNPs) with the minor allele frequency to be at least 29%. In order to obtain SNPs in the 5'-upstream region of the gene, total 20 SNPs described in publicly available database were screened, and we found one SNP out of 20 was useful for the present study. Thus, total six SNPs, one in 5'-upstream region, two in intron 2, one in exon 5, and two in exon 6, were employed to analyze the haplotype structure of the gene with DNAs derived from 54 individuals (108 alleles). The analysis revealed that only two of six SNPs showed significant linkage disequilibrium (r^2>0.5) between each other.

This haplotype information would be useful in disease-association studies where a contribution of CL-P1 gene has been suspected, especially in innate immunity defect or atherosclerosis. Two SNPs in exon 6, both lead to amino acid substitutions, could be attractive candidates to have some influence for disease susceptibility.

Key words: Single-nucleotide polymorphisms, Linkage disequilibrium, Haplotype, Collectin, Japanese population, host defense, amino acid substitution

Introduction

Collectins are a family of proteins that contain at least two characteristic structures, a collagen-like region and a carbohydrate recognition domain (CRD) (Drickamer et al. 1988). These lectins are found in vertebrates from avians to humans (Laursen et al. 1998). There are four groups of collectins: the mannan-binding protein (MBP) group, the surfactant protein A (SP-A) group (White et al. 1985), the surfactant protein D (SP-D) group (Persson et al. 1989), and the newly isolated group, collectin liver 1 (CL-L1) (Ohtani et al. 1999) and collectin placenta 1 (CL-P1) (Ohtani et al. 2001). MBP can destroy bacteria through activation of the complement pathway (Kawasaki et al. 1989) or opsonization via collectin receptors (Schweinle et al. 1989). MBP and conglutinin of the SP-D group are β-inhibitors of influenza A viruses that have hemagglutination inhibition and neutralization activities (Wakamiya et al. 1992; Hartley et al. 1992). SP-A amplifies the phagocytosis of bacteria by macrophages (Pikaar et al. 1995) and opsonizes herpes simplex virus (HSV) (van Iwaarden et al. 1992). SP-D agglutinates bacteria (Kuan et al. 1992) and has hemagglutination inhibition activity against influenza A virus (Hartshorn et al. 1994). These activities indicate that collectins play an important role in innate immunity. In addition, the type A scavenger receptor (SR-A) also contains a collagen-like domain, which forms a triple helical structure and binding sites (Kodama et al. 1990) that have a broad specificity for ligands. The primary function of scavenger receptors could be the destruction and neutralization of pathogens by endocytosis and phagocytosis. Recent knockout data show that SR-AI-deficient mice are sensitive to *Listeria monocytogenes* and herpes simplex virus type-1 infections (Suzuki et al. 1997). Thus, it appears that scavenger receptors also

have a role in innate immunity.

Recently, we cloned human CL-P1 gene (hCL-P1) and characterized its biological function, and found that it was a membrane-type collectin that functions as a scavenger receptor. Expression of this gene was found mainly in endothelial cells but not in monocyte-macrophage lineage cells. Furthermore, we found CL-P1 can bind and phagocytose bacteria and yeast as well as oxidized LDL (Ohtani et al. 2001). With these results, we hypothesized that this gene could play an important role in the human innate immunity system and/or atherosclerosis formation. In order to study this possibility, we are now collecting human samples those are susceptible to infection especially in childhood and also patients with atherosclerosis without lipid profile abnormalities. Haplotype association study would be the choice to this research purpose. Here, we reported the haplotype structure of the CL-P1 gene obtained from six single-nucleotide polymorphisms (SNPs) genotyping with 108 alleles in Japanese.

