Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4529–4532, June 1990
Genetics

Human genes for the α and β chains of complement C4b-binding protein are closely linked in a head-to-tail arrangement

(regulator of complement activation gene cluster/physical linkage)

FERNANDO PARDO-MANUEL*, JAVIER REY-CAMPOS*, ANDREAS HILLARP†, BJÖRN DAHLBÄCK†, AND SANTIAGO RODRIGUEZ DE CORDOBA*

*Unidad de Immunologia, Centro de Investigaciones Biologicas (Consejo Superior de Investigaciones Cientificas), Velazquez 144, 28006-Madrid, Spain; and †Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmo, Sweden

Communicated by Hans J. Müller-Eberhard, March 21, 1990 (received for review February 26, 1990)

ABSTRACT **C4b-binding protein (C4BP) is an important component in the regulation of the complement system and also binds the anticoagulant vitamin K-dependent protein S. These activities are performed by distinct, although structurally related, polypeptides of 70 kDa (α chain) and 45 kDa (β chain), respectively. In this report we have investigated the genetic relationships between these polypeptides. Using pulsed field gel electrophoresis analysis we demonstrate that the genes coding for the α (*C4BPα*) and β (*C4BPβ*) chains are closely linked within the regulator of complement activation gene cluster. In addition, we have determined that the 3′ end of the *C4BPβ* gene lies 3.5–5 kilobases from the 5′ end of the *C4BPα* gene. These findings support the concept that the *C4BPα* and *C4BPβ* genes are the result of a gene duplication event.**

C4b-binding protein (C4BP) is an abundant oligomeric plasma protein with a characteristic molecular heterogeneity. The major molecular form of C4BP is composed of eight chains, seven identical 70-kDa polypeptides (α chain) and one 45-kDa polypeptide (β chain), that are covalently linked by their C-terminal regions to give the molecule a spider-like structure (1–3). Alternative forms of C4BP include molecules lacking the β chain and heptamers of six α chains and one β chain (4).

C4BP is a regulator of complement activation. It binds to C4b, accelerates the decay of the classical pathway C3-convertase (C4b, 2a), and functions as a cofactor in the factor I-mediated proteolytic inactivation of C4b (1, 5–8). The complement regulatory functions of C4BP involve the α chain, whereas the β chain appears to be the binding site for the anticoagulant vitamin K-dependent protein S (2, 4).

The human genes coding for the α (*C4BPα*) and β (*C4BPβ*) polypeptide chains of C4BP have been cloned and their complete deduced amino acid sequences are known (9, 10). They both belong to a family of proteins with a characteristic structural organization based on a repetitive unit of 60 amino acids designated short consensus repeat (SCR). Because most of the proteins containing SCRs bind to C3b and/or C4b, this family is referred to as the superfamily of the C3b/C4b binding proteins. The typical SCR framework of conserved residues includes four cysteines, two prolines, one tryptophan, and several other partially conserved glycine and hydrophobic residues (11). Starting at their N-terminal ends the α and the β polypeptide chains of C4BP are composed of eight and three SCRs, respectively (9, 10). Both sequences end with a C-terminal nonrepeat region containing two cysteines that are thought to be involved in the interchain disulfide linkage.

The *C4BPα* gene has been physically linked to the genes coding for the C3b/C4b binding proteins MCP, CR1, CR2, DAF, and H (12–15). All of these proteins are complement-regulatory components. This group of genes, known as the regulator of complement activation (RCA) gene cluster (12), is located on the long arm of human chromosome 1 (1q32) (16, 17).

In this report we have investigated the mapping position of the *C4BPβ* gene. Our results demonstrate that the *C4BPβ* gene is located within the RCA gene cluster, where it lies adjacent to the 5′ end of the *C4BPα* gene.

