

ARTICLE

Molecular and biochemical characterisation of a novel sulphatase gene: Arylsulfatase G (ARSG)

Paola Ferrante¹, Silvia Messali^{1,3}, Germana Meroni¹ and Andrea Ballabio^{*,1,2}

¹Telethon Institute of Genetics and Medicine (TIGEM), Via P. Castellino 111, 80131 Naples, Italy; ²Medical Genetics, II University of Naples, Italy

Molecular analysis has provided important insights into the biochemistry and genetics of the sulphatase family of enzymes. Through bioinformatic searches of the EST database, we have identified a novel gene consisting of 11 exons and encoding a 525 aa protein that shares a high degree of sequence similarity with all sulphatases and in particular with arylsulphatases, hence the tentative name Arylsulfatase G (ARSG). The highest homology is shared with Arylsulfatase A, a lysosomal sulphatase which is mutated in metachromatic leukodystrophy, particularly in the amino-terminal region. The 10 amino acids that form the catalytic site are strongly conserved. The murine homologue of Arylsulfatase G gene product shows 87% identity with the human protein. To test the function of this novel gene we transfected the full-length cDNA in Cos7 cells, and detected an Arylsulfatase G precursor protein of 62 kDa. After glycosylation the precursor is matured in a 70 kDa form, which localises to the endoplasmic reticulum. Northern blot analysis of Arylsulfatase G revealed a ubiquitous expression pattern. We tested the sulphatase activity towards two different artificial substrates 4-methylumbelliferyl (4-MU) sulphate and *p*-nitrocatechol sulphate, but no arylsulphatase activity was detectable. Further studies are needed to characterise the function of Arylsulfatase G, possibly revealing a novel metabolic pathway.

European Journal of Human Genetics (2002) 10, 813–818. doi:10.1038/sj.ejhg.5200887

Keywords: sulphatases; lysosomal storage disorders; multiple sulphatase deficiency

Sulphatases are members of a highly conserved gene family and share extensive sequence homology and a high degree of structural similarity.^{1,2} Sulphated glycosaminoglycans, glycolipids, glycoproteins and hydroxysteroids are hydrolysed by sulphatases, each of which has exquisite specificity towards its individual substrate *in vivo*.^{3,4} A subset of sulphatases, each with very different natural substrates, is active *in vitro* against a common set of small aromatic substrates, hence the name arylsulphatases. This functional correlation reflects a high degree of predicted amino acid sequence similarity along the entire length of the enzymes. This amino acid sequence conservation

strongly suggests that the sulphatases are members of an evolutionarily conserved gene family sharing a common ancestor, which has undergone duplication events in several species.^{1,5} A more recent duplication has generated a cluster of sulphatase genes located on the distal short arm of the human X chromosome, in Xp22.3.⁶ Twelve human sulphatases have been identified and, on the basis of their subcellular localisation, they can be divided into two main categories: those acting at acidic pH sharing a lysosomal localisation, and those with a neutral pH optimum found in the endoplasmic reticulum and Golgi apparatus.² Most of these sulphatases have been identified by linking a clinical condition to a deficient function of one of these enzymes.^{7–9} Additional evidence that all sulphatases are functionally related comes from the study of a rare and intriguing genetic disease named multiple sulphatase deficiency (MSD) in which all known sulphate

*Correspondence: A Ballabio; Fax: +39-081-5609877;
E-mail: ballabio@tigem.it

³Current address: Istituto FIRC di Oncologia Molecolare (IFOM), Via Serio 21, 20134 Milan, Italy

Received 24 April 2002; revised 1 August 2002; accepted 2 August 2002

tases are deficient.² A major breakthrough in the understanding of the function of sulphatases has recently emerged from the observation that sulphatases undergo a common and unique co- or post-translational modification. This occurs in the endoplasmic reticulum and involves the oxidation of a conserved cysteine residue which is required for catalytic activity and is defective in MSD.¹⁰ All these data underline the importance of the sulphatase family and suggest that the identification of additional sulphatase genes creates novel opportunities to study human metabolism and disease mechanisms.

