Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Sequence Heterogeneity of Murine Complementary DNA Clones Related to the C4 and C4-Slp Isoforms of the Fourth Complement Component

M. Tosi, M. Levi-Strauss, Christiane Duponchel and T. Meo

Phil. Trans. R. Soc. Lond. B 1984 **306**, 389-394
doi: 10.1098/rstb.1984.0099

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1984 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Phil. Trans. R. Soc. Lond. B **306**, 389–394 (1984)
Printed in Great Britain

Sequence heterogeneity of murine complementary DNA clones related to the C4 and C4-Slp isoforms of the fourth complement component

By M. Tosi, M. Lévi-Strauss, Christiane Duponchel and T. Meo
Unité d'Immunogénétique, Institut Pasteur, 25 *rue du Docteur Roux*, 75724 *Paris Cédex* 15, *France*

Two classes of mRNA encoding the murine C4 protein were identified by sequence analysis of clones isolated from a liver complementary DNA library. The divergence found within a 357 base pair sequence available for comparison is limited to five nucleotide replacements located in the region corresponding to the carboxy-terminal end of the C4d peptide fragment. One of the nucleotide substitutions influences the presence of a site for the Hind III restriction endonuclease. That this restriction site indeed discriminates the two non-allelic genes encoding the mouse C4 and C4-Slp isoforms has been demonstrated by Southern blot analysis and nucleotide sequencing at the genomic level. Circumstantial evidence supports the identification of the gene lacking the Hind III site in the region corresponding to the carboxy-terminal end of the C4d fragment as the one encoding the C4-Slp isotype.

Introduction

The fourth component of complement is present in the serum of individual mice of certain strains in two serologically distinguishable forms: C4 and C4-Slp (sex limited protein). The expression of the two proteins is controlled differently. In contrast to C4, the synthesis of C4-Slp shows a marked dependence on testosterone induction and is characteristically absent in the males of some inbred strains (Shreffler *et al.* 1981). Generally, the structural and quantitative markers of C4 and C4-Slp proteins map to and define the *S* region of the H-2 major histocompatibility complex. The inheritance of such markers is best explained by postulating two non-allelic structural genes encoding two isotypic forms of C4. Several types of expression variants have been described for both proteins (Atkinson *et al.* 1982). To analyse the molecular mechanisms that characterize these genetic variants we sought to isolate molecular clones corresponding to the messengers encoding each of the two C4 proteins. To this end, we constructed in the plasmid vector pBR322 a complementary DNA library from a size enriched liver mRNA preparation from male mice of the strain B10.W7R. This H-2 congenic strain synthesizes both isoforms of C4 but expresses C4-Slp in large amounts and in a testosterone-independent mode (Hansen & Shreffler 1976).

Results and discussion

Twenty C4 clones have been isolated by screening 5000–10000 colonies using as hybridization probes first the insertion of the human C4 cDNA clone Alu/7 described by Carroll & Porter (1982), and then DNA fragments prepared from the mouse C4 clones isolated in the first round of screening. Chakravarti *et al.* (1983) showed that the sequence heterogeneity found in a pool of human C4 proteins resides at least in part in the carboxy-terminal half of the α chain. Assuming that these amino acid replacements identify stretches of the protein less subject to

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

functional constraints, we focused our attention on the corresponding portion of the mouse C4 clones with the hope of finding sequence markers that would be characteristic of each C4 isotype. Moreover, the quantitative difference described by Karp (1983) in the N-linked carbohydrates of the carboxy-terminal half of the C4 and C4-Slp α chains could be accounted for by point mutations affecting the glycosylation sites of this region. Three cDNA insertions representative of the clones that overlap the region encoding the carboxy-terminal half of the α chain are illustrated in figure 1. The clones are aligned, on the basis of nucleotide sequences,

