Vol. 200, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 15, 1994 Pages 202-205

THE PUTATIVE FIFTH HUMAN SERUM AMYLOID A PROTEIN (SAA)-RELATED GENE *"SAA5"* IS DEFINED BY *SAA3*

Grant C. Sellar and Alexander S. Whitehead#

Department of Genetics and Biotechnology Institute, Trinity College, Dublin 2, Ireland

Received February 24, 1994

The four well characterised members of the human serum amyloid A protein (SAA) gene family are clustered on chromosome 11p15.1. The acute phase SAA genes, *SAA1* and *SAA2*, are hyperinducible in response to inflammatory stimuli, whereas *SAA4* is only minimally induced, and *SAA3* is a pseudogene. We recently demonstrated that the GSAA4 sequence, reported by others (Sack, G.H. Jr. and Talbot, C.C. Jr., 1992. Biochem. Biophys. Res. Comm. 183, 362-366), and misidentified as corresponding to the *SAA4* locus, maps to the 11p15 region and speculated that it may be in close proximity to a distinct fifth SAA locus: *"SAA5"*. In this report we have used vectorette PCR in combination with direct sequencing and computer based homology searches of the nucleotide sequence databases to establish that the putative fifth SAA-related locus, *"SAA5,"* is defined by *SAA3* and therefore does not represent a distinct SAA gene.

The human serum amyloid A protein (SAA) gene family comprises four well characterized members located on chromosome 11 (1) on band p15.1 (2). *SAA1* and *SAA2* encode the hyperinducible acute phase SAAs (A-SAAs) SAA1 and SAA2, respectively; *SAA3* is a pseudogene; and *SAA4* encodes constitutive SAA (C-SAA). The four genes share a similar four exon organization although their intron sizes differ.

As a component of our goal to identify all members of the human SAA gene family we have continued our characterization of a putative fifth SAA locus: *"SAA5"* (1). We recently speculated that sequence derived from the human genomic clone GSAA4, reported by Sack and Talbot (3), may be in close proximity to a fifth distinct SAA locus: *"SAA5"* (1). GSAA4 was originally identified by Sack and Talbot (3) by cross-hybridization with a probe from the third and fourth exons of *SAA3*. The partial sequence of GSAA4 reported by these authors was misidentified as corresponding to that of the first exon of the *SAA4* gene which had originally been identified by Betts et al (4) and was subsequently sequenced in full by us (5). Of the 707bp of published GSAA4 sequence, an alignment of 300bp with the putative first exon of

#Author to whom correspondence should be addressed: Fax: 353-1-6798558.

ABBREVIATIONS:

SAA: serum amyloid A protein; PCR: polymerase chain reaction; YAC: yeast artificial chromosome; BLAST: basic local alignment search tool; UTR: untranslated region.

0006-291X/94 $5.00

the *SAA3* specific genomic clone GSAA1 (6) was also reported (3). Our own computer-based sequence analyses identified a repetitive element in the GSAA4 sequence (1). Despite its repetitive nature it was possible to design PCR (polymerase chain reaction; 7) primers from the sequence to yield a specific amplification product and to permit confirmation of its identity by direct sequencing. This locus specific PCR product was used to establish that the gene containing this sequence, like *SAAs1/2/3/4*, maps to the short arm of chromosome 11 (1). As part of the above study the 707bp of sequence presented by Sack and Talbot (3) was used in the orientation reported to search the nucleotide sequence databases using the BSEARCH programme (8) which was available to us on our VAX mainframe at that time. No SAA sequence from any species was present in the sequences comprising the top thirty scores, all of which contained repetitive Alu sequences. However, based on the isolation of GSAA4 with an *SAA3* gene probe and our definitive mapping of the sequence defined by GSAA4 to the *SAA* region of chromosome 11, we nevertheless speculated that the sequence described is in close proximity to an SAA coding sequence and may mark the position of a distinct fifth SAA superfamily locus: *"SAA5"* (1).

In this report we have used vectorette PCR (9) to extend directly from the 707bp of reported GSAA4 sequence. Direct sequencing of this product and further computer based homology analysis using the original GSAA4 sequence revealed that *"SAA5"* is defined by *SAA3*.

METHODS

Vectorette PCR and Direct Sequencing

DNA in agarose prepared from a "SAA5" positive yeast artificial chromosome (YAC) clone was digested with SmaI and a vectorette library constructed by the method of Riley et al. (9). Vectorette PCR was performed between the universal vectorette primer 224 (9; 5'-CGAATCGTAACCGTTCGTACGAGAATCGCT-3') and the locus specific *"SAA5"* primer HSAA5 PR7 (5'-GTACTAGGGATATAGAATTC-3) designed from the most 3' sequence of GSAA4 (3). PCR amplification conditions were 94°C for 7 minutes followed by 40 cycles of 94°C for 1'; 55°C for 2'; 72°C for 5'. The PCR product was excised from SeaPlaque (FMC) agarose, purified through a MAGIC PCR column (Promega) and sequenced directly using Sequenase Version 2 (United States Biochemicals). Oligonucleotide primers were synthesised using a PCR-MATE DNA Synthesiser Model 391 (Applied Biosystems).

SAA Hybridization Analysis

The vectorette PCR product was digested with the restriction enzymes EcoRI, HindIII, PstI and XbaI and fragments separated through a 1% (w/v) agarose gel in 1X TBE. The gel was then Southern blotted onto Gene Screen Plus membrane (NEN/DUPONT) using a PosiBlot apparatus (Stratagene) following the manufacturers' instructions. The resulting blot was prehybridized in 10% (w/v) dextran sulphate/1M NaCl/1% (w/v) SDS at 65°C and hybridized overnight with a probe specific for the third exon of human SAA3. Following washes according to Gene Screen Plus protocols, hybridizing signals were visualized after autoradiography for 1hr at room temperature or -70°C. Signals were removed from the membrane as recommended by the manufacturer and subsequently rehybridized with a probe specific for the coding regions of human SAA1 and SAA2. Probes were labelled by incorporation of [α-^{32}P]dCTP using a Random Prime labelling kit (Boehringer) and unincorporated radioactivity removed over a prepacked G-50 Nick" column (Pharmacia).

Nucleotide Sequence Analysis

Simple sequence analyses were performed using PCGENE and associated software (Intelligenetics). Database homology searches were performed on the VAX mainframe using the Basic Local Assignment Search Tool (BLAST; 10).

RESULTS and DISCUSSION

A SmaI vectorette library was constructed from a YAC clone from the ICI human YAC library (11) identified by locus specific PCR as containing *"SAA5"* (unpublished data). This YAC clone also contains *SAA3* and *SAA4* (unpublished data). From the published restriction map of GSAA4 (3), a ~2kb PCR product should be obtained from amplification between the universal vectorette primer 224 (9) and the locus specific *"SAA5"* primer HSAA5 PR7 designed from the most 3' sequence of GSAA4 (3). A single product of the expected size was obtained using this approach. Restriction analysis of this product with the enzymes used by Sack and Talbot (3) in their restriction map of GSAA4, i.e. EcoRI, HindIII, PstI, XbaI, was consistent with its being derived from the GSAA4 or *"SAA5"* locus. A Southern blot of the digested vectorette PCR product was hybridized with a probe specific for the third exon of *SAA3* and subsequently with a *SAA1/2* probe covering the coding regions of human *SAA1* and *SAA2*, which does not cross-hybridize with the third exon of *SAA3*. Both probes have been used in the mapping of a human SAA YAC contig and will be reported in that study (unpublished data). This hybridization analysis established the presence of sequence with homology to the third exon of *SAA3*, and also confirmed that the product contains additional SAA-like coding sequence. To identify the nature of this SAA coding sequence the vectorette PCR product was purified and directly sequenced (data not shown) using HSAA5 PR7 as a sequencing primer. The sequence obtained was then analysed using PCGENE and associated software (Intelligenetics). Direct translation of this sequence in three reading frames did not reveal any SAA or SAA-like protein sequence. However, when the sequence was inverted and complemented and then translated, the derived amino acid sequence was readily identifiable as SAA-like. Further examination of the sequence identified the sequence as being derived from the fourth exon and 3' untranslated region (UTR) of human *SAA3*. However, the data originally presented by Sack and Talbot (1992) indicated that we should have identified coding sequence from the second exon of this SAA-like locus from the strand as sequenced. Our data strongly suggested that *"SAA5"* in fact corresponds to *SAA3*, and that the clone GSAA4 contains the 3' portion of the *SAA3* gene.

Further examination of the published restriction map of GSAA4 (3) showed that it is an almost exact inversion (over those enzymes in common) of the 3' region of the human *SAA3* clone GSAA1 originally reported by the same authors (6). Since our original nucleotide sequence database searches were conducted, we have acquired the BLAST program (10) which automatically performs the search using both strands. A BLAST search with the 707bp GSAA4 sequence identified *SAA3* as the top match. Nucleotides 707 to 461 from GSAA4 (3) align on the minus strand with a score of 1235 and 100% identity over 247 residues with nucleotides 6534 to 6780 from the 3' UTR of clone GSAA1 (6; HSSAA1A; accession number X13895) including the polyadenylation signal AATAAA. This result was entirely consistent with the sequencing data obtained from the vectorette PCR product.

In summary, in this study we have definitively established that the putative fifth human SAA locus, *"SAA5"*, is defined by *SAA3*. However, we consider it likely that additional SAA genes exist and we are continuing to search the p15.1 region of human chromosome 11 to identify these.

ACKNOWLEDGMENTS

We wish to thank Ms. Clarissa Uhlar for technical assistance and Dr. Andrew Lloyd, Irish National Center for BioInformatics, for useful discussions on database searches. This work was supported by a Wellcome Trust Program grant. GCS is a Beit Memorial Research Fellow.

REFERENCES

1. Sellar, G.C., and Whitehead, A.S. (1993). Genomics 16, 774-776.
2. Sellar, G.C., Jordan, S.A., Bickmore, W.A., Fantes, J.A., Van Heyningen, V., and Whitehead, A.S. (1994). Genomics 19, (in press).
3. Sack, G.H. Jr. and Talbot, C.C. Jr. (1992). Biochem. Biophys. Res. Comm. 183, 362-366.
4. Betts, J.C., Edbrooke, M.R., Thakker, R.V., and Woo, P. (1991). Scand. J. Immunol. 34, 471-482.
5. Steel, D.M., Sellar, G.C., Uhlar, C.M., Simon, S., DeBeer, F.C., and Whitehead, A.S. (1993) Genomics 16, 447-454.
6. Sack, G.H., Jr., and Talbot, C.C., Jr. (1989). Gene 84, 509-515.
7. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Science 239, 487-491.
8. Higgins, D.G. (1989). Genetics Department, Trinity College, Dublin 2, Ireland.
9. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C., and Markham, A.F. (1990). Nucleic Acids Res. 18, 2887-2890.
10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). J. Mol. Biol. 215, 403-410.
11. Anand, R., Riley, J.H., Butler, R., Smith, J.C., and Markham, A.F. (1990). Nucleic Acids Res. 18, 1951-1956.