

Identification of the autoantigen SART-1 as a candidate gene for the development of atopy

Amanda P. Wheatley¹, Daniel J. Bolland², Jane E. Hewitt², Jane C. Dewar¹ and Ian P. Hall^{1,*}

¹Division of Therapeutics and ²Institute of Genetics, University Hospital of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK

Received April 29, 2002; Revised and Accepted June 28, 2002

The possibility that immune responses to autoantigens may contribute to the development of atopic disease has been largely ignored. In this paper, we describe the chromosomal localization of the gene for squamous cell carcinoma-associated reactive antigen for cytotoxic T cells (SART-1). The *SART-1* gene localized to a region of 11q12–13 showing strong linkage to atopy in previous studies. Further analysis of this gene revealed the presence of at least 20 exons of varying lengths and four novel single-nucleotide polymorphisms, one of which resulted in an amino acid substitution. Association analysis in families recruited on the basis of affected sib pairs for asthma reveal significant association for both coding region polymorphisms with atopy. We therefore hypothesize that polymorphic variation within the *SART-1* gene may account for individuals developing atopy.

INTRODUCTION

Whilst sensitization to external allergens such as house dust mite (HDM) is common in the general population, relatively little is known about the role of autoantigens in atopic diseases such as asthma and eczema. Autoantigens have been shown to be present in some patients with atopic dermatitis (1). The squamous cell carcinoma-associated reactive antigen for cytotoxic T cells 1 (SART-1) is a leucine zipper DNA-binding protein of 800 amino acids, and is normally found in the nucleus of proliferating epithelial cells (2). In addition, a shortened form (259 amino acids) is expressed in the cytosol of most squamous cell carcinomas (3). This 259-amino-acid form of SART-1 has been found in serum from some patients with atopic dermatitis (4). The shortened form may be released into the extracellular matrix upon tissue damage, potentially resulting in its recognition by antibodies. In one study, antiserum raised to recombinant SART-1 was found to react with cytoplasmic protein exhibiting a broad cellular and tissue reactivity (4) in the skin, lung, gastrointestinal tracts, muscle and brain. We therefore hypothesized that *SART-1* may be an important candidate gene for atopy.

RESULTS

Initial sequence comparison of a partial clone (GenBank accession no. AF109680) (5) revealed almost complete homology to the cDNA database sequences for *SART-1* [GenBank accession nos AB006198 (3) and Y14314 (4)]. We

mapped the human *SART-1* gene using the GeneBridge 4 whole-genome radiation hybrid panel (6). Following screening of the panel, data were analyzed using the Rmapper server at <http://carbon.wi.mit.edu:8000/cgi-bin/contig/rmapper.pl>. Clone AF109680 was thereby mapped to chromosome 11q12–13. The microsatellite marker D11S97 in this region shows strong linkage to atopy, and subsequent studies have confirmed linkage of this region to atopy in some but not all studies (7–12). This region also contains the gene for the high-affinity IgE receptor FcεR1-β, which is located close to *SART-1* (Fig. 1A).

Having localized the gene for SART-1 to a chromosomal locus with good evidence for linkage to atopy, we next undertook a screen of the coding region of this gene to define the degree of polymorphic variation within the gene. Database comparison of the three available sequences for SART-1 suggested that polymorphic variation might be present at basepairs 774, 1463 (a 3 bp deletion) and 1926 (from the translational start site). Direct sequencing of PCR products across the base-pair 774 and 1454 sites revealed the presence of two common single-nucleotide polymorphisms (SNPs). We were unable, however, to identify any individual with a 3 bp deletion at base pair 1463, despite sequencing this region in genomic DNA from 10 different subjects. Comparison of genomic sequence with the cDNA sequence on the database revealed the presence of several introns within this region of *SART-1*. Further analysis of the intron/exon arrangement of the whole *SART-1* gene was therefore performed using a combination of *in silico* and PCR/sequencing approaches with genomic DNA as template. The probable intron/exon arrangement of the *SART-1* gene is shown in Figure 1B, although to date we have

*To whom correspondence should be addressed. Tel: +44 1159709905; Fax: +44 1159422232; Email: ian.hall@nottingham.ac.uk