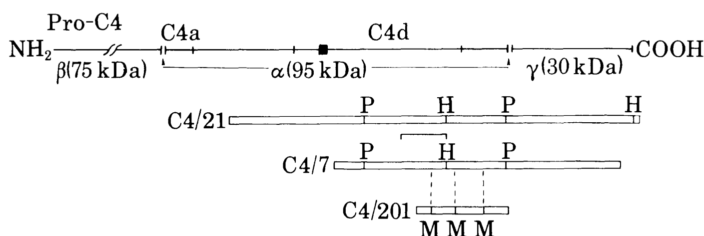

FIGURE 1. Alignment of the Pst I insertions of three representative mouse C4 cDNA clones. The upper line shows the order of the three chains of native C4 in the pro-C4 molecule. The unlabelled segment of the alpha chain on the carboxy-terminal side of C4d corresponds to the human α_4 fragment. A solid square marks the location of the thiolester bond. A square bracket above the insertion of clone pMC4/7 shows the extent (300 base pairs) of the human C4 probe used in the first screening. Restriction endonuclease sites are: Pst I (P), Hind III (H) and Mbo I (M). Note that sequence analysis of the region of clone pMC4/7 encoding the α–γ junction showed the presence of two Pst I sites separated by 41 nucleotides.

to the structure of the human precursor pro-C4. This single chain polypeptide comprises the β, α and γ chains, in that order, characteristic of native C4 (Goldberger *et al.* 1980; Parker *et al.* 1980; Karp *et al.* 1981). Clone pMC4/7 was the first to be submitted to sequence analysis for verifying its authenticity as well as for identifying the encoded portion of the protein on the basis of the amino acid sequence available for human C4. This 1900 base pair insertion is our mouse reference clone. It includes the 458 base pair cDNA sequence of the clone described by Ogata *et al.* (1983) that spans the α–γ polypeptide junction of mouse C4. At its 5′ end clone pMC4/7 goes past the 300 base pair of the human probe used for screening. On the basis of the available protein sequence data of the amino-terminal portion of the human C4d fragment, this clone appears to end 27 amino acids before the cysteine residue of the thiolester bond responsible for the covalent-binding reactions of C4 (Campbell *et al.* 1981; Harrison *et al.* 1981). The octapeptide containing this active site is highly conserved in two other C4 related proteins: C3 and α_2-macroglobulin (Sim & Sim 1981). We have determined the nucleotide sequence of this region in the cDNA insertion of clone pMC4/21, which appears to differ only in length from the insertion of clone pMC4/7. The third clone illustrated in figure 1, pMC4/201, is the only representative of an alternative C4 mRNA sequence found among eight clones overlapping the insertion of pMC4/201. A nucleotide sequence comparison was made between clones pMC4/7 and pMC4/201 for the two Mbo-I fragments occupying the central part of clone pMC4/201. Figure 2 presents the nucleotide and the deduced amino acid sequence of the message strand of clone pMC4/7 between the first and third Mbo-I sites indicated in the previous figure. The sequence was unambiguously determined on both strands of DNA fragments subcloned in phage M13 (Sanger *et al.* 1980). The last codon of the sequence we report is numbered from the first position of the tetraargine peptide proposed to mark the junction be-

Downloaded from rstb.royalsocietypublishing.org

tween the C4 α and γ chains (Ogata *et al.* 1983). The sequence presented in figure 2 probably contains the site of cleavage by the endopeptidase Factor I at the carboxy-terminal end of the C4d segment (see figure 1). The amino terminal sequence of the fragment (α_4) adjacent to the carboxy-terminus of C4d has been partly determined by Press & Gagnon (1981) for human C4. They identified an arginine at the carboxy-terminal position of human C4d and nine amino acids at the first 11 positions of the α_4 fragment. The best match we could find within the amino acid sequence presented in figure 2 with the sequence reconstructed for the human C4d–α_4

