

BASIC RESEARCH

Characterization of human gene encoding SLA/LP autoantigen and its conserved homologs in mouse, fish, fly, and worm

Chun-Xia Wang, Andreas Teufel, Uta Cheruti, Joachim Grötzinger, Peter R Galle, Ansgar W Lohse, Johannes Herkel

Chun-Xia Wang, Ansgar W Lohse, Johannes Herkel, I. Department of Medicine, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
Chun-Xia Wang, Shandong University Shandong Provincial Hospital, Jinan, Shandong Province, China
Andreas Teufel, Uta Cheruti, Peter R Galle, I. Department of Medicine, Johannes Gutenberg University, Mainz, Germany
Joachim Grötzinger, Institute of Biochemistry, Christian-Albrechts-University, Kiel, Germany
Supported by the Deutsche Forschungsgemeinschaft (SFB 548)
Co-first-authors: Chun-Xia Wang and Andreas Teufel
Correspondence to: Johannes Herkel, I. Department of Medicine, University Medical Centre Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany. jherkel@uke.uni-mainz.de
Telephone: +49-40-42803-3910 Fax: +49-40-42803-8531
Received: 2005-05-20 Accepted: 2005-08-26

mimicry.

© 2006 The WJG Press. All rights reserved.

Key words: Autoimmune hepatitis; Autoantigen; Genomics; Bioinformatics

Wang CX, Teufel A, Cheruti U, Grötzinger J, Galle PR, Lohse AW, Herkel J. Characterization of human gene encoding SLA/LP autoantigen and its conserved homologs in mouse, fish, fly, and worm. *World J Gastroenterol* 2006; 12(6): 902-907

<http://www.wjgnet.com/1007-9327/12/902.asp>

Abstract

AIM: To approach the elusive function of the SLA/LP molecule, we have characterized genomic organization and conservation of the major antigenic and functional properties of the SLA/LP molecule in various species.

METHODS: By means of computational biology, we have characterized the complete SLA/LP gene, mRNA and deduced protein sequences in man, mouse, zebrafish, fly, and worm.

RESULTS: The human SLA/LP gene sequence of approximately 39 kb, which maps to chromosome 4p15.2, is organized in 11 exons, of which 10 or 11 are translated, depending on the splice variant. Homologous molecules were identified in several biological model organisms. The various homologous protein sequences showed a high degree of similarity or homology, notably at those residues that are of functional importance. The only domain of the human protein sequence that lacks significant homology with homologous sequences is the major antigenic epitope recognized by autoantibodies from autoimmune hepatitis (AIH) patients.

CONCLUSION: The SLA/LP molecule and its functionally relevant residues have been highly conserved throughout the evolution, suggesting an indispensable function of the molecule. The finding that the only non-conserved domain is the dominant antigenic epitope of the human SLA/LP sequence, suggests that SLA/LP autoimmunity is autoantigen-driven rather than being driven by molecular

INTRODUCTION

Autoantibodies to soluble liver antigen (SLA)^[1] and to liver/pancreas (LP)^[2] have been described as specific markers for autoimmune hepatitis^[3]. Recently, identity of the SLA and LP antigens has been demonstrated, and the target antigen of SLA/LP autoantibodies has been cloned^[4]. Using the recombinant SLA/LP molecule, the strict specificity of SLA/LP autoantibodies as markers for AIH has been confirmed^[5].

The primary biological function of SLA/LP remains unclear. Because the SLA/LP molecule was found to be associated with the UGA tRNP(Ser)Sec complex^[6,7], which facilitates the co-translational incorporation of selenocysteine into proteins, it has been speculated that the SLA/LP molecule may have a role in selenoprotein metabolism; the specialized UGA tRNA is initially charged with serine to form seryl-tRNA, which then is enzymatically converted to selenocysteine-tRNA^{Sec}. However, there is no direct experimental evidence for such a role of the SLA/LP molecule so far. Nevertheless, a fold recognition study predicted the SLA/LP tertiary structure by comparison to known protein structures to be that of a pyridoxal phosphate (PLP)-dependent transferase^[8], which is compatible with a role in selenoprotein metabolism. The active site was proposed to be a cavity with a channel, formed by dimerization of two SLA/LP molecules^[8,9]. In the three-dimensional model, five amino acids of monomer A (L88, T89, F92, T94, and S118) as well as two amino acids of monomer B (P251 and G252) were involved in dimerization and the amino acids critical for binding and orientation of the co-enzyme PLP were

identified to be G50, S51, S165, D186, H189, K224^[8] (residue numbering according to GenBank accession number NP_722547).