MATERIALS AND METHODS

Preparation of Genomic DNA and Restriction Enzyme Digestion. Human genomic DNA was prepared from an Epstein–Barr virus-transformed B-cell line as described (18). Briefly, cells were collected by pelleting, washed twice with phosphate-buffered saline (PBS: 20 mM sodium phosphate, pH 7.2/150 mM NaCl), and resuspended at 30×10^6 cells per ml in PBS. An equal volume of melted low-gelling-temperature agarose (SeaPlaque; FMC), 1.6% in PBS at 50°C, was mixed with the cell suspension and immediately poured into ice-cooled 20 × 6 × 10 mm molds. When set, the agarose blocks were cut into 1-mm-thick slices that were incubated in 0.5 M EDTA, pH 8.0/1% SDS with 2 mg of proteinase K per ml at 55°C for 2 days. The agarose slices were washed twice in TE (10 mM Tris·HCl, pH 8.0/1 mM EDTA) at room temperature for 30 min, twice in TE with 1 mM phenylmethylsulfonyl fluoride (PMSF) at 55°C for 30 min, and then two more times with TE at room temperature. Plugs of 6 × 3 × 1 mm (containing ≈1.5 μg of DNA) were cut from the slices and subjected to restriction enzyme digestions. These were done in a 100-μl volume for 4–8 hr with 12–30 units of enzyme per digestion. In the double digestion experiments, when enzymes were not buffer compatible, the agarose plugs were washed twice with 1 ml of H_2O before the second enzyme digestion.

Pulsed Field Gel Electrophoresis (PFGE) Analysis. PFGE was done at 14°C in 13 × 13 cm 1.5% agarose gels using a 120° angle alternating homogeneous electrical field apparatus constructed according to Chu *et al.* (19) and a Pulsewave 760 electrophoretic pulse switcher (Bio-Rad). The gels were run in 0.25× TBE (1× TBE: 89 mM Tris base/89 mM boric acid/2 mM EDTA) at 180 V (voltage gradient of 6 V/cm) for a total of 40 hr with pulse lengths of 90 s for the first 20 hr and 45 s during the last 20 hr. *Saccharomyces cerevisiae* (strain 344-12A) chromosomes and λ DNA concatemers were used as molecular weight markers.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: C4BP, C4b-binding protein; RCA, regulator of complement activation; PFGE, pulsed field gel electrophoresis; SCR, short consensus repeat; PMSF, phenylmethylsulfonyl fluoride.

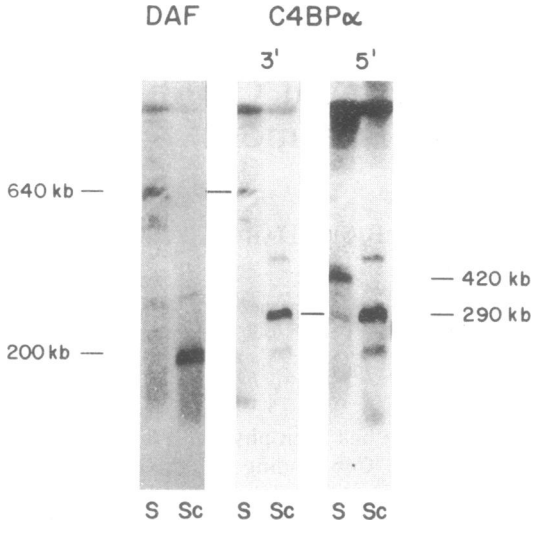

FIG. 1. Pulsed field gel analysis of human genomic DNA using DAF-, 3'-C4BPα-, and 5'-C4BPα-specific cDNA probes. Sizes of the restriction fragments hybridizing with each probe are indicated in kb. S, *Sal* I; Sc, *Sac* II.

Transfer and Hybridization. The gels were stained with 0.5 μg of ethidium bromide per ml after running and then UV irradiated for 10 s on a 254-nm wavelength UV-transilluminator before alkaline transfer to Biotrace-RP nylon membranes (Gelman). The membranes were prehybridized in 5× SSC (1× SSC: 15 mM sodium citrate, pH 7/150 mM NaCl), 1% SDS, 2% commercial skim milk, and 200 μg of denatured salmon testes DNA per ml at 68°C for 3–8 hr and then hybridized for 24–40 hr under the same conditions with probes ^{32}P-labeled by the random oligopriming method (20). After high stringency washes (last wash in 0.2× SSC/1% SDS at 68°C) the membranes were exposed for 1–3 days with Kodak RX-Omat films at −70°C. Before reprobing, the membranes were stripped by washing in 0.4 M NaOH at 45°C for 30 min and then neutralized in 0.2 M Tris·HCl, pH 7.5/0.1× SSC/1% SDS at 45°C for 30 min.