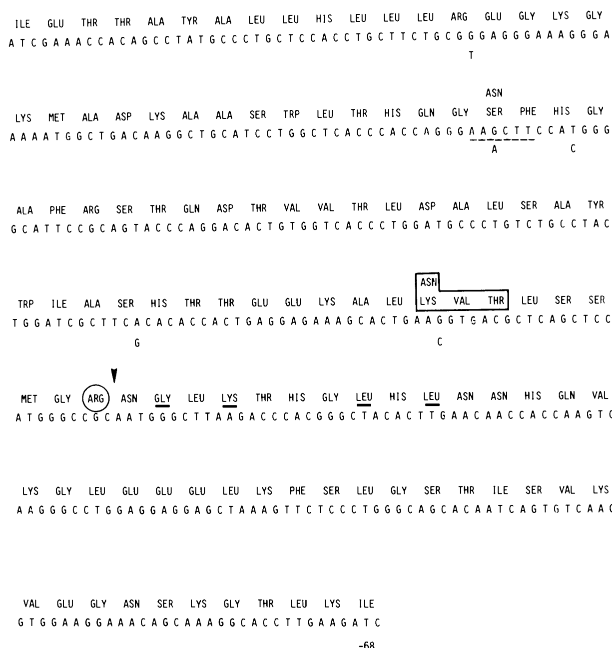

FIGURE 2. Nucleotide sequence comparison of cDNA clones pMC4/7 and pMC4/201 in the region covered by the two Mbo I fragments shown in figure 1. The sequence of the message strand of clone pMC4/7 is shown with the deduced amino acid sequence. The distance from the junction between the α and γ chain is indicated below the last amino acid residue. Nucleotide substitutions found in clone pMC4/201 are indicated below the pMC4/7 sequence, and the corresponding amino acid replacements are shown above the protein sequence deduced from clone pMC4/7. The Hind III site present in the pMC4/7 sequence is underlined by a broken line. The boxed amino acid residues mark a potential N-linked glycosylation site in the protein encoded by pMC4/201. An arrowhead shows the putative junction between the C4d and the α_4 fragments.

junction is marked by the arrowhead which identifies the putative Factor I-mediated cleavage site. This endopeptidase is also responsible for a double cleavage of the activated C3 complement component (Harrison & Lachmann 1980). However, no similarity has emerged in the four sequences carboxy-terminal to the two sites of action of Factor I, both on human C3 and C4 α chains (Davis & Harrison 1982). This suggests that there are no obvious functional constraints that should maintain sequence homology between the amino-termini of the human and murine C4 α_4 peptides. The four underlined residues in figure 2 are the only positions, out of the nine comparable, that are conserved with respect to the human α_4 sequence of Press & Gagnon (1981). In conclusion, the expected location of the α_4 amino-terminus within this mouse pro-C4 sequence is based on the consideration of the size of this peptide estimated in man to be roughly 90 residues long (Press & Gagnon 1981), and the identification of its carboxy-terminus within the tetraarginine peptide (Ogata *et al.* 1983) that we find at 111 amino acid residues