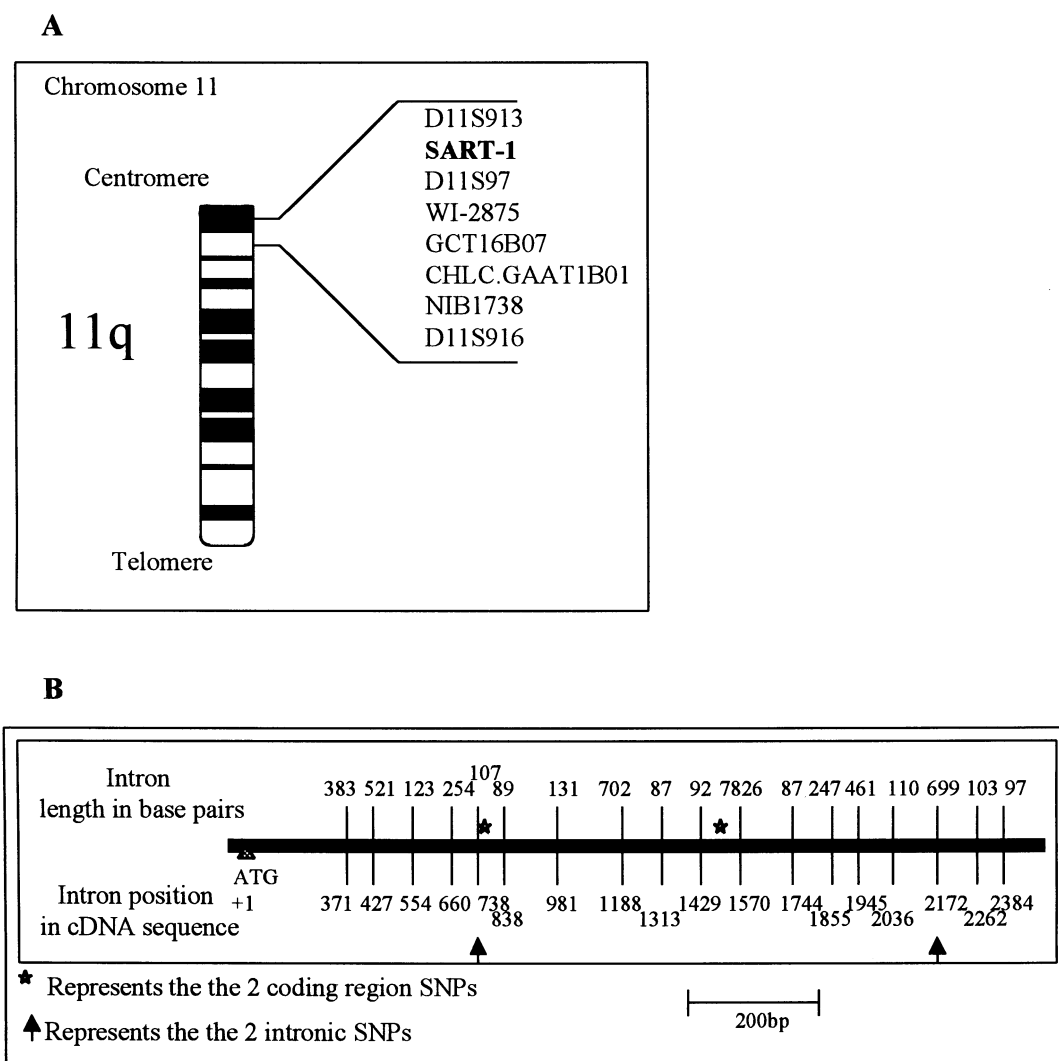


Figure 1. (A) Location of *SART-1* and relevant markers on chromosome 11q. (B) Intron positions and lengths within *SART-1* determined by a database search using the *SART-1* cDNA sequence.

been unable to amplify the 5' end of the coding region (ATG to +500 bp) and a second region between base pairs 1427 and 1855, from genomic DNA. This is probably because large introns (and possibly further polymorphic variation) may exist within these regions. BLAST searching of the High-Throughput Genomic Sequence (HTGS) database using the published human cDNA sequence (GenBank accession no. AB006198) identified a number of working draft sequences containing part of the human gene (GenBank accession nos AP000592, AP000586, AP001191 and AP001201). Comparison allowed our deduced genomic structure to be confirmed and the regions for which no sequence data could be obtained to be inferred. The human gene therefore contains at least 20 exons, with intron sizes ranging from 88 bp to >7.8 kb. Two further SNPs were identified within introns 6 and 17, as indicated in Figure 1B. Subsequent to completion of this study, a search of the *SART-1* locus using dbSNP confirmed three of the four SNPs identified in the present study (rs556643,

rs688862 and rs660118) and in addition the presence of 13 further SNPs located in intronic regions of the gene.