The immune mechanisms that drive SLA/LP autoimmunity and the pathogenetic relevance of autoimmunity to SLA/LP are currently not clear. The main antigenic region crucial for the recognition of SLA/LP autoantibodies was identified to locate between amino acids 390 and 428^[4,10], thus being easily accessible by antibodies^[8]. This region contains sequence homologies with various microbial antigens, including proteins of Rickettsia species, human herpesvirus 6, and cytomegalovirus^[4]; however, the homologous microbial sequences are only poorly recognized by SLA/LP autoantibodies^[10]. Nevertheless, molecular mimicry of homologous proteins from other species, such as parasites, may be a trigger of SLA/LP autoimmunity.

Due to the human genome project and other genomic sequencing groups, sequences of many different species are now publicly available. While some chromosomes are completely assembled, most genomic fragments are currently still preliminary and await characterization. To approach the elusive biological function of the SLA/LP molecule, we have characterized here the human SLA/LP gene and homologous sequences of other species by means of computational biology. The obtained results support the hypothesis that SLA/LP autoimmunity is autoantigen-driven rather than being driven by molecular mimicry^[10].

MATERIALS AND METHODS

Identification of the human SLA/LP gene

We have previously identified the human SLA/LP protein and deposited the corresponding mRNA sequence into the GenBank^[11] with the accession numbers NM_016955 and NM_153825. BLASTing the human mRNA sequence against the human genome draft sequence revealed the genomic localization^[12,13].

Identification of SLA/LP homologs in other species

Homologs corresponding to the human SLA/LP were identified by BLAST query with the human SLA/LP protein sequence (NP_722547) against the non-redundant databases of mouse (*Mus musculus*), zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*) and the nematode worm (*Caenorhabditis elegans*).

Exon/intron structure and chromosomal localization

Exon/intron structure and chromosomal localization were identified by BLASTing the SLA/LP mRNA sequences of each homolog against the corresponding genome draft sequences of the particular species.

Locus analysis

Data regarding the genomic loci analysis were obtained from the GenBank^[11].

Analysis of protein structure and conservation

The amino acid sequence was deduced from the

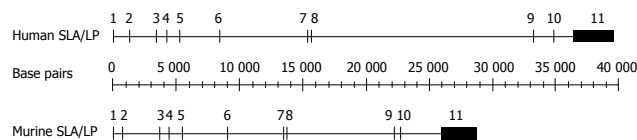


Figure 1 Organization of SLA/LP genes in man and mouse. The exon/intron structure of the human and mouse SLA/LP gene is represented schematically. Exon sequences (numbered boxes) are highly conserved and of similar size and number in man and mouse. Intron sequences, marked as lines, differ in both species.

obtained mRNA sequences. The alignment of SLA/LP homologous sequences was initially performed using the ClustalW algorithm^[14]. The sequence alignments in Figure 3 were performed using the longer human SLA/LP variant (CAB89517). The theoretical isoelectric point and molecular weight was calculated with the pI/Mw tool on the Expasy server^[15,16]. Protein structures and domains were analyzed using two independent algorithms, Pfam^[17] and NCBI Conserved Domain Search^[18].

RESULTS

Human SLA gene and corresponding mRNA and protein sequences

We have previously submitted the human mRNA sequences of two alternative splice variants of SLA/LP to GenBank with the accession numbers NM_016955 and NM_153825. Using the human SLA/LP protein sequence as a query for BLAST searches against the human mRNA database, additional matching mRNA sequences (splice variants) were not identified. Matching the human mRNA sequence to the available human genomic draft sequence (htgs), the corresponding gene was mapped to chromosome 4p15.2. The gene was found to span a genomic region of approximately 39 kb and to be organized in 11 exons (Figure 1). Most exon sizes were rather small, ranging between 91 and 187 bp, with the exception of exon 11 at the 3' end, which covered 3 224 bp. The two splice variants were fully identical, except for the absence of exon 2 (149 bp) in the shorter variant. Both variants have a large 2 930 bp 3' untranslated region; the 5' untranslated region of the two variants is relatively short with 155 and 174 bp, respectively. The amino acid sequences of the two splice variants, as deduced from their mRNA sequence, differ only in their amino-terminal residues (Figure 2). The theoretical molecular weight and isoelectric point of the longer splice variants is 48.84 kDa and 8.64, respectively; that of the shorter variant is 46.85 kDa and 8.52, respectively.