Downloaded from rstb.royalsocietypublishing.org

downstream in clone pMC4/7. Moreover, the endopeptidase Factor I seems to act on the C4 α chain with a trypsin-like specificity for a peptide bond carboxy-terminal to an arginine residue. In the sequence of figure 2 appear three arginines, only one of which is adjacent to an undecapeptide that shares four positions out of the nine comparable with the sequence of human α_4. The length of the putative murine α_4 fragment appears to exceed by some 20 amino acids the size deduced by Press & Gagnon (1981) for the homologous human fragment. At present, the significance of this discrepancy can be at best considered questionable. However, the inconsistency could be accounted for by the proteolytic trimming of the α_4 peptide at its carboxyl-terminus, proposed to occur in the murine and human C4 plasma forms (Karp *et al.* 1982; Chan *et al.* 1983). The sequence of clone pMC4/201 was found identical to that of pMC4/7 except for five non-contiguous nucleotide substitutions which are indicated in figure 2 with symbols underlying the sequence of pMC4/7. Only the second and fifth nucleotide replacements produce an amino acid change and both introduce a new asparagine residue in the peptide sequence corresponding to clone pMC4/201. It is noteworthy that the second of the two new asparagines creates with the contiguous valine and threonine residues a new potential glycosylation site (boxed in figure 2). Interestingly, the nucleotide replacement that produces a serine to asparagine substitution in clone pMC4/201 also marks the loss of the Hind III restriction site which is underlined by a broken line in figure 2 and is also characteristic of clone pMC4/21 (figure 1) as well as of five others overlapping this region. We surmise that divergence at this Hind III site between the C4 mRNAs of strain B10.W7R is a general marker of the isotypic variation between C4 and C4-Slp, and that the presence of this restriction site is in all likelihood specific for the *C4* (that is, not for the *C4-Slp*) gene in most mouse strains. Evidence supporting these conclusions would be more straightforward had the two C4 cDNA sequences been derived from one of the more common laboratory strains. In fact, we have found that $H\text{-}2^{w7}$ and the other wild derived haplotypes $H\text{-}2^{w16}$ and $H\text{-}2^{w19}$, an unusual feature of which is the testosterone-independent expression of the C4-Slp protein, exceptionally bear multiple copies of the *C4* gene (Lévi-Strauss, 1984).

This discrepancy in gene numbers is evident in the Southern blot analysis presented in figure 3. The first and second lane show Hind III digests of DNA from strains B10.W7R and B10.HTT ($H\text{-}2^{k}$) hybridized with the 3′ most Pst-I fragment of the pMC4/7 insertion (figure 1). By using genomic clones isolated by M. Steinmetz (Basel Institute for Immunology) from an AKR/J library, we have determined by quantitative Southern blot analysis that indeed $H\text{-}2^{k}$ carries only two copies of the *C4* gene as most of the common *H-2* haplotypes (Lévi-Strauss 1984). By comparison $H\text{-}2^{w7}$ appears to carry three to four additional copies. Although B10.W7R might express more than two species of C4 mRNA, the Hind III site in the C4d sequence of clone pMC4/7 is not a peculiarity of this strain. A comparison of the hybridization patterns between the second and third lane of the Southern blot presented in figure 3 shows that the two *C4* genes in the $H\text{-}2^{k}$ haplotype differ exactly for the presence and absence of a Hind III site in the region intercepted by the two hybridization probes used. Lanes 2 and 3 contain Hind III digests of B10.HTT liver DNA hybridized respectively with the 3′ most Pst I fragment of the pMC4/7 insertion and with the larger Pst I-Hind III fragment of the same insertion (figure 1). A larger hybridizing fragment of about 20 kilobases is detected with both cDNA probes, independently of their location with respect to the Hind-III site. This fragment defines the $H\text{-}2^{k}$ *C4* gene that lacks a Hind III site in the region probed. Conversely, the change in size of the other restriction fragment detected by the two probes, respectively 3′ and

Downloaded from rstb.royalsocietypublishing.org

5′ to the Hind III site of clone pMC4/7 demonstrates the presence of a Hind III site in the second *C4* gene. The sequences corresponding in this gene to the C4d region are distributed over a 5′ Hind III fragment of 13.5 kilobases and a 3′ Hind III fragment of 4.6 kilobases. We have determined that the Hind III site separating these two fragments corresponds precisely to the one shown in the sequence presented in figure 2 by limited sequence analysis of the corresponding genomic clone. A survey of Hind III digests of DNA from conventional *H-2* haplotypes hybridized with probes similar to the ones used in figure 3 indicates that the exonic Hind III site described above discriminates between the two *C4* gene copies commonly found in these strains. In particular, this is evident for the *C4* genes of the *H-2*^{d} haplotype for which a detailed molecular map has recently been published (Chaplin *et al.* 1983).