The allelic frequencies of the two identified SNPs within the *SART-1* coding region in a Caucasian population of 211 individuals are 774C 0.55, 774T 0.45, 1454C 0.52 and 1454G 0.48. The SNP at base pair 774 (C→T) is degenerate (Thr258). However, the SNP at base pair 1454 (G→C) results in a substitution of a glycine for alanine in codon 485. Both of these SNPs are in Hardy-Weinberg equilibrium in the population studied (see below). However, strong linkage disequilibrium exists between the two polymorphisms ($\chi^2 = 229$, 1df, $P < 0.0001$) with the common haplotypes being 774T/1454G (Gly485) and 774C/1454C (Ala485).

Having localized the *SART-1* gene to a candidate region of atopy on 11q12-13 and identified polymorphic variation within this gene, we next looked for association of the polymorphic variants with atopy in a nuclear family sample (51 families, 211 individuals) recruited from the Nottingham area on the basis of

affected sib pairs for asthma. Significant association with atopy defined as one or more positive skin prick tests to common allergen was observed within these families for both the Ala485 variant of *SART-1* ($P=0.008$, Tables 1 and 2) and the C allele of the base-pair 774 SNP. No significant association was seen with asthma *per se*, suggesting that *SART-1* may be implicated in the development of atopy but not asthma. These data are consistent with linkage studies, which have demonstrated significant linkage to this region for atopy but not asthma (8,9,13). We also attempted to perform a transmission disequilibrium analysis for these SNPs but, because of the high prevalence of atopy within the families studied, the number of informative transmissions was very low and no significant transmission disequilibrium was found (Gly485, 24 versus 26 transmissions, $P=NS$; C774, 28 versus 24 transmissions, $P=NS$).

DISCUSSION

These data taken together with linkage taken from other studies suggest that *SART-1* may be a strong candidate gene for the development of atopy. However, the true potential role of polymorphism in *SART-1* will only become clear after replication studies have been undertaken. Whilst atopy is usually considered to be driven by immunoglobulin E (IgE)-mediated response to external allergens, it remains far from

clear why atopy develops in some individuals but not others, despite similar exposure to allergen. We hypothesize that an autoimmune response to autoantigens such as *SART-1* may also be involved in the development of atopic disease and that therapeutic strategies designed to target these responses may be effective in the management of atopic disease. Autoantibodies may result either because of cross-reactivity of epitopes of different polymorphic forms of autoantigens with exogenous antigens, or because of differing antigenicity of autoantigens such as *SART-1* due to the polymorphism within the gene: in the latter model, autoantigen release occurs as a consequence of initial tissue damage following antigen exposure. Polymorphic variation within autoantigens such as *SART-1* would thus contribute to the intensity of autoantibody formation and hence in part dictate the development and the severity of atopy. At present, no functional data are available on the potential for the polymorphisms identified in the current study to alter immunogenicity of *SART-1*: future studies on this subject are required.

MATERIALS AND METHODS

Sample

The study sample consisted of 51 families ($n=211$) from the Nottingham area recruited on the basis of two siblings having doctor-diagnosed asthma. For the analyses, the diagnosis of asthma was confirmed using a previously validated questionnaire. Atopy was defined as one or more positive skin prick tests to common allergens >3 mm over control: the allergens tested were Der p1, cat dander, grass pollen, and *Aspergillus fumigatus*. A positive histamine control was included for all individuals. EDTA blood samples were obtained from individuals for DNA extraction and subsequent genotype analysis.

Mutation screening

A database search was first carried out using the *SART-1* sequence. Itoh *et al.* (3) published the full-length cDNA sequence of *SART-1* (GenBank accession no. AB006198), described as an antigen recognized by cytotoxic T cells of a squamous cell carcinoma patient. Valenta *et al.* (4) published a DNA sequence for an expressed human autoantigen named *HOMS-1* (GenBank accession no. Y14314) with almost complete sequence identity to *SART-1*; this sequence is unlikely to represent the full-length cDNA, since it leaves out the first 38 amino acids compared with the Itoh *et al.* (3) sequence. The human genomic clone identifying *SART-1* to chromosome 11 (GenBank accession no. AF109680) contains a 94-amino-acid region of *SART-1* spanning amino acids 637–731 of the protein, interrupted by three intronic regions.