Materials and Methods

Identification of SNPs

In order to identify useful SNPs in the gene region, amplified DNAs derived from ten Japanese volunteers (20 alleles) were sequenced. PCR primers were chosen at approximately 500bp intervals, with about 80bp overlap (Shinohara et al. 2001). Total sixteen primer sets were designed based on the information obtained from public genome database (AC022545, AP000925, AP001005, http://www.ncbi.nlm.nih.gov/) to amplify all 10 exons, 5'-flanking regions, and introns 1, 4, 6, 7, and 9. Other introns were sequenced in the regions nearby exon/intron boundaries. After amplification, sequencing reaction was performed and analyzed on an ABI 310 genetic analyzer (PE Biosystems, Foster City, CA, USA). SNPs were detected by alignment of sequencing data of ten individuals. To obtain SNPs in the 5'-upstream region, 20 SNPs described in the publicly available SNPs database of TSC, The SNP Consortium Ltd (http://snp.cshl.org/index.html), were analyzed with Japanese 20 alleles.

SNPs genotyping

SNPs were genotyped with three PCR based methods; single strand conformational polymorphism (SSCP), SNaPshot®, and sequencing. PCR was performed in a volume of 12.5μl containing 20ng genomic DNA, 75mM Tris HCL (pH 8.8), 20mM $(NH_4)_2SO_4$, 0.01% Tween 20, 1.5mM MgCl2, 200μM deoxyribonucleotide triphosphate (dNTPs), 10pmol of each primer, and 0.25 units of Taq polymerase. PCR amplification was performed through 30 cycles of 94ºC for 30 sec, 49-62ºC for 30 sec, and 72ºC for 30sec, depending on the region analyzed, with a final extension step of 5min at 72ºC in a Gene Amp PCR9700 System (PE Biosystems, Foster City, CA, USA).

For SSCP analysis, each 2.5μl PCR products were mixed with formamide dye and heat denatured, and then applied to polyacrylamide gel electrophoresis. After electrophoresis, the gel was silver stained (Wako Pure Chemical Co Ltd., Tokyo, Japan) to detect polymorphic bands. SNaPshot® ddNTP primer extension method was performed following the manufacture's instruction (PE Biosystems, Foster City, CA, USA). In order to obtain purified template for SNaPshot® and sequencing, dNTPs and primers were removed from amplified genomic DNA by means of ExoSAP-IT (USB corp., Cleveland, Ohio, USA). In case of SNaPshot® method, reaction mixtures were electrophoresed on an ABI 310 genetic analyzer and data were analyzed with GeneScan® software (PE Biosystems, Foster City, CA, USA). For haplotype analysis, total 54 DNA samples (108 alleles) of healthy volunteers, recruited in our medical school, and were genotyped. All of them were medical students and gave their informed consent. The Ethics Committee of Asahikawa Medical College approved this study.

Haplotype analysis

Estimation of haplotype frequency was performed by the maximum-likelihood method using SNPAlyze® (DYNACOM, Yokohama, Japan, http://www.dynacom.co.jp/).

Results and discussion

Sequencing the CL-P1 gene region of ten Japanese volunteers (20 alleles) identified five SNPs, two in intron 2 (SNP2, SNP3), one in exon 5 (SNP4), and two in exon 6 (SNP5, SNP6) (Fig. 1, Table 1). Considering the importance in gene regulation, screening of 20 SNPs in the 5'-upstream region described in the SNP consortium Ltd database by sequencing was performed. One SNP out of 20 (SNP1) revealed to be useful for our purpose. After all, total six SNPs were utilized for genotyping and haplotype analysis. Genotype and allele frequencies of six SNPs and the location in the gene were shown in Table 1. All genotype frequencies conformed to Hardy-Weinberg equilibrium. Genotyping of 54 individuals (108 alleles) revealed that all SNPs are quite common, with the minor allele frequency to be at least 29%.

Out of six SNPs, three (SNP4 to 6) locate in the coding region, one silent mutation in exon 5 (SNP4: T267T) and two missense mutations in exon 6 (SNP5: S522P, SNP6: G606S) (Table 1). As shown in Fig. 1, SNP5 in exon 6 is in the collagen-like domain, that has characteristic Gly-X-Y repetitive pattern, where X and Y can be any amino acid but are frequently prolines or hydroxyprolines. SNP5, either serine or proline, locating in the X position, could have some effect in the stability of collagen structure. On the other hand, SNP6 in exon 6, either glycine or serine, locates in the neck domain. Hydrophobic amino acids in the neck domain contribute to form a triple alpha-helical coiled-coil. Each alpha-helix interacts with a neighboring CRD (Sheriff et al. 1994). C-terminal CRDs associated with the N-terminal collagen-like domain and recognize pathogens, so an amino acid alteration in the neck region might influence the association between them.

The result of linkage disequilibrium (LD) analysis was summarized in Table 2. Among the SNPs, only SNP4 and SNP5 showed significant LD (r^2=0.823). Haplotype frequencies among 108 alleles were calculated (Table 3). Top eight haplotypes whose frequencies over 5% accounted for 67% of all haplotypes.

CL-P1 is a member of the collectin family, which is considered to play significant roles in innate immunity. Classical collectins are soluble, but CL-P1 is membrane-bound. CL-P1 might bind and control not only bacteria and yeasts but also modified LDLs in the vascular space. The collagen-like domains in human and mouse CL-P1, which have the highest identity (96%) described to date, may play the most important role in these biological functions (Ohtani et al. 2001).

In conclusion, these six polymorphisms, their haplotype, and state of LD will be useful for investigation of a possible relationship between genetic variation at the human Cl-P1 locus and human diseases, especially in association studies between specific polymorphisms and susceptibility to innate immunity disease or atherosclerosis.

Acknowledgements

We are grateful to Katsura Nakanishi, Ayako Sado for their technical assistance. This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Medical Genome Science" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Drickamer, K (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins J Biol Chem 263: 9557-9560

Hartley CA, Jackson DC, and Anders EM (1992) Two distinct serum mannose-binding lectins function as beta inhibitors of influenza virus: identification of bovine serum beta inhibitor as conglutinin. J Virol 66: 4358-4363

Hartshorn KL, Crouch EC, White MR, Eggleton P, Tauber AI, Chang D, and Sastry K (1994) Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. J Clin Invest 94: 311-319

Kawasaki N, Kawasaki T, and Yamashina I (1989) A serum lectin (mannan-binding protein) has complement-dependent bactericidal activity. J Biochem (Tokyo) 106: 483-489

Kodama T, Freeman M, Rohrer L, Zabrecky J, Matsu-daira P, and Krieger M (1990) Type I macrophage scavenger receptor contains alpha-helical and collagen-like coiled coils. Nature 343: 531-535

Kuan SF, Rust K, and Crouch E (1992) Interactions of surfactant protein D with bacterial lipopolysaccharides Surfactant protein D is an Escherichia coli-binding protein in bronchoalveolar lavage. J Clin Invest 90: 97-106

Laursen SB, Dalgaard TS, Thiel S, Lim BL, Jensen TV, Juul-Madsen HR, Takahashi A, Hamana T, Kawakami M, and Jensenius JC (1998) Cloning and sequencing of a cDNA encoding chicken mannan-binding lectin (MBL) and comparison with mammalian analogues. Immunology 93: 421-430

Ohtani K, Suzuki Y, Eda S, Kawai T, Kase T, Yamazaki H, Shimada T, Keshi H, Sakai Y, Fukuoh A, Sakamoto T, Wakamiya N (1999) Molecular cloning of a novel human collectin from liver (CL-L1). J Biol Chem 274: 13681-13689

Ohtani K, Suzuki Y, Eda S, Kawai T, Kase T, Keshi H, Sakai Y, Fukuoh A, Sakamoto T, Itabe H, Suzutani T, Ogasawara M, Yoshida I, Wakamiya N (2001) The membrane-type collectin CL-P1 is a scavenger receptor on vascular endothelial cells. J Biol Chem 276: 44222-8

Persson A, Chang D, Rust K, Moxley M, Longmore W, and Crouch E (1989) Purification and biochemical characterization of CP4 (SP-D), a collagenous surfactant-associated protein. Biochemistry 28: 6361-636

Pikaar JC, Voorhout WF, van Golde LM, Verhoef J, Van Strijp JA, and van Iwaarden JF (1995) Opsonic activities of surfactant proteins A and D in phagocytosis of gram-negative bacteria by alveolar macrophages. J Infect Dis 172: 481-489

Schweinle JE, Ezekowitz RA, Tenner AJ, Kuhlman M, and Joiner KA (1989) Human mannose-binding protein activates the alternative complement pathway and enhances serum bactericidal activity on a mannose-rich isolate of Salmonella. J Clin Invest 84: 1821-1829

Sheriff S, Chang CY, Ezekowitz RA (1994) Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. Nat Struct Biol 1: 789-94

Shinohara Y, Ezura Y, Iwasaki H, Nakazawa I, Ishida R, Kodaira M, Kajita M, Shiba T, Emi M (2001) Linkage disequilibrium and haplotype analysis among ten single-nucleotide polymorphisms of interleukin 11 identified by sequencing of the gene. J Hum Genet 46: 494-7

Suzuki H, Kurihara Y, Takeya M, Kamada N, Kataoka M, Jishage K, Ueda O, Sakaguchi H, Higashi T, Suzuki T, Takashima Y, Kawabe Y, Cynshi O, Wada Y, Honda M, Kurihara H, Aburatani H, Doi T, Matsumoto A, Azuma S, Noda T, Toyoda Y, Itakura H, Yazaki Y, Horiuchi S, Takahashi K, Kruijt JK, Berkel TJC, Steinbrcher UP, Ishibashi S, Maeda N, Gordon S, Kodama T (1997) A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection. Nature 386: 292-296

Wakamiya N, Okuno Y, Sasao F, Ueda S, Yoshimatsu K, Naiki M, and Kurimura T (1992) Isolation and characterization of conglutinin as an influenza A virus inhibitor.

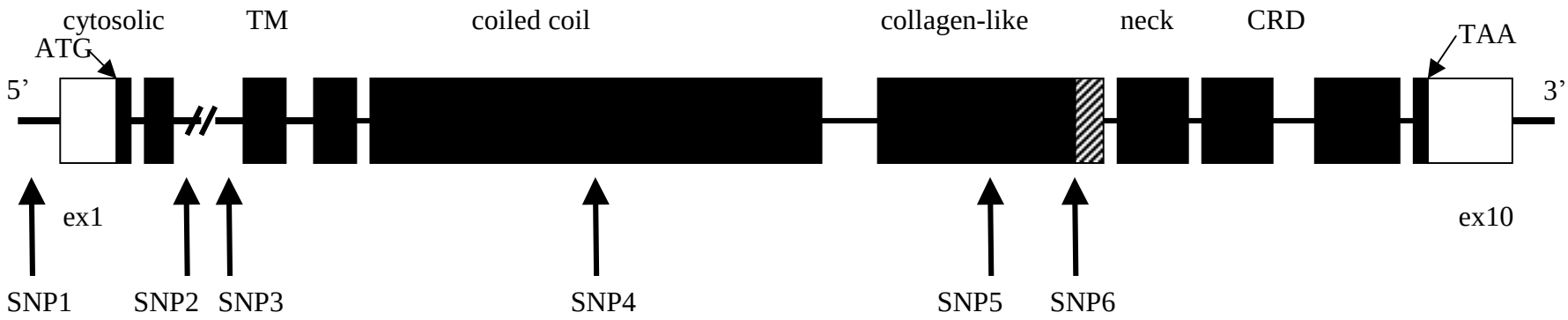
Biochem Biophys Res Commun 187: 1270-1278

White RT, Damm D, Miller, J, Spratt K, Schilling J, Hawgood S, Benson B, and Cordell, B (1985) Isolation and characterization of the human pulmonary surfactant apoprotein gene. Nature 317: 361-363

van Iwaarden JF, van Strijp JA, Visser H, Haagsman HP, Verhoef J, and van Golde LM (1992) Binding of surfactant protein A (SP-A) to herpes simplex virus type 1- infected cells is mediated by the carbohydrate moiety of SP-A. J Biol Chem 267: 25039-25043

Figure legends

Figure 1. Relationship between genomic structure and the location of single-nucleotide polymorphisms (SNPs) in the human CL-P1 gene. Vertical rectangles and horizontal lines represent exons and introns, respectively. Coding regions are shown with filled rectangles. Arrows indicate approximate position of each SNP. Functional domains are indicated above the gene. TM: transmembrane domain, CRD: carbohydrate recognition domain.



TM: transmenbrane domain
CRD: carbohydrate recognition domain
Black boxes indicate the coding region, and open boxes indicate 5’ untranslated region (UTR) and 3’ UTR.
Shaded portion represents a variant type, which lacks CRD.

Fig. 1. Relationship between genomic structure and the location of single-nucleotide polymorphisms (SNPs) in the human CL-P1 gene

Table 1. Genotype and allele frequencies of all six SNPs in the control subjects

Polymorphism	Nucleotide position	amino acid	Genotype			Allele		Allele frequency	
SNP1	-4078 [a]		TT	TG	GG	T	G	T	G
(5' flanking region)			22	24	8	68	40	0.63	0.37
SNP2	IVS2+814G>A [b]		GG	GA	AA	G	A	G	A
(intron 2)			21	24	9	66	42	0.61	0.39
SNP3	IVS2-1619A>G [b]		AA	AG	GG	A	G	A	G
(intron 2)			24	24	6	72	36	0.67	0.33
SNP4	801+1C>T [b]	T267T	CC	CT	TT	C	T	C	T
(exon 5)			28	21	5	77	31	0.71	0.29
SNP5	1564+1C>T [b]	S522P	CC	CT	TT	C	T	C	T
(exon 6)	IMS-JST062184		27	23	4	77	31	0.71	0.29
SNP6	1815+1A>G [b]	G606S	GG	GA	AA	G	A	G	A
(exon 6)	IMS-JST062186		23	27	4	73	35	0.68	0.32

a: Nucleotide position is identified from the translation initiation site.

b: The designation of other SNPs follows the recommendation of mutation nomenclature in the reference (Dunnen and Antonarakis 2000).

Table 2. Analysis of linkage disequilibrium for all possible two-way comparison among six SNPs

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
SNP1	D'	0.040	0.659	0.144	0.245	0.208
	r^2	0.001	0.119	0.005	0.014	0.012
	p	0.822	0.000	0.534	0.287	0.279
SNP2		D'	0.003	0.226	0.145	0.003
		r^2	0.000	0.032	0.013	0.000
		p	0.822	0.040	0.236	0.822
SNP3			D'	0.325	0.437	0.092
			r^2	0.092	0.167	0.008
			p	0.001	0.000	0.351
SNP4				D'	0.907	1.000
				r^2	0.823	0.195
				p	0.000	0.000
SNP5					D'	1.000
					r^2	0.195
					p	0.000

Table 3. Frequencies of haplotypes constructed from six SNPs

SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	
TG	AG	GA	TC	CT	GA	Frequency
G	G	A	C	C	G	0.155
T	G	G	T	T	G	0.114
T	G	A	C	C	G	0.091
T	G	A	C	C	A	0.074
T	G	G	C	C	A	0.060
T	A	A	C	C	G	0.059
G	A	A	T	T	G	0.053
T	A	G	C	C	A	0.053
G	A	A	C	C	A	0.049
G	G	A	C	C	A	0.046
T	A	A	C	C	A	0.038
T	A	A	T	T	G	0.037
T	G	A	T	T	G	0.032
G	A	A	C	C	G	0.026
T	A	G	T	T	G	0.021
T	G	G	C	T	G	0.019
G	A	G	T	T	G	0.016
T	G	G	C	C	G	0.013
G	A	G	C	C	G	0.012
T	A	G	C	C	G	0.012
G	A	A	T	C	G	0.009
T	A	A	T	C	G	0.009
					Total	1.000