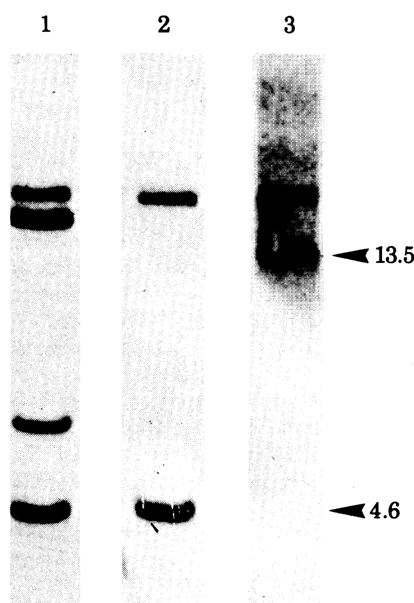

FIGURE 3. Comparison of mouse *C4* genes of the *H-2*^{w7} haplotype (lane 1) and of the *H-2*^{k} haplotype (lanes 2 and 3) by Southern blot analysis. Liver DNA (15 μg) from B10.W7R (lane 1) or from B10.HTT (lanes 2 and 3) was digested with Hind III and hybridized with the 3′ most Pst I fragment of the insertion of clone pMC4/7 (lanes 1 and 2) or with the longer Pst I-Hind III fragment isolated from the same clone (lane 3). The arrowheads point to two Hind III fragments located respectively 3′ (4.6 kilobase) and 5′ (13.5 kilobase) to the exonic Hind III site that characterizes the C4d encoding region of one of the two *C4* genes of the *H-2*^{k} haplotype.

The abundance of the cDNA class represented by clone pMC4/7 suggests that the presence of the Hind III site in the region encoding C4d identifies the C4 rather than the C4-Slp isotype because the former appears to be expressed at considerably higher levels (Ogata *et al.* 1983 and our own unpublished estimation based on immunoprecipitates of *in vitro* translation products). Interestingly, the sequence of the clone pMC4/201, here correlated with the *C4-Slp* gene, bears an additional potential glycosylation site which could contribute to the higher degree of glycosylation that distinguishes the carboxy-terminal half of the C4-Slp^{w7} chain from the corresponding part of the C4^{w7} protein (Karp 1983).

This work was supported in part by the Ministère de l'Industrie et de la Recherche under grant 83C 0855. We thank M. Carroll and R. Porter for the use of the human Alu/7 C4 clone.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

We wish to acknowledge the help of Elena Georgatsou in the last phase of this work. The editorial assistance of Jacqueline Badella and the laboratory help of Pierre Bourgarel are greatly appreciated.

References

Atkinson, J. P., Karp, D. R., Seeskin, E. P., Killion, C. C., Rosa, P. A., Newell, S. L. & Schreffler, D. C. 1982 H-2 region determined polymorphic variants of the C4, Slp, C2 and B complement proteins: A compilation. *Immunogenetics* **16**, 617–623.

Campbell, R. D., Gagnon, J. & Porter, R. R. 1981 Amino acid sequence around the thiol and reactive acyl groups of human complement component C4. *Biochem. J.* **199**, 359–370.

Carroll, M. C. & Porter, R. R. 1983 Cloning of a human complement component C4 gene. *Proc. natn. Acad. Sci. U.S.A.* **80**, 264–267.

Chakravarti, D. N., Campbell, R. D. & Gagnon, J. 1983 Amino acid sequence of a polymorphic segment from fragment C4d of human complement component C4. *FEBS Lett.* **154**, 387–390.

Chan, A. C., Mitchell, K. R., Munns, T. W., Karp, D. R. & Atkinson, J. P. 1983 Identification and partial characterization of the secreted form of the fourth component of human complement: Evidence that it is different from major plasma form. *Proc. natn. Acad. Sci. U.S.A.* **80**, 268–272.

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R. & Seidman, J. G. 1983 Molecular map of the murine *S* region. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6947–6951.