Sequence alignment of these published sequences using BLAST (14) allowed *in silico* screening (15) for mutation detection. A number of SNPs and a 3 bp deletion were observed that might suggest the presence of true polymorphisms or could represent sequencing errors. Further analysis was performed as outlined above for the two potential polymorphic sites identified *in silico*. The first potential polymorphism

Table 1. Association for the atopy trait with the *SART-1* loci using the sib-pair program (17)

	Affected	Unaffected	Total no. of alleles
<i>SART-1</i> locus 1454:			
Allele (C)	127 (0.53)	62 (0.40)	189
Allele (G)	111 (0.47)	94 (0.60)	205
Total no. of alleles	238	156	394
<i>SART-1</i> locus 774:			
Allele (C)	115 (0.50)	63 (0.38)	178
Allele (T)	115 (0.50)	101 (0.62)	216
Total no. of alleles	230	164	394

Locus 1454: contingency Pearson's $\chi^2=7.0$ ($P=0.008$).

Locus 774: contingency Pearson's $\chi^2=5.2$ ($P=0.023$).

Table 2. Association of the asthma trait with the *SART-1* loci using the sib-pair program

	Affected	Unaffected	Total no. of alleles
<i>SART-1</i> locus 1454:			
Allele (C)	131 (0.48)	59 (0.48)	190
Allele (G)	141 (0.52)	65 (0.52)	206
Total no. of alleles	272	124	396
<i>SART-1</i> locus 774:			
Allele (C)	117 (0.44)	62 (0.48)	179
Allele (T)	151 (0.56)	66 (0.52)	217
Total no. of alleles	268	128	396

Locus 1454: contingency Pearson's $\chi^2<0.01$ ($P=0.91$).

Locus 774: contingency Pearson's $\chi^2=0.8$ ($P=0.37$).

contained the 3 bp deletion resulting in the Gln488 deletion observed in the Valenta *et al.* (4) sequence. The second potential polymorphism was the degenerate SNP C→T at nucleic acid position 774 from the translational start site within codon 258. Neither of these positions were present in our partial clone (GenBank accession no. AF109680).

Sequencing

Genomic DNA was extracted from whole blood using standard techniques (QiAMP kit, Qiagen). Polymerase chain reaction (PCR) was carried out for fragments spanning the two sites of interest. Hot-start PCR was used, and 32 cycles of 94°C, 61°C and 72°C for 90 s, each followed by a final extension temperature of 72°C for 10 min, were carried out for the PCR across the 774 bp region. Primer sequences spanning this region were (forward) 5'-GGAGCCGGCAGCTGCAGAA-3' and (reverse) 5'-CCACGCTCTCGTCCTCGGCATA-3', producing a fragment of 785 bp rather than the expected 349 bp due to the presence of a previously unknown intron. Primer sequences spanning the 3 bp deletion at 1463 were (forward) 5'-GAAGGAGCCTGTGCCTCAG-3' and (reverse) 5'-CACTGTCTCGCAGCTGCTG-3', giving a PCR product of 208 bp. PCR conditions were 35 cycles and the same temperature as above. The PCR products were then purified (Wizard PCR purification system, Promega) for direct sequencing. Automated sequencing (ABI Prism) using the upstream PCR amplicon as the sequencing primer was carried out for 10 samples (i.e. 20 chromosomes) at each location.

Direct sequencing failed to confirm the presence of the Gln488 deletion in any individual: however, a non-degenerate SNP G→C was observed within codon 485, resulting in an amino acid substitution of glycine for alanine (nucleotide position 1454 from translational start site). Direct sequencing confirmed the presence of the second polymorphism (C→T at nucleotide position 774).

Genotyping

Allele-specific oligonucleotide (ASO) hybridization for each of these two novel polymorphisms within *SART-1* was used for genotyping the family population as previously described (16). PCR fragments produced using the above conditions were dot-blotted onto Nylon N+ membrane (Amersham) for use as template in the assay. Probes were designed for specificity, containing a single base difference between the two allelic variants across each polymorphism. The probe sequences were (for the 774 SNP) 5'-GGGCCTCACCGTGGAGCATG-3' and 5'-GGGCCTCACTGTGGAGCATG-3' and (for the 1454 SNP) 5'-GGAGCTCCACCGCCGGCGTCC-3' and 5'-GGAGCTCCACCGCCGGGGTCC-3'.

Statistics

A sib-pair program for simple genetic analysis (SIB-PAIR version 0.99.3) (17) was used to analyze the family data obtained for each polymorphism. The individual phenotypes (asthma and atopy) were tested separately for an association to

the genotypes of *SART-1*. Contingency Pearson χ^2 values and transmission disequilibrium data were produced. An estimated haplotype (EH) program was used to test for linkage disequilibrium between the two polymorphisms: the double-heterozygous individuals are included in the analysis using this program, which uses a gene-counting EM algorithm to estimate haplotype combinations in these individuals.