Mouse SLA gene and corresponding mRNA and protein sequences

The SLA/LP gene and genomic organization was highly conserved in mouse. BLASTing the human SLA/LP protein sequence against the mouse protein database, revealed a mouse homolog with the accession number NP_766078. The human and mouse SLA/LP protein sequences are identical at 85% of the residues (Figure 2). Both, the NCBI and Pfam conserved protein domain

Human short	MS - TSYGCFWRR 11	
Mouse short	MNPESF AAGE RR 12	
	* . * . . *	
Human long	MDSNNFLGNCVGEREGRVASALVARRHYR 30	
Mouse long	MDSNNFLGNCVGEREGRVASALVARRHYR 30	

Human SLA/LP	FIHGIGRSGDISAVQPKAAGSSLLNKITNSLVLD I I K L A G V H T V A N C F V V P M A T	65/84
Mouse SLA/LP	FIHGIGRSGDISAVQPKAAGSSLLNKITNSLVLD I I K L A G V H T V A N C F V V P M A T	66/84

Human SLA/LP	GMSLTLCLFLTLRHKRPKAKYIIWPRIDQKSCFKSMI TAGFEPVVIENVLEGDEL	119/138
Mouse SLA/LP	GMSLTLCLFLTLRHKRPKAKYIIWPRIDQKSCFKSMV TAGFEPVVIENVLEGDEL	120/138

Human SLA/LP	RIDLKAVEAKVQELGPDCLIC I H S T T S C F A P R V P D R L E E L A V I C A N Y D I P H I V N	173/192
Mouse SLA/LP	RIDLKAVEAK I Q E L G P E H I L C L H S T T A C F A P R V P D R L E E L A V I C A N Y D I P H V V N	174/192

Human SLA/LP	NAYGVQSSKCMHLIQQGARVGRIDAFVQSLDKNFMPVPGGAIAGFNDSFIQEI	227/246
Mouse SLA/LP	NAYGLQSSKCMHLIQQGARVGRIDAFVQSLDKNFMPVPGGAIAGFNDFIQDI	228/246

Human SLA/LP	SKMYPGRASASPSLDVLTLLSLGCSGYRLLKERKEMFVYLSYLSNQI KKLSEA Y N	281/300
Mouse SLA/LP	SKMYPGRASASPSLDVLTLLSLGCSGYRLLKERKEMFVYLSYLSNQI KKLSEA Y N	282/300

Human SLA/LP	ERLLHTPHNPISLAMTLKTLDEHRDKAVTQLGSMFLTRQVSGARVVPVPLGSMQTV	335/354
Mouse SLA/LP	ERLLQTPHNPISLAMTLKTLDEHRDKAVTQLGSMFLTRQVSGARAVPLGNVQTV	336/354

Human SLA/LP	SGYTFRGFMSHTNNYP CAYLNA A S A I G M K M Q D V D L F I N R D R C L K A V R K E R S K E	389/408
Mouse SLA/LP	SGHTFRGFMSHADNYP CAYLNA A A I G M K M Q D V D L F I K R L D K C L N I V R K E Q T R A	390/408

Human SLA/LP	---SDDNYDK T E D V D I E E M A L K L D N V L L D T Y Q D A S S	422/441
Mouse SLA/LP	SVVSGADRNKAEADIEEMALKLDDVLDVGVQGPAL	426/444

Figure 2 SLA/LP splice variants in man and mouse. The amino acid sequences of the long and short SLA/LP splice variants in humans and their putative homologs in mouse are aligned. The protein variants differ only in their amino-terminal sequence, 11 or 30 residues in the human protein and 12 or 30 residues in the murine protein, as depicted in the upper panel. The other residues (lower panel) are identical in the respective splice variants within each species.