Davis III, A. E. & Harrison, R. A. 1982 Structural characterization of Factor I mediated cleavage of the third component of complement. *Biochemistry, Wash.* **21**, 5745–5749.

Goldberger, G., Abraham, G. N., Williams, J. & Colten, H. R. 1980 NH_2-terminal sequence analysis of pro-C4, the precursor of the fourth component of guinea pig complement. *J. biol. Chem.* **255**, 7071–7073.

Hansen, T. H. & Shreffler, D. C. 1976 Characterization of a constitutive variant of the murine serum protein allotype, Slp. *J. Immunol.* **117**, 1507–1513.

Harrison, R. A. & Lachmann, P. J. 1980 The physiological breakdown of the third component of human component. *Mol. Immunol.* **17**, 9–20.

Harrison, R. A., Thomas, M. L. & Tack, B. F. 1981 Sequence determination of the thiolester site of the fourth component of human complement. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7388–7392.

Karp, D. R., Parker, K. L., Atkinson, J. P., Shreffler, D. C. & Capra, J. D. 1981 Characterization of the murine C4 precursor (pro-C4): evidence that the carboxy-terminal subunit is the C4 γ-chain. *J. Immunol.* **126**, 2060–2061.

Karp, D. R., Shreffler, D. C. & Atkinson, J. P. 1982 Characterization of the M_r difference between secreted murine fourth component of complement and the major plasma form: Evidence for carboxy-terminal cleavage of the α chain. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6666–6670.

Karp, D. R. 1983 Variation in the glycosylation of the murine sex-limited protein: Comparison with the fourth component of murine complement. *J. Immunol.* **131**, 1405–1410.

Lévi-Strauss, M. 1984 (In preparation.)

Ogata, R. T., Shreffler, D. C., Sepich, D. S. & Lilly, S. P. 1983 cDNA clone spanning the α–γ subunit junction in the precursor of the murine fourth complement component (C4). *Proc. natn. Acad. Sci. U.S.A.* **80**, 5061–5065.

Parker, K. L., Capra, J. D. & Shreffler, D. C. 1980 Partial amino acid sequences of the murine fourth component of complement (C4): Demonstration of homology with human C4 and identification of the amino-terminal subunit in Pro-C4. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4275–4278.

Press, E. M. & Gagnon, J. 1981 Human complement component C4: Structural studies on the fragments derived from C4b by cleavage with C3b inactivator. *Biochem. J.* **199**, 351–357.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. 1980 Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. mol. Biol.* **143**, 161–178.

Shreffler, D. C., Atkinson, J. P., Brown, L. J., Parker, K. L. & Roos, M. H. 1981 Genetics, Structure and Function of Murine S Region Gene Products. In *Immunobiology of the major histocompatibility complex*, 7th int. Convoc. Immunol., Niagara Falls, N.Y., 1980, pp. 78–88. Basel: Karger.

Sim, R. B. & Sim, E. 1981 Autolytic fragmentation of complement components C3 and C4 under denaturing conditions. A property shared with α_2-macroglobulin. *Biochem. J.* **193**, 129–141.