DNA Probes. The human probe for the β chain of C4BP (C4BPβ) has been described previously (10). It is 886-base-pair (bp)-long cDNA and contains a single *Bam*HI restriction site that upon digestion generates two fragments of 360 bp and 526 bp corresponding to the 3' and 5' ends of the C4BPβ cDNA, respectively. The human probe for the α chain of C4BP (C4BPα) is a 2190-bp-long cDNA clone. This clone contains a single internal *Eco*RI restriction site located 547 bp from the 5' end. This 547-bp fragment will be referred to as the 5'-C4BPα probe. Similarly, a *Kpn* I restriction site located 458 bp from the 3' end was used to generate the 3'-C4BPα probe. The human probe for DAF has been described previously (21) and was a generous gift of I. Caras (Genentech).

Genomic Clones. A group of overlapping genomic clones spanning the 5' end of the *C4BPα* gene was used in hybridization experiments with the C4BPβ probes. They were originally isolated from a human leukocyte genomic library constructed in the EMBL3 vector (Clontech; HL 1006d) using the 2190-bp-long human C4BPα cDNA probe and will be described in detail elsewhere.

RESULTS

Previous work on the organization of the human RCA gene cluster (14) has revealed that (*i*) the *CR1*, *CR2*, *DAF*, and *C4BPα* genes are included, in this order, in a 640-kilobase (kb) *Sal* I fragment; (*ii*) the *DAF* and *C4BPα* genes lie within the same 800-kb *Mlu* I and 475-kb *Sfi* I fragments, which are different from those containing the *CR1* and the *CR2* genes; and (*iii*) the *DAF* and *C4BPα* genes can be separated into different *Sac* II fragments.

The *C4BPα* gene can be 3' → 5' oriented in relation to the *DAF* gene based on the existence of an internal *Sal* I restriction site. As shown in Fig. 1, cDNA probes specific for the 3' (3'-C4BPα) and 5' (5'-C4BPα) ends of *C4BPα* (see *Materials and Methods* for a description of the different probes) hybridize with two different *Sal* I fragments of 640 kb and 420 kb, respectively. In agreement with previous experiments (14, 22), the DAF-cDNA probe hybridizes with the 640-kb *Sal* I fragment but not with that of 420 kb. Both 5'-C4BPα and 3'-C4BPα probes hybridize with the same *Sac* II fragment of 290 kb, which, as expected, is different from that of 200 kb hybridizing with the DAF-cDNA probe. From these results we conclude that the *C4BPα* gene is oriented with the 3' end toward the *DAF* gene.

Fig. 2 shows that the non-crosshybridizing cDNA probes

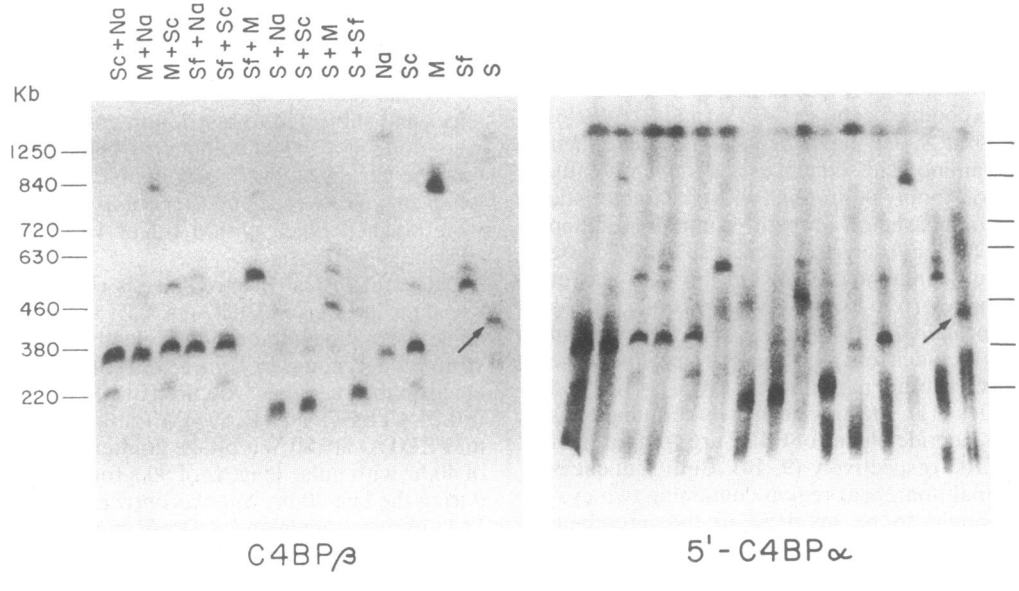

FIG. 2. Pulsed field gel analysis of human genomic DNA using C4BPβ- and 5'-C4BPα-specific cDNA probes. Sizes of some representative yeast chromosomes are indicated in kb. Arrows show the 420-kb *Sal* I fragment hybridizing with both probes. S, *Sal* I; Sc, *Sac* II; Sf, *Sfi* I; M, *Mlu* I; Na, *Nae* I.

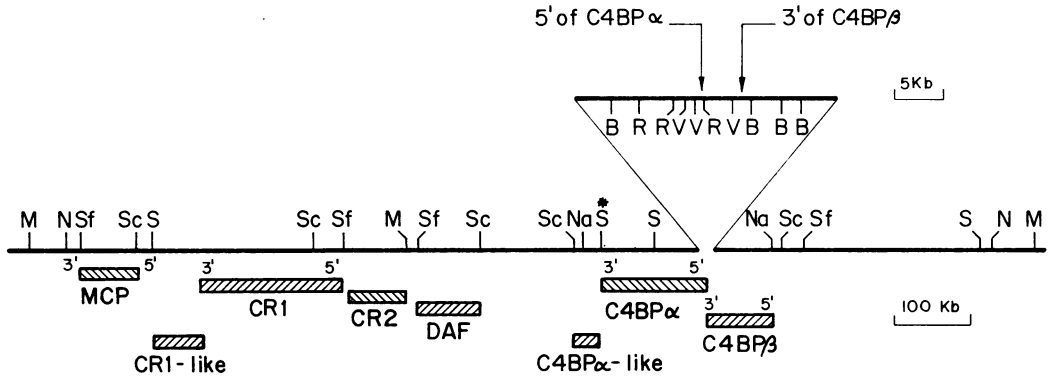

FIG. 3. Restriction map of the human RCA gene cluster showing the position of the *C4BPβ* gene. The bars for the *CR1*-like and *C4BP*-like genes are only to indicate that they have been mapped to the 3' end of the *CR1* and *C4BPα* genes, respectively (23, 24). S, *Sal* I; Sc, *Sac* II; Sf, *Sfi* I; M, *Mlu* I; N, *Not* I; Na, *Nae* I; R, *Eco*RI; B, *Bam*HI; V, *Eco*RV. The asterisk labels a "hyposensitive" *Sal* I restriction site.

5'-C4BPα and C4BPβ hybridize with identical genomic fragments generated by the digestion with *Sfi* I, *Mlu* I, *Sac* II, *Nae* I, or *Sal* I. The smallest of these restriction fragments is 180 kb long and results from the combined digestion with *Sal* I and *Nae* I. The 800-kb *Mlu* I and the 475-kb *Sfi* I fragments hybridize also with a DAF-cDNA probe (not shown). Most informative in this experiment is the finding that the 420-kb *Sal* I fragment that hybridizes with 5'-C4BPα also hybridizes with the C4BPβ probe. These results demonstrate that the *C4BPβ* gene lies in the vicinity of the 5' end of the *C4BPα* gene.

Since all restriction enzymes used in the PFGE analysis described above gave identical hybridization patterns using the 5'-C4BPα or the C4BPβ probe, we postulate that these two genes are adjacent. To test this possibility, a group of overlapping genomic clones spanning the 5' end of the *C4BPα* gene was analyzed for the presence of *C4BPβ* sequences. These clones were originally isolated from a human genomic library constructed in the EMBL3 vector (Clontech) using the human C4BPα cDNA probe (see *Materials and Methods*). One of these clones (G562), with a 21-kb insert containing the first (5') exon of the *C4BPα* gene (unpublished data), hybridizes with the C4BPβ probe. Detailed analysis of clone G562 by restriction mapping and hybridization with probes specific for the 3' and 5' ends of the *C4BPβ* gene demonstrated that the 3' end of *C4BPβ* lies in a 1.5-kb *Eco*RV/*Bam*HI fragment that is separated by a segment of 3.5 kb from the 450-bp *Eco*RV/*Eco*RI fragment containing the 5' end of *C4BPα* gene. Thus, the *C4BPβ* gene is oriented with the 3' end next to the 5' end of the *C4BPα* gene. Both genes have, therefore, the same 5' → 3' orientation, which is also the same as reported for the human *MCP* and *CR1* genes (22, 23). These experiments also provide an estimate of the distance between the *C4BPα* and the *C4BPβ* genes between 3.5 kb and 5 kb. Fig. 3 summarizes these data and provides an up-to-date diagram of our current understanding of the organization of this region of the human RCA gene cluster.

DISCUSSION

This report demonstrates that the genes coding for the α and β chains of human C4BP are closely linked within the RCA gene cluster. The 3' end of the *C4BPβ* gene has been located 3.5–5 kb from the 5' end of the *C4BPα* gene (Fig. 3).

C4BP is unique among RCA proteins in that it is organized by functionally distinct polypeptides that arrange a peculiar spider-like quaternary structure (3). Both α and β polypeptides belong to the superfamily of C3b/C4b binding proteins and, except for the C-terminal region, their sequences are organized by SCRs (9, 10). The non-SCR C-terminal region is important in organizing the quaternary structure. This region is similar in the α and β chains (10), probably reflecting a closer evolutionary relationship between these two polypeptides than that already inferred by the presence of SCRs. The fact that they are encoded by contiguous genes indicates that the *C4BPα* and *C4BPβ* genes are most likely the result of a gene duplication event.

The hypothesis that *C4BPα* and *C4BPβ* originated by gene duplication is consistent with the current view concerning the generation and evolution of the RCA gene cluster. With the observations reported here, an interesting picture that delineates certain regions of the RCA gene cluster as active areas of gene duplication is emerging. One of these areas is the "C4BP-RCA subregion" where, in addition to the *C4BPα* and *C4BPβ* genes, a *C4BPα*-like gene has been mapped adjacent to the 3' end of the *C4BPα* gene (24). Another of these regions occurs at the "CR1-RCA subregion" including the *CR1*-like, *CR1*, and *CR2* genes (25–27). The ancestors of the *CR1* and *CR2* genes most likely originated before man and mouse diverged (28). When the *C4BPα* and *C4BPβ* genes appeared is currently unknown.

Interestingly, the human and mouse C4BP α chains present distinct non-SCR C-terminal regions (9, 29). The major difference is that the murine C4BPα polypeptides lack the two cysteinyl residues involved in the disulfide linkage of the human C4BP chains. To determine whether there is a murine homolog of the human *C4BPβ* and to characterize its C-terminal region should, therefore, be informative in relation to the evolution of the *C4BPα* and *C4BPβ* genes.

Finally, a possible consequence of the close linkage between *C4BPα* and *C4BPβ* is that they share regulatory sequences, as it has already been discussed for the closely linked genes coding for the human complement components C2 and factor B (30). This hypothesis is particularly attractive in the case of *C4BPα* and *C4BPβ* because they code for subunits of the same molecule.

We thank Drs. P. Sanchez-Corral and M. A. Penalva for carefully reading and criticizing the manuscript. This work was partially supported by a grant from the Spanish government (Direccion General de Investigacion Cientifica y Tecnica, PM88-0002) and a grant from the Swedish Medical Research Council (B89-13X-07143-05C).

1. Scharfstein, J., Ferreira, A., Gigli, I. & Nussenzweig, V. (1978) *J. Exp. Med.* **148**, 207–222.
2. Hillarp, A. & Dahlback, B. (1988) *J. Biol. Chem.* **263**, 12759–12764.
3. Dahlback, B., Smith, C. A. & Muller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3461–3465.
4. Hillarp, A., Hessing, M. & Dahlback, B. (1989) *FEBS Lett.* **259**, 53–56.

5. Fujita, T., Gigli, I. & Nussenzweig, V. (1978) *J. Exp. Med.* **148,** 1044–1051.
6. Fujita, T. & Nussenzweig, V. (1979) *J. Exp. Med.* **150,** 267–276.
7. Nagasawa, S. & Stroud, R. M. (1980) *Mol. Immunol.* **17,** 1365–1372.
8. Gigli, I., Fujita, T. & Nussenzweig, V. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6596–6600.
9. Chung, L. P., Bentley, D. R. & Reid, K. B. M. (1985) *Biochem. J.* **230,** 133–141.
10. Hillarp, A. & Dahlback, B. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1183–1187.
11. Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) *Immunol. Today* **7,** 230–234.
12. Rodriguez de Cordoba, S., Lublin, D., Rubinstein, P. & Atkinson, J. P. (1985) *J. Exp. Med.* **161,** 1189–1195.
13. Rey-Campos, J., Rubinstein, P. & Rodriguez de Cordoba, S. (1987) *J. Exp. Med.* **166,** 246–252.
14. Rey-Campos, J., Rubinstein, P. & Rodriguez de Cordoba, S. (1988) *J. Exp. Med.* **167,** 664–669.
15. Bora, N. S., Lublin, D. L., Kumar, B. V., Hockett, R. D., Holers, V. M. & Atkinson, J. P. (1989) *J. Exp. Med.* **169,** 597–602.
16. Weiss, J. H., Morton, C. C., Bruns, G. A., Weiss, J. J., Klickstein, L. B., Wong, W. W. & Fearon, D. T. (1987) *J. Immunol.* **138,** 312–319.
17. Lublin, D., Lemons, R. S., Le Beau, M. M., Holers, V. M., Tykocinski, M. L., Medof, M. E. & Atkinson, J. P. (1987) *J. Exp. Med.* **165,** 1731–1736.
18. Kenwrick, S., Patterson, M., Speer, A., Fischbeck K. & Davies, K. (1987) *Cell* **48,** 351–357.
19. Chu, G., Vollrath, D. & Davis, R. W. (1986) *Science* **234,** 1582–1584.
20. Feinberg, A. I. & Volgestein, B. (1983) *Anal. Biochem.* **132,** 6–13.
21. Caras, I. W., Davitz, M. A., Rhee, L., Weddell, G., Martin, D. W., Jr., & Nussenzweig, V. (1987) *Nature (London)* **325,** 545–547.
22. Carroll, M. C., Alicot, E. M., Katzman, P. J., Klickstein, L. B., Smith, J. A. & Fearon, D. T. (1988) *J. Exp. Med.* **167,** 1271–1280.
23. Hourcade, D., Bora, N., Holers, V. M. & Atkinson, J. P. (1989) *Complement Inflammation* **6,** 348.
24. Rey-Campos, J., Marshall, P., Rubinstein, P. & Rodriguez de Cordoba, S. (1989) *Complement Inflammation* **6,** 393–394.
25. Wong, W. W., Cahill, J. M., Rosen, M. D., Kennedy, C., Bonaccio, E. T., Morris, M. J., Wilson, J. G., Klickstein, L. B. & Fearon, D. T. (1989) *J. Exp. Med.* **169,** 847–863.
26. Hourcade, D., Miesner, D. R., Atkinson, J. P. & Holers, V. M. (1988) *J. Exp. Med.* **168,** 1255–1270.
27. Weiss, J. J., Toothaker, L. E., Smith, J. A., Weiss, J. H. & Fearon, D. T. (1988) *J. Exp. Med.* **167,** 1047–1066.
28. Kurtz, C. B., Paul, M. S., Aegerter, M., Weiss, J. J. & Weiss, J. H. (1989) *J. Immunol.* **143,** 2058–2067.
29. Kristensen, T., Ogata, R. T., Chung, L. P., Reid, K. B. M. & Tack, B. F. (1987) *Biochemistry* **26,** 4668–4674.
30. Wu, L.-C., Morley, B. J. & Campbell, R. D. (1987) *Cell* **48,** 331–342.