REFERENCES

- Natter, S., Seiberler, S., Hufnagl, P., Binder, B., Hirshl, A., Ring, J., Abeck, D., Schmidt, T., Valent, P. and Valenta, R. (1998) Isolation of cDNA clones coding for IgE autoantigens with serum IgE from atopic dermatitis patients. *FASEB J.*, **12**, 1559–1569.
- Gotoh, M., Shichijo, S., Hoshino, T., Imai, Y., Imaizumi, T., Inoue, Y., Takasu, H., Yamaoka, T. and Itoh, K. (1998) Sequence analysis of genes encoding rodent homologues of the human tumour-rejection antigen SART-1. *Jpn J. Cancer Res.*, **89**, 849–854.
- Itoh, K., Nakao, M., Imai, Y., Toh, Y. and Yamana, H. (1995) Tumour rejection antigens expressed on human squamous cell carcinoma. *Hum. Cell*, **8**, 149–154.
- Valenta, R., Natter, S., Sieberler, S., Wichlas, S., Maurer, D., Hess, M., Pauelka, M., Grote, M., Ferreira, F., Szepefalusi, Z. *et al.* (1998) Molecular characterisation of an autoantigen, Hom s 1, identified by serum IgE from atopic dermatitis patients. *J. Invest. Dermatol.*, **111**, 1178–1183.
- Bolland and Hewitt (1999) Accession number AF109680.
- Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillet, D., Muselet, D., Prud'Homme, J., Dib, C., Auffray, C. *et al.* (1996) A radiation hybrid map of the human genome. *Hum. Mol. Genet.*, **5**, 339–346.
- Cookson, W., Sharp, P., Faux, J. and Hopkin, J. (1989) Linkage between immunoglobulin E responses underlying asthma and rhinitis and chromosome 11q. *Lancet*, **333**, 1292–1295.
- Shirakawa, T., Hashimoto, T., Furuyama, J. and Morimoto, K. (1994) Linkage between severe atopy and chromosome 11q13 in Japanese families. *Clin. Genet.*, **46**, 228–232.
- Hizawa, N., Yamaguuchi, E., Furuya, K., Ohnuma, N., Kodama, N., Kojima, J., Ohe, M. and Kawakami, Y. (1995) Association between high serum total IgE levels and D11S97 on chromosome 11q13 in Japanese subjects. *J. Med. Genet.*, **32**, 363–369.
- Ullbrecht, M., Eisenhut, T., Bonisch, J., Kruse, R., Wjst, M., Heinrich, J., Wichmann, H., Weiss, E. and Albert, E. (1997) High serum IgE concentrations: association with HLA-DR and markers on chromosome 5q31 and chromosome 11q13. *J. Allergy Clin. Immunol.*, **99**, 828–836.
- Cox, H., Moffatt, M., Faux, J., Walley, A., Coleman, R., Trembath, R., Cookson, W. and Harper, J. (1998) Association of atopic dermatitis to the beta subunit of the high affinity immunoglobulin E receptor. *Br. J. Dermatol.*, **138**, 182–187.
- Deichmann, K., Starke, B., Schlenker, S., Heinzmann, A., Sparholt, S., Forster, J. and Kuehr, J. (1999) Linkage and association studies of atopy and the chromosome 11q13 region. *J. Med. Genet.*, **36**, 379–382.
- Cookson, W., Young, R., Sandford, A., Moffatt, M., Shirakawa, T., Sharp, P., Faux, J., Julier, C., Souef, P.L., Nakamura, Y. *et al.* (1992) Maternal inheritance of atopic IgE responsiveness on chromosome 11q. *Lancet*, **340**, 381–384.
- Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Sunyaev, S., Hanke, J., Aydin, A., Wirkner, U., Zastrow, I., Reich, J. and Bork, P. (1999) Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. *J. Mol. Med.*, **77**, 754–760.
- Hall, I., Wheatley, A., Wilding, P. and Liggett, S. (1995) Association of Glu 27 β 2-adrenoceptor polymorphism with lower airway reactivity in asthmatic subjects. *Lancet*, **35**, 1213–1214.
- Duffy, D. (1997) A program for simple non-parametric genetic analysis. Includes IBD and IBS based AMP, Haseman-Elston sib-pair, TDT and association analyses. *Am. Soc. Hum. Genet.*, **61**, A197.