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

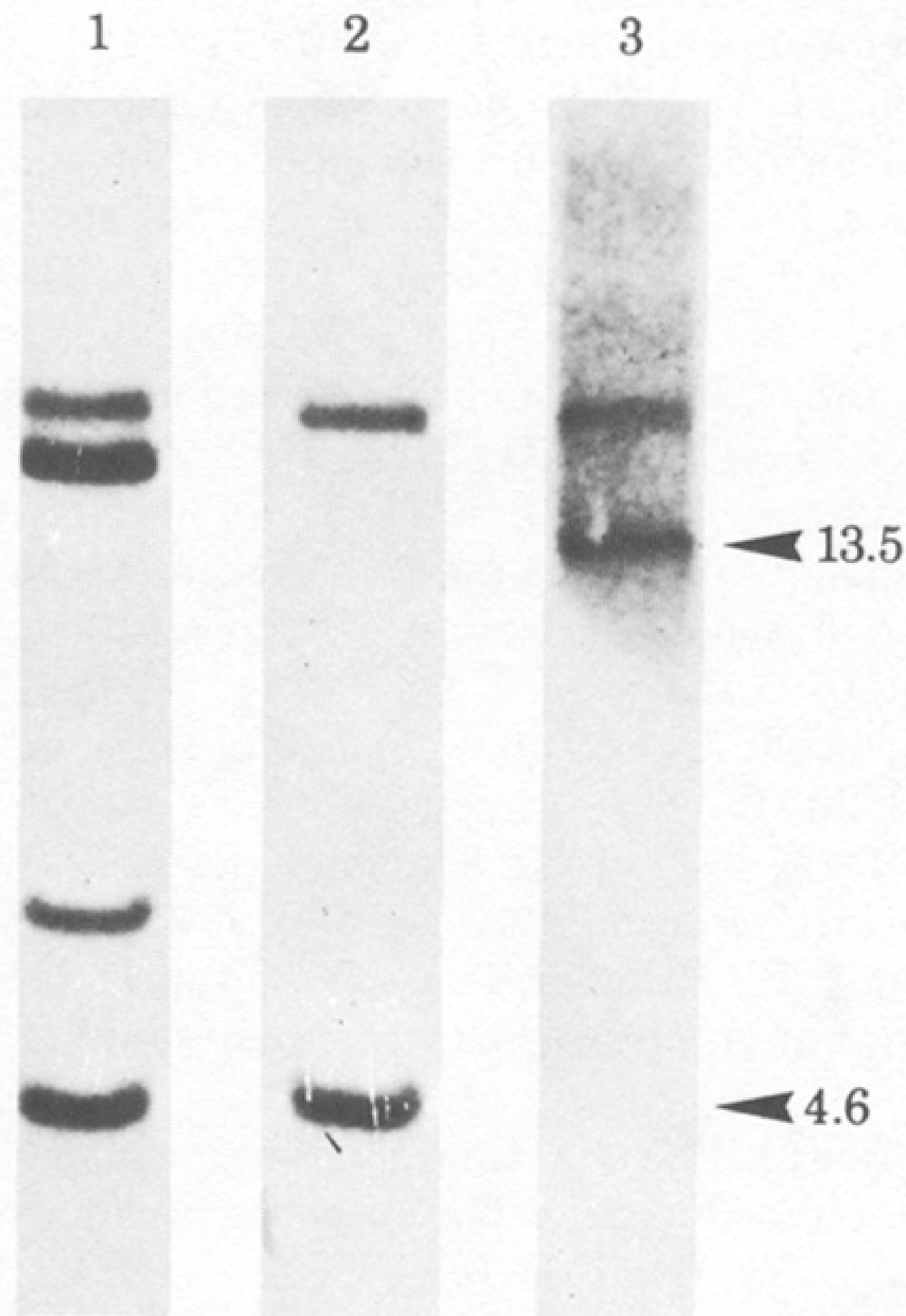

FIGURE 3. Comparison of mouse *C4* genes of the *H-2*^{w7} haplotype (lane 1) and of the *H-2*^{k} haplotype (lanes 2 and 3) by Southern blot analysis. Liver DNA (15 μg) from B10.W7R (lane 1) or from B10.HTT (lanes 2 and 3) was digested with Hind III and hybridized with the 3′ most Pst I fragment of the insertion of clone pMC4/7 (lanes 1 and 2) or with the longer Pst I-Hind III fragment isolated from the same clone (lane 3). The arrowheads point to two Hind III fragments located respectively 3′ (4.6 kilobase) and 5′ (13.5 kilobase) to the exonic Hind III site that characterizes the C4d encoding region of one of the two *C4* genes of the *H-2*^{k} haplotype.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES