Vol. 165, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1989 Pages 138-144

MOLECULAR CLONING OF THE cDNA CODING FOR PROLINE-RICH PROTEIN (PRP): IDENTITY OF PRP AS C4b-BINDING PROTEIN

Tetsuya MATSUGUCHI, Seiichi OKAMURA*, Teijiro ASO, Teizo SATA and Yoshiyuki NIHO

The First Department of Internal Medicine and Cancer Center, Faculty of Medicine, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812, Japan

Received October 11, 1989

Summary: Proline-rich protein (PRP) is a plasma protein with a high proportion of proline residues and possessing lipid-binding properties. In order to clarify its structure, a human liver cDNA library was screened using anti-PRP antiserum. Several overlapping phage cDNA clones were isolated and the total nucleotide sequence of the cDNA, 2178 bp in length, was analyzed. The amino acid composition of PRP deduced from the cDNA was essentially the same as that reported for PRP. In a homology search, the cDNA sequence was almost completely the same as the previously reported cDNA sequence of C4b-binding protein. Furthermore, the reported molecular weights of the two proteins under both reduced and unreduced conditions were quite alike. These findings indicate that PRP is identical with C4bp.

Proline-rich protein (PRP) is a glycoprotein that was first isolated from lipoprotein-free plasma by absorption to lecithin-stabilized triglyceride emulsion (Intralipid) (1). PRP, which is also present in chylomicron, has been determined to have a molecular weight (MW) of 74,000 by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions, but in native plasma it behaves as a larger particle (1). The mean plasma concentration of PRP was reported to be 21.7 mg/dl and is closely related to serum cholesterol of lipoprotein and serum triglyceride (1). Recently, it was shown that PRP is an acute-phase reactant, and that its level rises in parallel with erythrocyte sedimentation rate (ESR) and C-reactive protein (2). Although little is known about its structure and function, amino acid analysis of PRP has shown a large proline content (1). In this study, we isolated and analyzed the cDNA for PRP and revealed that it is identical with that for C4b-binding protein (C4bp) (3, 4). Our present report also extends the previously reported nucleotide sequence of C4bp cDNA.

*To whom correspondence should be addressed.

ABBREVIATIONS: PRP, proline-rich-protein; C4bp, C4b-binding protein.

0006-291X/89 $1.50

MATERIALS AND METHODS

PRP and antisera for PRP: Purification of PRP and production of its antiserum have been reported previously (1). Briefly, the dialyzed 1.21-density infranatant fraction of plasma was mixed with 0.2% Intralipid and incubated at 37°C for 1 h. After centrifugation, the supernatant lipid layer was subjected to chromatography on a 4% agarose gel column. The purified emulsion particles were extracted with ethanol-diethylether, and re-subjected to chromatography on a 4% agarose gel column. The eluate was further purified on an alkaline 3.75% polyacrylamide gel. The PRP band was cut out and emulsified with Freund's complete adjuvant and injected intradermally into New Zealand albino rabbits. The injection was repeated three times at intervals of 2 weeks and the animal was bled 10 days after the last injection. The raised antiserum was preliminarily shown to give a single precipitation line against whole human serum.

Isolation of cDNA clones: A human liver cDNA library was purchased from Clontech Lab. Inc. (Palo Alto, CA) and screened with the PRP antiserum according to the method of Huynh *et al.*(5). The insert DNA of the positive clone (λPRP2) was labeled with [^{32}P]CTP by random priming (6) and used as a probe to rescreen the same library by the nitrocellulose replica filter preparation technique (7). Filters were hybridized for 16 h at 42 °C in a buffer containing 50% formamide (8).

DNA Sequence Analysis: DNA sequencing was carried out by the chain-termination DNA sequencing method (9) with T7 DNA polymerase (Sequenase; United States Biochemical Corp., Cleveland, OH)(10).

Comparison of Nucleotide Sequences: The EMBL computer program was used for analysis of the nucleotide sequence data.

RESULTS

Approximately 40,000 phage clones from the human liver cDNA library were screened with rabbit anti-PRP antiserum and one positive cDNA clone was isolated and designated λPRP2 (Fig.1). Using the insert DNA of λPRP2 as a DNA probe, about 10,000 clones of the same library were rescreened and four overlapping additional cDNA clones (λPRP4, λPRP6, λPRP7, λPRP8) were isolated (Fig.1).

Nucleotide sequence analysis was performed on the five overlapping cDNA clones (Fig.1). The connected cDNA sequence, 2,178 bp in length, included all

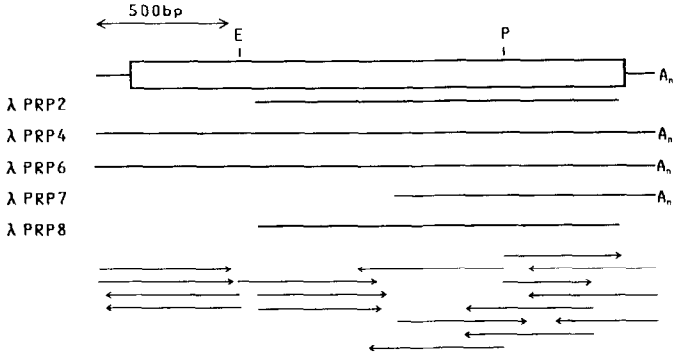

Fig. 1. Structure of human PRP cDNA. The coding region is represented by an open box. The isolated phage cDNA clones are represented by thick lines. The sequencing strategy is also shown. The restriction enzyme sites used in the preparation of the fragments for sequencing (E:*Eco*RI, P:*Pst*I) are indicated.

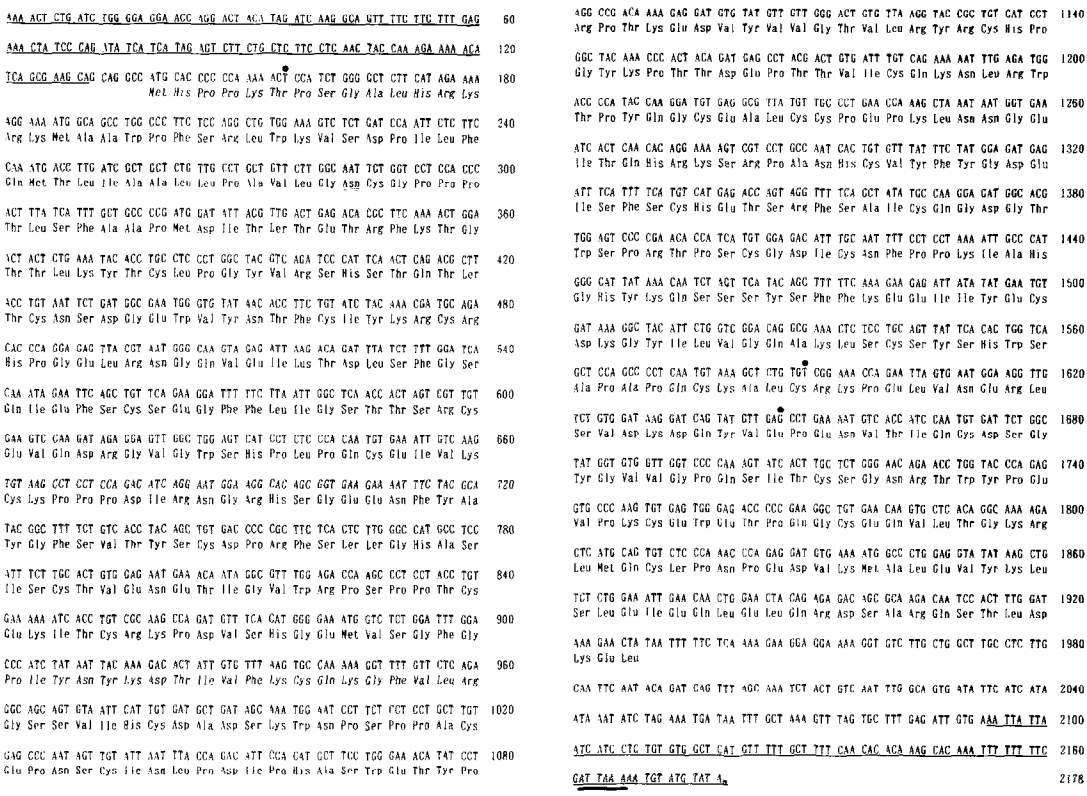

Fig. 2. The entire nucleotide sequence of PRP cDNA deduced from the sequences of the overlapping fragments shown in Fig.1. The sequences of the 5' and 3' untranslated regions which were not previously published for C4bp cDNA are underlined. The deduced amino acid sequence is shown below. The reported N-terminal amino acid (11) of C4bp is double-underlined. The nucleotides which were different from those in previous reports (3,4) are dotted. The used polyadenylation signal is thickly underlined.

of the 3' non-translated region as shown by the run of adenine residues at the 3' end of the cDNA (Fig.2). The position of the poly(A) tract was constant among three independent cDNA clones (λPRP4, λPRP6, λPRP7), suggesting no variation in the length of the 3' non-translated region. The longest open reading frame was for 597 amino acids (AAs) and the first ATG codon was 139 nucleotides (nt) from the 5' end of the sequence. The length of the 3'non-translated region was 246 nt. The putative polyadenylation signal was ATTAAA located at residues 2,162-2,167 or 17 nt upstream from the poly(A) tract. The same kind of one base mutation of the commonest polyadenylation signal (AATAAA) has been reported to be valid in various genes (11).

Figure 2 also shows the deduced amino acid sequence of the protein encoded in the longest open reading frame starting at the ATG at nucleotide 139. When the amino acid composition of the deduced protein was compared with the previously reported amino acid analyses of PRP (1,2), they were basically

Table 1. Comparison of amino acid (AA) composition. A: AA composition deduced from the isolated cDNA sequence of PRP. B: Composition excluding the leader peptides. C: AA composition of PRP reported by Sata *et al.*(1). D: AA composition of PRP reported by Funakoshi *et al.*(2). Tryptophan residue is excluded in A and B, as it could not be detected by the analysis method used in the previous reports (1, 2).

mol/10^3mol of amino acids	Samples			
	A	B	C	D
Lysine	58	56	59	53
Histidine	29	28	27	27
Arginine	53	50	51	47
Aspartic acid	75	80	90	88
Threonine	70	72	72	69
Serine	87	89	91	98
Glutamic acid	106	113	127	122
Proline	89	85	89	86
Half-cystine	62	67	34	55
Glycine	70	72	87	86
Alanine	36	32	41	38
Valine	56	58	46	42
Methionine	12	7	6	10
Isoleucine	50	50	40	36
Leucine	67	59	64	62
Tyrosine	43	46	38	41
Phenylalanine	36	35	38	41

the same, although a signal peptide appeared to be included in our present data (Table 1).

When the isolated PRP cDNA sequence was compared with sequences in the EMBL Data Library, it shared almost complete homology with the cDNA sequence for C4b-binding protein (C4bp) (3,4). There were only three differences between our nucleotide sequence and the reported C4bp cDNA sequence, a T for C substitution at nucleotide 1,590, a G for A substitution at 1,644, and a T for TT substitution at 156. The first two substitutions do not change the coded amino acids, because they involve the third nucleotide of the codons. The last substitution of T for TT, which was revealed by repeated sequencing of this region in both directions using several different fragments (Fig.1), changes the frame of the leader peptide making the ATG codon at nucleotide 139 a candidate for the first methionine.

Our newly isolated cDNA sequence is longer than the previously reported C4bp cDNA sequences (3, 4). It adds 131 nucleotides to the 5'-end, and includes 86 more nucleotides of the 3'-untranslated region extending to the poly(A) tract. The N-terminal amino acid of the C4bp structural protein is reported to be asparagine, which is the 49th amino acid of our deduced polypeptide (12). This indicates that the leader peptide is as long as 48 amino acids. The amino acid composition of C4bp excluding its leader peptide was calculated (Table 1, column B), and found to be also very close to that already reported for PRP (1, 2).

DISCUSSION

PRP was first isolated in 1976 as a plasma protein showing affinity for lipid and containing large amount of proline (1). In spite of its appreciable concentration in plasma, little is known about its structure and function. In order to clarify the structure, we attempted to clone its cDNA. As the liver had been shown immunohistochemically to produce PRP (2), we screened a liver cDNA library to isolate PRP cDNA clones. The putative PRP cDNA was obtained using a λgt11 expression vector system and a specific antiserum for PRP. In addition to immunological detection, the amino acid composition deduced from the cDNA was basically the same as that of PRP reported previously (Table 1). Furthermore, the MW of the protein estimated from the cDNA sequence was about the same as that of PRP. From these data we concluded that the isolated cDNA was for PRP.

On sequence comparison it was almost the same as the reported cDNA sequence for C4bp, indicating that PRP and C4bp are the same protein, but purified by different methods (1,13). In addition to their immunological relatedness, the reported molecular weights (MWs) of PRP and C4bp are quite alike. In previous reports (1,2), PRP showed a MW of 74,000 by SDS-PAGE but a much heavier MW of >1,000,000 (1) or 352,000 (2) by polyacrylamide gel electrophoresis (PAGE) under non-reduced conditions without (1) or with (2) urea. From these data, PRP was thought to exist in a polymerized form in plasma. On the other hand, C4bp was reported to have a MW of about 70,000 and 550,000 by PAGE under reduced and non-reduced conditions, respectively, and was thought to exist as a heptamer (14). As it is difficult to determine the exact MW of high-molecular-weight proteins by PAGE, the two proteins seem to have very similar MWs as a monomer and both exist in polymerized form in plasma. Furthermore, the mean concentrations of these two proteins in plasma were reported to be quite similar, 21.7 mg/dl for PRP (1) and about 20 mg/dl for C4bp (15). All these findings suggest that PRP is identical with C4bp.

The newly isolated cDNA was different from the previously reported C4bp cDNA sequence (3,4) at three nucleotides. The T for C substitution at nucleotide 1,590 and the G for A substitution at 1,644 do not change the coded amino acids because they involve the third nucleotide of each codon (Fig.2) and we considered that these are polymorphic sites. The same kinds of polymorphism have also been reported for C4bp cDNA previously (3). On the other hand, a T for TT substitution at nucleotide 156 will change the reading frame of the reported leader peptide (4) and make the ATG at 139 a candidate for the initiation codon. Though there are two other universal translation initiation codons 3'-downstream from this codon (at nucleotides 187 and 244), the ATG at 139 is most likely to be the point of initiation because it has

been reported that translation begins at the 5'-proximal AUG triplet in about 90% of eukaryotic mRNAs (16).

C4bp was first described in 1978 as a regulator of the classical pathway of the complement system. C4bp contains seven binding sites for C4b (14). It accelerates the dissociation of C2a from the C3 convertase (C4bC2a, EC 3.4.21.43) and also functions as a cofactor for enzyme factor I in the degradation of C4b (17, 18). C4bp also has a single binding site for the vitamin K-dependent protein S (19) and was recently demonstrated to be one of the regulators of the coagulation system (20, 21).

The identification of PRP and C4bp indicates that C4bp has a lipid binding capacity and that at least a proportion of it is present in chylomicron. C4bp, which is an important regulator of the complement system as well as of the coagulation system, may also be related to the metabolism of cholesterol, because the concentration of PRP is significantly correlated with that of cholesterol in lipoprotein of very low (VLDL) or low (LDL) density (1).

The PRP content of serum was reported to be related to the levels of CRP and ESR with or without inflammation or infection (2). Following acute myocardial infarction, the PRP level increases, and reaches a maximum on day 7, subsequently decreasing gradually to a normal level by day 20, just like the course of the CRP level (2). These findings also indicate that C4bp, which is the same as PRP, is one of the acute-phase reactants. The increase of C4bp during inflammation seems to be rational, as C4bp decreases C3 convertase activity and may limit the spread of inflammation.

Acknowledgments: We are grateful to Drs. M. Yamamoto and M. Furuichi, Medical Institute of Bioregulation of our University, for technical advice. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan (Nos. 60770949, 61015074, 61440054, 63015063) and from the Fukuoka Anti-cancer Society.

REFERENCES

1. Sata,T., Havel,R.J., Kotite,L., and Kane,J.P. (1976) Proc.Natl.Acad.Sci. USA. 73, 1063-1067.
2. Funakoshi,M., Sasaki,J., and Arakawa,K. (1988) Biophys.Biochim.Acta 963, 98-108.
3. Chung,L.P., Bentley,D.R., and Reid,K.B.M. (1985) Biochem.J. 230, 133-141.
4. Lintin,S.J., Lewin,A.R., and Reid,K.B.M. (1988) FEBS Lett. 232, 328-332.
5. Huynh,T.V., Young,R.A., and Davis,R.W. (1985) In DNA Cloning (D.M.Glover, Ed.), Vol.I, pp.49-78. IRI Press, Oxford.
6. Feinberg,A.P., and Vogelstein,B. (1983) Analyt.Biochem. 132, 6-13.
7. Grosveld,F.G., Dahl,H.H.M., de Boer,R., and Flavell, R.A. (1981) Gene 13, 227-237.
8. Bernards,R., and Flavell,R.A. (1980) Nucleic Acids Res. 8, 1521-1534.
9. Sanger,F., Miklen,S. and Coulson,A.R. (1977) Proc.Natl.Acad.Sci.USA. 74, 5463-5467.
10. Tabor,S., and Richardson,C.C. (1987) Proc.Natl.Acad.Sci.USA. 84, 4767-4771.

11. Leff,S.E., Rosenfeld,M.G., and Evans,R.M. (1986) Ann.Rev.Biochem. 55, 1091-1117.
12. Chung,L.P., Gagnon,J., and Reid,K.B.M. (1985) Mol.Immunol. 22, 427-435.
13. Scharfstein,J., Ferreira,A., Gigli,I., and Nussenzweig.V. (1978) J.Exp. Med. 148, 207-222.
14. Dahlback,B., Smith,C.A., and Muller-Eberhard,H.J. (1983) Proc.Natl.Acad. Sci. USA. 80, 3461-3465.
15. Dahlback,B. (1983) Biochem.J. 209, 847-856.
16. Kozak,M. (1978) Cell 15, 1109-1123.
17. Fujita,T., Gigli,I., and Nussenzweig,V. (1978) J.Exp.Med. 148, 1044-1051.
18. Fujita,T., and Nussenzweig,V. (1979) J.Exp.Med. 150, 267-276.
19. Dahlback,B., and Stenflo,J. (1981) Proc.Natl.Acad.Sci.USA. 78, 2512-2516.
20. Comp,P.C., Nixon,R.R., Cooper,M.R., and Esmon,C.T. (1984) J.Clin.Invest. 74, 2082-2086.
21. Dahlback,B. (1986) J.Biol.Chem. 261, 12022-12027.