Homo sapiens	---MDSNNFLGNCVGEREGRVASALVARRHYRF I HGIGRSGD IS AVQPKAAGSSLLNK	56
Mus musculus	---MDSNNFLGNCVGEREGRVASALVARRHYRF I HGIGRSGD IS AVQPKAAGSSLLNK	56
Danio rerio	---MDSNNFLGNCVGEREGRVASALVARRHYRL I HGIGRSGD IAAVQPKAAGSSLLNK	56
Drosophila m.	MLASLDSNNYP HKVGLGEREAR I ACKLVARRHYNFGHGIGRSGDLL EAQPKAAGSTLLAR	60
C. elegans	-----MIPV G A G E R E G R V L T P L V Q R L H S N L T H G I G R S G N L L E I Q P K A L G S S M L A C	50
	* * * * * * * * * * * * * * * * * * * *	
Homo sapiens	I T N S L V L D I I K L A G V H T V A N C F V V P M A T G M S L T L C F L T L R H K R P K A K Y I I W P R I D Q K S C F	116
Mus musculus	I T N S L V L N V I K L A G V H S V A S C F V V P M A T G M S L T L C F L T L R H K R P K A K Y I I W P R I D Q K S C F	116
Danio rerio	I T N S V L V D V L K L T G V R S V S C F V V P M A T G M S L T L C F L T L R H R R P K A R Y I L W P R I D Q K S C F	116
Drosophila m.	L T N A L I L D L I R G I G L P S C A G C F L V P M C T G M T L T L C Q S L R K R R P G A R Y V L W S R I D Q K S C F	120
C. elegans	L S N E F A K H A L H L L G L H A V K S C I V V P L C T G M S L S L C M T S W R R R R P K A K Y V V W L R I D Q K S S L	110
	:: * * * * * * * * *	
Homo sapiens	K S M I T A G F E P V V I E N V L E G D E L R T D L K A V E A K V Q E L G P D C I L C H S T T S C F A P R V P D R L E	176
Mus musculus	K S M V T A G F E P V V I E N V L E G D E L R T D L K A V E A K I Q E L G P E H I L C L H S T T A C F A P R V P D R L E	176
Danio rerio	K S M V T A G F E P V V I E N V L E G D E L R T N L E E V E R K I E E F G A E N T L C V H S T T S C F A P R V P D R L E	176
Drosophila m.	K A I T A T G L V P V I P C L I K G E S L N T N V D L F R E K I K S L G V D S I L C L Y T T S C F A P R N S D D I A	180
C. elegans	K S I Y H A G F E P I I V E P I R D R D S L I T D V E T V N R I I E Q R G E E - I L C V M T T S C F A P R S P D N V E	169
	* * * * * * * * *	
Homo sapiens	E L A V I C A N Y D I P H I V N N A Y G V Q S S K C M H L I Q Q G A R V G R I D A F V Q S L D K N F M P V P G G A I I A	236
Mus musculus	E L A V I C A N Y D I P H V N N A Y G L Q S S K C M H L I Q Q G A R V G R I D A F V Q S L D K N F M P V P G G A I I A	236
Danio rerio	E L S V L C A K H D I P H I V N N A Y G V Q P S K C M H L I Q Q G A R V G R I D A F V Q S L D K N F V P V P G G A I I A	236
Drosophila m.	E V S K L S K W Q I P H L V N N A Y G L Q A K E I V N Q L E C A N R V G R I D Y F V Q S S D K N L L V P V G S A I V A	240
C. elegans	A I S A I C A A H D V P H L V N N A Y G L Q S E E T I R K I A A A H E C G R V D A V V Q S L D K N F Q V P V G G A I A	229
	:: * * * * * * * *	
Homo sapiens	G F N D S F I Q E I S K M Y P G R A S A S P S L D V L I T L L S L G S N G Y K L L K E R K E M F S Y L S N Q I K K L S	296
Mus musculus	G F N E P F I Q D I S K M Y P G R A S A S P S L D V L I T L L S L G S G Y R K L L K E R K E M F V Y L S T Q L K K L A	296
Danio rerio	G F D E N F I Q E I S K I Y P G R A S A S P S L D V L I T L L T L G A N G Y K L L A D R K E L Y G H L A Q E L S A L A	296
Drosophila m.	S F N E S V L H D V A S T Y A G R A S G S Q S L D V L M T L L S L G R N G F R L L F D Q R G E N F N Y L R E N L R K F A	300
C. elegans	A F K Q N H I Q S I A Q S Y P G R A S V P S R D L V L T L L Y Q Q S A F L E P F G K Q K Q M F L K M R R K I S F A	289
	. * * * * * * * *	
Homo sapiens	E A Y N E R L L H T P H N P I S L A M T L K T L D E H R D K A V T Q L G S M L F T R Q V S G A R V V P L G S M Q - T V S	355
Mus musculus	E A H N E R L L Q T P H N P I S L A M T L K T I D G H H D K A V T Q L G S M L F T R Q V S G A R A V P L G N V Q - T V S	355
Danio rerio	A R H G E R L L K T P H N P I S L A M S L N H L E A H S S A V T Q L G S M L F T R Q V S G A R V V P L G V Q Q - T V S	355
Drosophila m.	E P R G E I V I D S R F N S I S L A I T L A T L A G D Q M K S I T K L G S M L H M R G V S G A R V I V P G Q N K - T I D	359
C. elegans	E N I G E C V Y E P E N E I S S A M T L S T I P P A K Q - - - T L F G S I L F A K G I T G A R V V T S S Q S K T T I E	346
	* * * * . . . * * * * * * *	
Homo sapiens	G Y T F R G F M S H T N N Y P C A Y L N A A S A I G M K M Q D V D L F I K R L D R C L K A V R K E R S K E - - - S D D N	412
Mus musculus	G H T F R G F M S H A D N Y P C A Y L N A A A I G M K M Q D V D L F I K R L D K C L N I V R K E Q T R A S V S G A D	415
Danio rerio	G H T F S G F M S H S E A Y P C P Y L N A A S A I G I T K G D V I V C M K R L G K C L K I L K E K S D P - - - - A	409
Drosophila m.	G H E F L G K - - - - -	366
C. elegans	G C E F I N F G S H T T E Q H G G Y L N I A C S V G M T D H E L E E L F T R L T S S Y A K F V R E L A K E - - - D E R I	403
	* * * * * * * *	
Homo sapiens	Y D K T E D V D I E E M A L K L D N V L L D T Y Q D A S S	441
Mus musculus	R N K A E D A D I E E M A L K L D D V L D V G V Q G P A L	444
Danio rerio	D L E A E D G E L E S P Q R S T E L R V - - - - -	430
Drosophila m.	- - - - -	
C. elegans	N S S G R R I P I N E S F D M E N D - - - - -	421

Figure 3 Alignment of SLA/LP protein sequences. The SLA/LP amino acid sequences of various eukaryotic species were aligned by ClustalW. The presumed residues that mediate binding to the cofactor pyridoxal phosphate are highlighted in bold and underlined; residues that function in dimerization are highlighted in bold and italic. The immunodominant epitope sequence (390-428) of the human SLA/LP protein is marked in bold.

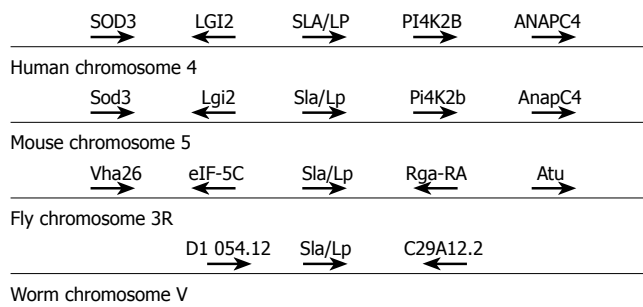


Figure 4 Chromosomal localization of SLA/LP genes. The genes adjacent to the SLA/LP loci on human chromosome 4, mouse chromosome 5, fly chromosome 3R, and worm chromosome V are represented schematically.

algorithms, predicted the existence of SLA/LP structure between amino acids 61 and 458 and amino acids 61 and 459, respectively. BLASTing the protein sequence against the translated mouse genome database revealed the corresponding mRNA NM_172490.1. BLASTing the mRNA sequence against the draft sequence of the mouse genome, the corresponding SLA/LP gene was mapped to mouse chromosome 5qC1. The genomic region spanned 28.5 kb and the mouse SLA/LP gene was organized into 11 exons like the human homolog (Figure 1). Exon sizes ranged between 91 and 2 927 bp with exon 11 being by far the largest exon. The 3' untranslated region was 2 629 bp and the 5' untranslated region was 19 bp. Although not filed in the GenBank, two murine splice variants are very likely; the mouse genomic organization and exon sequences are very similar to the human gene and a splicing signature similar to that in human beings is present in the homologous mouse sequence. The presumed splice variants in the mouse are shown in Figure 2, aligned to the corresponding human proteins. The theoretical molecular weight and isoelectric point of the longer splice variants is 48.48 ku and 8.86, respectively; that of the shorter variant is 46.5 ku and 8.71, respectively.

Zebrafish SLA gene sequence and corresponding mRNA

The SLA gene and genomic organization was highly conserved in fish. BLASTing the human SLA/LP protein sequence against the zebrafish protein database, revealed a homolog with the accession number NP_956448. The zebrafish sequence was identical to the human protein sequence in 69% of the residues. Both, the NCBI and Pfam conserved protein domain algorithms, predicted the existence of SLA/LP structure between amino acids 61 and 459. BLASTing the protein sequence against the translated zebrafish genome database revealed the corresponding mRNA NM_200154. Chromosomal mapping was not possible, since the genomic fragment containing the SLA/LP sequence was not yet assembled into the genomic context. The corresponding gene could only be mapped to the fragment scaffold Zv4_NA8411.1. Thus a complete exon/intron structure was not yet available, and potential splice variants could not be identified.

Drosophila SLA gene sequence and corresponding mRNA

The SLA gene and genomic organization were also

conserved in *Drosophila*. BLASTing the human SLA/LP protein sequence against the *Drosophila* protein database revealed a *Drosophila* homolog (two isoforms) with the accession numbers AAF51994 and AAS65099. The *Drosophila* protein sequence has 42% identity compared to the human protein sequence. However, this may be an underestimation, as it remains unclear whether the recently annotated *Drosophila* SLA/LP sequence is complete. Both, the NCBI and Pfam conserved protein domain algorithms predicted the existence of SLA/LP structure between amino acids 61 and 413 and amino acids 61 and 428, respectively. BLASTing the protein sequence against the translated *Drosophila* genome database revealed the corresponding mRNA NM_141299.1. BLASTing the mRNA sequence against the draft sequence of the fruit fly genome, the corresponding SLA/LP gene was mapped to *Drosophila* chromosome 3R. The genomic region of the currently annotated *Drosophila* SLA/LP gene spanning approximately 1.6 kb was found to be organized into (at least) 4 exons. Exon sizes ranged between 114 and 1 021 bp with exon 3 being by far the largest exon. A short 3' UTR was shown to consist of 115 bp. Thus far, a 5' UTR was not identified.

***C. elegans* SLA/LP gene sequence and corresponding mRNA**

The SLA/LP gene and genomic organization were also conserved in *C. elegans*. BLASTing the human SLA/LP protein sequence against the *C. elegans* protein database revealed similarity to *C. elegans* protein D1054.13 with the accession number CAA98446. The *C. elegans* protein sequence had 42% identity compared to the human protein sequence. Again, both the NCBI and Pfam conserved protein domain algorithms predicted the existence of SLA/LP structure between amino acids 53 and 450. BLASTing the protein sequence against the translated *C. elegans* genome database revealed the corresponding mRNA NM_073360. BLASTing the mRNA sequence against the draft sequence of the *C. elegans* genome, the corresponding SLA/LP gene was mapped to *C. elegans* chromosome V. The genomic region spanned 1.8 kb and the *C. elegans* SLA/LP gene was organized in 7 exons. Exon sizes ranged between 67 bp for exon 3 and 636 bp for exon 5. A rather short 5' UTR contained 43 bp, a 3' UTR could not be found in the currently available sequence. Thus, it remains unclear, whether the currently available sequence represents the complete protein sequence of *C. elegans*.

Conservation of SLA/LP homologs

We aligned the protein sequences of the SLA/LP homologs in human, mouse, fish, fly, and worm (Figure 3). The mammalian sequences were the two closest related sequences. The invertebrate sequences of *Drosophila* and *C. elegans* sequence were most closely related to each other; the zebrafish sequence was equally related to the mammalian and invertebrate sequences.

Chromosomal localization of the human, mouse, fly, and worm SLA/LP genes

We then studied the neighboring genes in man, mouse, fly, and worm (Figure 4). Upstream of the human SLA/

LP sequence on human chromosome 4 was the gene for extracellular-superoxide dismutase 3^[19] and the Leucine rich gene, glioma-inactivated 2^[20]. Downstream of the SLA/LP locus was the phosphatidylinositol 4-kinase type-II beta (PI4K2B)^[21] and anaphase promoting complex, subunit 4^[22]. The same structure was found for mouse chromosome 5 that contains the murine SLA/LP locus. The genomic structure of SLA/LP loci in fly and worm did not share these similarities to the mammalian genomes. The fly SLA/LP homolog was flanked by the gene *eIF-5C*, which is involved in long-term memory^[23] and the gene for a 26 ku E-subunit of the vacuolar ATPase^[24], upstream of the SLA/LP locus. Two currently uncharacterized genes, Rga-RA (account number AAF51992) and Atu (account number AAF51991), are located downstream of the fly SLA/LP gene. The worm SLA/LP sequence was flanked by a gene (5L286) encoding protein of unknown function, upstream, and the serpentine receptor class H (srh-184)^[25], downstream of the SLA/LP locus.

Conservation of the main antigenic region and main structural features of folding and dimerization

The dominant autoepitope of SLA/LP autoantibodies had been identified to locate between amino acids 390 and 428 of the longer human SLA/LP protein sequence variant (NP_722547)^[4,10]. After aligning the homologous protein sequences (Figure 3), we studied the degree of sequence conservation in the immunodominant antigenic region. Human and mouse sequences showed 57% identity and 71% similarity. All other sequences did not display a considerable level of similarity compared to the human epitope sequence.

We then studied the conservation of the residues that seem to function in protein folding and dimerization^[8] (Figure 3). The lysine residue at position 224, which is crucial for covalent binding of the coenzyme PLP (pyridoxal-5'-phosphate) was conserved in all homologous sequences investigated. Of the other five residues directly in contact with PLP, G50, S51, and H189 were identical in all the homologs, whereas S165 was replaced by alanine in mouse and D186 was replaced by glutamine in *Drosophila* and aspartic acid in *C. elegans*. Of the residues of monomer A involved in dimerization, only L88 was conserved in all homologous sequences. T89 was conserved in most species, but replaced by serine in *C. elegans*. F92 was replaced by leucine in *Drosophila* and by methionine in *C. elegans*, and otherwise conserved. T94 was conserved in vertebrates, but replaced by serine in non-vertebrates. S118 was replaced by arginine in *Drosophila*. The contact residues of monomer B, P251 and G252, which mediate dimerization, were conserved in all the species.

DISCUSSION

SLA/LP autoantibodies have been demonstrated to be highly specific markers of autoimmune hepatitis^[4,5]. A role of SLA/LP autoimmunity in the pathogenesis of AIH is likely, at least in a subgroup of AIH patients. However, our current understanding of both, immunogenicity and biological function of the SLA/LP molecule is rather limited. To approach the elusive function and

immunogenicity of SLA/LP, we have characterized the genomic organization and sequence conservation of the human SLA/LP molecule and homologous proteins in other eukaryotic species.

For all homologs, mRNA and genomic sequences were available from the GenBank^[12]. However, in contrast to human and mouse protein and mRNA sequences, which seemed to be complete, some genomic sequences have yet to be completely assembled, and thus did not allow for complete and final characterization of the genomic organization. This was especially the case for fish, where no chromosomal assignment could be made. Nevertheless, genomic characterization in man, mouse, fly, and worm was possible.

The two mammalian SLA/LP genes were similar in organization of the exon/intron structure (Figure 1) and genomic localization (Figure 4); exon/intron structure and genomic localization in fish, fly, and worm was clearly different from the mammalian sequences. Nevertheless, all homologs displayed a high degree of similarity and identity in their amino acid sequences (Figure 3). The similarity between the most distant species - man and worm - was as high as 42%. The highest degree of similarity was found between the human and the mouse amino acid sequences (Figure 2); both species also have a highly similar exon/intron structure in the SLA/LP gene (Figure 1). Moreover, both mammalian species seem to generate similar variant proteins (Figure 2) by differential splicing of exon 2. Whether splice variants exist in non-mammalian species is currently not clear. Previously, the residues, which seem to mediate dimerization and binding of PLP, have been identified by molecular modeling^[8]. These residues are highly conserved among the various species studied (Figure 3). Four of six residues that are crucial for PLP-binding are identical in all the species (G50, S51, H189, K224); the other two residues are homologous (S/A165, D/Q185). Two of the seven residues that mediate dimerization were identical in all the studied species (L88, G252), the other residues were equivalent (T/S89, T/S94, S/A118, P/A251), except for F92, which was conserved in man, mouse and fish, but substituted by leucine in fly and methionine in worm.

The query of human protein databases with the human SLA/LP sequence did not reveal any functional homologs to other proteins and the primary biological function of the SLA/LP protein remains unclear. However, the extraordinary degree of conservation, notably of the presumed functional residues, suggests that the SLA/LP protein has an important biological function that relies on PLP binding. The possible role of the smaller splice variant remains obscure. It may be speculated that the smaller protein variant may act as an inhibitor of the elusive function of the larger variant; the molecular signature of the smaller variant is not compatible with a PLP-dependent transferase, as predicted for the larger variant^[8].

The autoimmune response to SLA/LP is not random, but displays a remarkable degree of uniformity: SLA/LP autoantibodies are of a preferred dominant subtype (IgG1) and recognize the same dominant antigenic region, which was mapped to a carboxy-terminal domain between residues 399 and 428 of the longer splice variant (accession

number NP_722547)^[10]. Thus, SLA/LP autoimmunity has features of a highly selected immune response and seems to be driven by the same mechanism in all the patients. Of note, the antigenic region is located within the only domain of the molecule that was not conserved among the various species (Figure 3). Thus, it appears that SLA/LP autoimmunity is specific for the self-antigen and not for homologous sequences from other eukaryotic species. The possibility that SLA/LP autoimmunity might be driven by homologous proteins from parasites is hence quite unlikely. SLA/LP-homologous proteins are only found in eukaryotes and archaeobacteria, but not in eubacteria^[10]. Nevertheless, a few homologous sequences from bacterial or viral proteins with some degree of similarity to the antigenic epitope of the SLA/LP protein do exist; however, these are not recognized by SLA/LP autoantibodies^[10]. Likewise, the corresponding sequence of an archaeobacterial SLA/LP-homolog is also not recognized by SLA/LP autoantibodies^[10]. Therefore, SLA/LP autoimmunity in patients is very likely driven by the self-SLA/LP molecule rather than by a mechanism that involves molecular mimicry.

ACKNOWLEDGMENTS

Chun-Xia Wang and Andreas Teufel contributed equally. The authors thank A. Schindler for her comments on the manuscript.

REFERENCES

- Manns M, Gerken G, Kyriatsoulis A, Staritz M, Meyer zum Buschenfelde KH. Characterisation of a new subgroup of autoimmune chronic active hepatitis by autoantibodies against a soluble liver antigen. *Lancet* 1987; **1**: 292-294
- Stechemesser E, Klein R, Berg PA. Characterization and clinical relevance of liver-pancreas antibodies in autoimmune hepatitis. *Hepatology* 1993; **18**: 1-9
- Kanzler S, Weidemann C, Gerken G, Lohr HF, Galle PR, Meyer zum Buschenfelde KH, Lohse AW. Clinical significance of autoantibodies to soluble liver antigen in autoimmune hepatitis. *J Hepatol* 1999; **31**: 635-640
- Wies I, Brunner S, Henninger J, Herkel J, Kanzler S, Meyer zum Buschenfelde KH, Lohse AW. Identification of target antigen for SLA/LP autoantibodies in autoimmune hepatitis. *Lancet* 2000; **355**: 1510-1515
- Baeres M, Herkel J, Czaja AJ, Wies I, Kanzler S, Cancado EL, Porta G, Nishioka M, Simon T, Daehrich C, Schlumberger W, Galle PR, Lohse AW. Establishment of standardised SLA/LP immunoassays: specificity for autoimmune hepatitis, worldwide occurrence, and clinical characteristics. *Gut* 2002; **51**: 259-264
- Gelpi C, Sontheimer EJ, Rodriguez-Sanchez JL. Autoantibodies against a serine tRNA-protein complex implicated in co-translational selenocysteine insertion. *Proc Natl Acad Sci U S A* 1992; **89**: 9739-9743
- Costa M, Rodriguez-Sanchez JL, Czaja AJ, Gelpi C. Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP(Ser)Sec complex recognized by autoantibodies from patients with type-1 autoimmune hepatitis. *Clin Exp Immunol* 2000; **121**: 364-374
- Kernebeck T, Lohse AW, Grotzinger J. A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology* 2001; **34**: 230-233
- Scarsdale JN, Kazanina G, Radaev S, Schirch V, Wright HT. Crystal structure of rabbit cytosolic serine hydroxymethyltransferase at 2.8 Å resolution: mechanistic implications. *Biochemistry* 1999; **38**: 8347-8358
- Herkel J, Heidrich B, Nieraad N, Wies I, Rother M, Lohse AW. Fine specificity of autoantibodies to soluble liver antigen and liver/pancreas. *Hepatology* 2002; **35**: 403-408
- GenBank website: <http://www.ncbi.nlm.nih.gov>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403-410
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389-3402
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673-4680
- Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 1993; **14**: 1023-1031
- Bjellqvist B, Basse B, Olsen E, Celis JE. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 1994; **15**: 529-539
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res* 2004; **32**: D138-D141
- Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 2003; **31**: 383-387
- Stern LF, Chapman NH, Wijsman EM, Altherr MR, Rosen DR. Assignment of SOD3 to human chromosome band 4p15.3-->p15.1 with somatic cell and radiation hybrid mapping, linkage mapping, and fluorescent in-situ hybridization. *Cytogenet Genome Res* 2003; **101**: 178
- Nagase T, Kikuno R, Ohara O. Prediction of the coding sequences of unidentified human genes. XXI. The complete sequences of 60 new cDNA clones from brain which code for large proteins. *DNA Res* 2001; **8**: 179-187
- Balla A, Tuymetova G, Barshishat M, Geiszt M, Balla T. Characterization of type II phosphatidylinositol 4-kinase isoforms reveals association of the enzymes with endosomal vesicular compartments. *J Biol Chem* 2002; **277**: 20041-20050
- Yu H, Peters JM, King RW, Page AM, Hieter P, Kirschner MW. Identification of a cullin homology region in a subunit of the anaphase-promoting complex. *Science* 1998; **279**: 1219-1222
- Dubnau J, Chiang AS, Grady L, Barditch J, Gossweiler S, McNeil J, Smith P, Buldoc F, Scott R, Certa U, Broger C, Tully T. The staufen/pumilio pathway is involved in Drosophila long-term memory. *Curr Biol* 2003; **13**: 286-296
- Guo Y, Wang Z, Carter A, Kaiser K, Dow JA. Characterisation of vha26, the Drosophila gene for a 26 kDa E-subunit of the vacuolar ATPase. *Biochim Biophys Acta* 1996; **1283**: 4-9
- Robertson HM. The large srh family of chemoreceptor genes in Caenorhabditis nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res* 2000; **10**: 192-203

S- Editor Guo SY L- Editor Elsevier HK E- Editor Liu WF