Materials and methods

Plasmids and constructs

ARSG expression vectors were obtained by cloning both the full-length and the coding sequence of ARSG cDNA into *EcoRI*-digested pcDL and *EcoRI/Sal*-digested pmt 21 vectors.¹¹ ARSG (Ala) and ARSG (Pro) (see below) were both cloned in the two different vectors.

Cos7 transfection and protein extract

Cos7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FCS, 100 g penicillin/ml, and 100 g streptomycin/ml, at 37°C in 5% CO₂ atmosphere. Ten micrograms of wild-type ARSG cDNA expression vector constructs were introduced into Cos7 cells, by electroporation using a BioRad Gene Pulser apparatus, and were seeded in 75-cm² flasks. The cells were harvested 24–72 h after transfection and were resuspended in extraction buffer (150 mM NaCl; 100 mM Tris-HCl, pH 7.5; and 1% Triton X-100) and cell lysates were clarified by centrifugation at 13 000 rpm to remove cell debris. The supernatant was mixed with SDS-PAGE sample buffer, boiled, and loaded on the gel.

Production of polyclonal antibody

The region of amino acids 389–525 of ARSG was fused to the six-histidine tag bacterial expression vector pQE (Qiagen) and was produced in DH5 α cells, after induction with isopropyl- β -D-thiogalactoside (IPTG). The His-tagged protein was purified on a NiNTA agarose column (Qiagen) and was used to immunise rabbits. ARSG antiserum was precipitated with ammonium sulphate and the antibody was purified on a protein A-Sepharose column.

Immunoblot

Ten micrograms of soluble cellular protein was boiled for 5 min in sample buffer, electrophoresed through 10% SDS-PAGE, and electroblotted onto PVDF membrane (Amersham). The membrane was treated with 5% dry milk in Tris-buffer saline with the addition of Tween 20 (TTBS (20 mM Tris-HCl, pH 7; 50 mM NaCl; and 0.1% Tween 20)) to inhibit nonspecific binding. Anti-ARSG antibody was used at a 1:700 dilution in TTBS. Visualisation of antibody binding was performed with a secondary anti-rabbit IgG antibody conjugated with peroxidase.

Immunofluorescence

Immunofluorescence was performed on paraformaldehyde (PFA)-fixed transfected Cos7 cells. Cells were permeabilised with 0.2% Triton X-100, blocked with porcine serum, and incubated with anti-ARSG antibody (1:700). Staining was obtained after incubation with secondary fluorescein isothiocyanate (FITC)- or tetramethylrhodamine isothiocyanate (TRITC)-conjugated isotype-specific antibody. Specific ER staining was obtained even by double staining with a commercially available anti-ER polyclonal antibody, ERAB (K-20) (Santa Cruz Inc). This antibody recognises a peptide mapping near the carboxy terminus of ERAB (endoplasmic reticulum amyloid beta-peptide-binding protein.)

Endoglycosidase H and F treatment

For glycosylation analysis, protein samples derived from transfected Cos7 cells were digested overnight with endoglycosidase H and N-glycosidase F (Boehringer Mannheim) in an appropriate buffer (50 mM phosphate buffer, pH 7.4; 50 mM EDTA; 1% NP40; 0.1% SDS; 1% β -mercaptoethanol; 4 μ g of aprotinin per ml; 2 μ g of leupeptin per ml; 2 μ g of pepstatin per ml) and were analysed by SDS-PAGE.

Enzyme assay

4-MU sulphate assay Cos7 cells were transfected with 10 μ g of either ARSG (Ala) or ARSG (Pro) cDNA constructs by electroporation and harvested by trypsinisation 72 h after transfection for enzyme assay. Cells were resuspended in 0.1 M Tris-HCl (pH 7.5), 1% Triton X-100, and 150 mM NaCl. The incubation mixture was prepared with different pH buffer, from pH 4 to 9. The enzymatic assay was run at 37°C for 2 h, the reaction was stopped adding 1 ml of glycine-carbonate buffer (pH 10.7) and fluorescence was determined on Hoefer TKO 100 fluorometer.

P-nitrocatechol sulphate assay Cos7 cells were transfected with 10 μ g of ARSG cDNA constructs by electroporation and harvested by trypsinisation 72 h after transfection for enzyme assay. Cells were resuspended in 0.1 M Tris-HCl (pH 7.5), 1% Triton X-100, and 150 mM NaCl. The incubation mixture was prepared with two different pH buffer, pH 4.9 and 7.5. The enzymatic assay was run at 37°C for 2 h, the reaction was stopped adding 700 μ l of NaOH (0.64 N) and absorbance was determined at 515 nm on spectrophotometer.

mRNA analysis

Northern blots containing human RNAs from several tissues were purchased from Clontech and hybridised with a cDNA fragment of 885 bp as a probe, using the conditions recommended by the manufacturer (Clontech). Washing conditions were 2 \times SSC, 0.05% SDS at 58°C (Wash Solution I) and 0.1 \times SSC, 0.1% SDS at room temperature (Wash Solution II). RT-PCR experiments were performed with oligonucleotide primers from ARSG that amplify a 360 bp product on samples from different murine tissues. PCR

was carried out for 40 cycles. Annealing temperature was 60°C for ARSG primers.

Results

Identification of the ARSG gene

Bioinformatic searches of the EST and nr (non-redundant) databases identified a novel cDNA sequence that shares a high degree of homology with all sulphatases and in particular with arylsulphatases, hence the name of Arylsulfatase G (ARSG). This gene encodes a 525 amino-acid protein that shows the highest level of homology (37% identity and 50% similarity) with Arylsulfatase A (Figure 1). Sequence alignment between this novel sulphatase with ARSA and ARSB revealed a high degree of amino-acid similarity along

the entire length of the protein, particularly in the amino-terminal region.^{12,13} Moreover, the 10 residues that form the catalytic site of the protein are strongly conserved (Figure 1). ARSG cDNA (KIAA 1001) was generated by the large transcript identification project of the Kazusa Institute, Japan. The availability of mapped ESTs and of genomic sequences corresponding to the ARSG transcripts allowed us to map this gene to 17q23-q24 and to define the intron-exon boundaries of the 11 exons that compose the gene (Table 1). The gene spans the genomic region from 72 287 512 to 72 383 112 (95 600 bp). Comparison among the sequence of the KIAA 1001 clone with the available database sequences revealed a nucleotide change. The 501 codon GCA (Ala) in the genomic and EST sequences

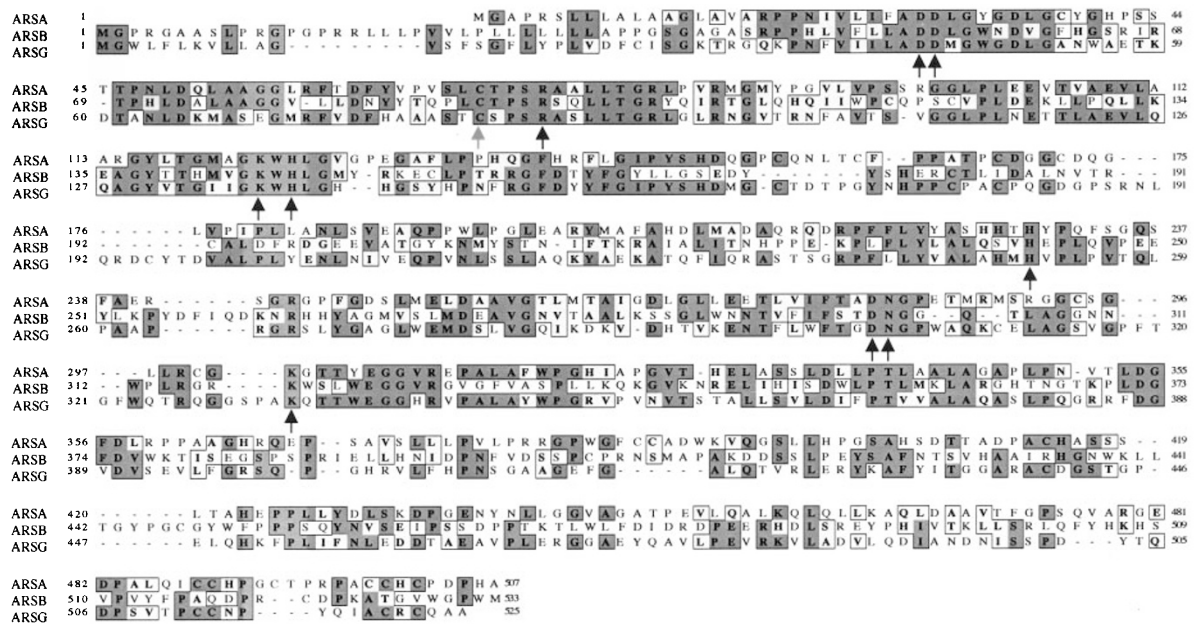


Figure 1 Multiple alignment of human arylsulphatases (A, B, G). Dark grey boxing indicates residues that are identical in at least two out of the three proteins; light grey indicates residues that are conserved in at least two out of the three proteins. Amino acids that are conserved in the catalytic pocket of all sulphatases are indicated with black arrows. The grey arrow indicates the conserved cysteine, contained in the consensus signature that undergoes conversion to 2-amino-3-oxopropionic acid in all eukaryotic sulphatases.

Table 1 Exon-intron boundaries of ARSG gene

Exon	Size (bp)	cDNA position	5' splice site	3' splice site	
1	218	457–674	GGGAATGAG	gtgagtctt	tcctctcag
2	188	675–862	GGATAATAG	gtaactctg	tttctacag
3	48	863–910	ACTTCCGTG	gtaagaatt	ctattccag
4	112	911–1022	ACCATCAAG	gtaatgctg	tccacacag
5	138	1023–1160	GCGTGCAAG	gtgaggagt	ctgtttcag
6	197	1161–1357	GGTTTACAG	gtaaagtag	ggtttctag
7	81	1358–1438	CTCGTCAAG	gtaagggc	ctcdtcacag
8	109	1439–1547	CTTGTTAAG	gtatgagac	tttccgcag
9	121	1548–1668	GGGCACAGG	gttagtggg	ctttctcag
10	91	1669–1759	ACATTACCG	gtgagtggg	cctctctag
11	275	1760–2034			

(BG994996-BG744786-BE740712-AV654375-BM708401) is substituted by CCA (Pro) in the KIAA1001 clone. To test if this mutation affects the localisation and/or the activity of the protein we mutagenised the KIAA 1001 clone. All the experiments described below have been performed with both the original (ARSG Ala) and the mutated (ARSG Pro) clones and no different results were obtained. Moreover, through protein–protein blast (blastp) searches, we have identified the ARSG murine orthologue that shows 87% of identity with the human protein. In agreement with the human genomic and EST sequences, the mouse ARSG presents an alanine in the position corresponding to the human 501 residue.

Characterisation of ARSG in transfected Cos7 cells

We raised polyclonal antibody against the C-terminal portion of the ARSG protein. The region of amino acids 389–525 was fused to a six-histidine tag bacterial expression vector and was then expressed in *DH5 α* cells and used to immunise rabbits. The antiserum was raised against a region where the sulphatases share the lowest degree of homology to prevent cross-reaction with other members of the family. This antiserum (anti-ARSG Q57) was used in Western blotting analysis of protein extracts from Cos7 cells transfected with tagged and non-tagged vectors, and a band of 70 kDa was identified (Figure 2). The amino-acid sequence of the predicted ARSG protein reveals the presence of four putative N-glycosylation sites (asparagine residues 117, 215, 356, and 497) (Figure 2). Treatment with endoglycosidase H and F reduced the size of the 70 kDa polypeptide by 8 kDa, suggesting that the increase in molecular size observed in the maturation of ARSG is due to N-glycosylation (Figure 3). Assuming that the average mass of an oligosaccharide is 2 kDa, all four N-glycosylation sites maybe utilised, accounting for the 8 kDa shift observed in SDS–PAGE.

Intracellular localisation of ARSG

Several sulphatases identified thus far are lysosomal proteins. Non lysosomal sulphatases include STS, which has a microsomal localisation, ARSD and ARSF which are located in the endoplasmic reticulum and ARSE which is located in the Golgi apparatus.^{1,2} The subcellular localisation of ARSG (Ala) in transfected Cos7 cells was established by immunofluorescence techniques using a *myc*-tagged construct, recognised by anti-*myc* antibody. The use of a specific anti-ARSG antibody confirmed this result and revealed a typical reticular distribution of the protein. The reticular localisation was also confirmed by double staining with a specific anti-ER antibody, ERAB (K-20) (Figure 4). The same results were obtained with the ARSG (Pro) (data not shown).

Enzymatic analysis

To test a putative arylsulphatase activity of the protein, ARSG (Ala) and ARSG (Pro) cDNA constructs were transiently transfected into Cos7 cells by electroporation. After

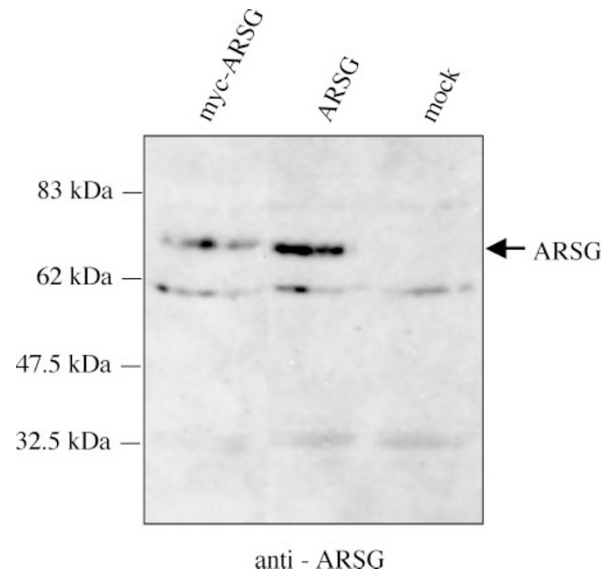


Figure 2 Western blotting analysis: soluble cellular proteins from cells extracts were analysed by SDS–PAGE followed by immunoblot with anti-ARSG antibody. A band of 70 kDa was detectable in Cos7 cells transfected with *myc*-ARSG (lane 1) and with ARSG (lane 2).

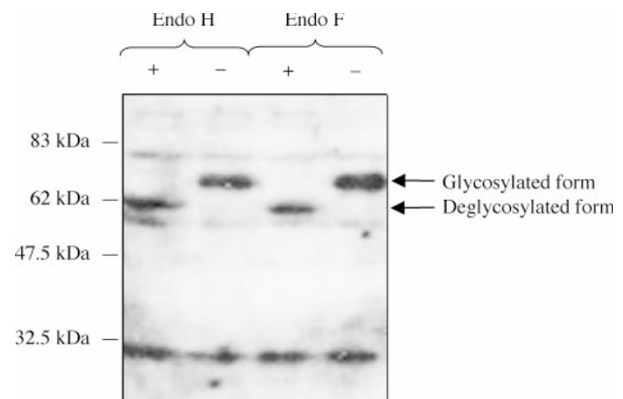


Figure 3 Glycosylation of the ARSG gene product. Cos7 cells were transfected with *myc*-ARSG, cell extracts were subjected to endoglycosidase H and N-glycosidase F treatment and were analysed by SDS–PAGE, together with the untreated extracts. After endoglycosidase treatment, the 70 kDa polypeptide was converted into the 62 kDa form.

72 h the total cell extracts were assayed at different pH (pH range from pH 4 to 9) for their ability to hydrolyse the sulphate group from 4-methylumbelliferyl (4-MU) sulphate. As control we used Cos7 cells transfected with ARSE cDNA construct.¹¹ No arylsulphatase activity was detectable in the extracts obtained from cells transfected with the constructs (Figure 6). These results indicate that in these conditions the ARSG gene product is unable to hydrolyse the 4 MU sulphate. The enzymatic assays were

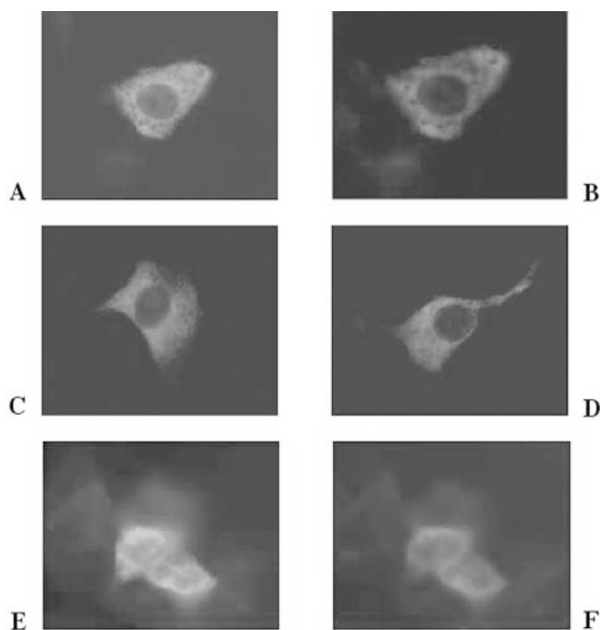


Figure 4 Localisation of *ARSG* in transfected cells. Cos 7 cells were transfected with myc-*ARSG* (A, B) and with *ARSG* (C–F). (A, B) Staining using anti-Myc antibody, followed by FITC anti-mouse antibody, and anti-*ARSG* antiserum, followed by TRITC anti-rabbit antibody, respectively. (C, D) Staining using purified anti-*ARSG* antibody, followed by FITC anti-rabbit antibody. (E, F) Double staining immunofluorescence using anti-*ARSG* antibody, followed by FITC anti-rabbit antibody, and anti-ER antibody, followed by TRITC anti-goat antibody respectively.

also performed using the *p*-nitrocatechol sulphate, another artificial substrate. We obtained the same result with no activity toward this substrate (data not shown).

ARSG expression pattern

Northern blot analysis revealed a low level and ubiquitous expression pattern of human *ARSG* mRNA. Pancreas, kidney and brain showed slightly higher levels of expression (data not shown). Detection of murine *ARSG* mRNA by RT-PCR allowed us to investigate the expression pattern of this gene in mouse. Reverse-transcribed RNA samples from different tissues were amplified by PCR using oligonucleotide primers generated from the murine cDNA sequence. A ubiquitous expression pattern was observed similarly to what we detected in human tissues (Figure 5). In addition to RT-PCR and Northern blot analysis, we performed RNA *in situ* hybridisation experiments. Analysis of sagittal and coronal sections of E12.5 and E14.5 mouse embryos revealed a low level and ubiquitous expression pattern confirming the Northern blotting and RT-PCR results (data not shown).

Discussion

In most instances the identification of the sulphatase genes resulted from the biochemical characterisation of a known

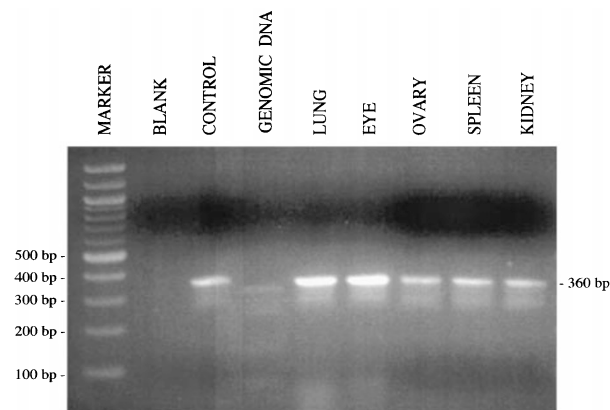


Figure 5 RT-PCR analysis of murine *ARSG* gene. Oligonucleotide primers from murine *ARSG* cDNA were used for the amplification of reverse-transcribed total RNA from different mouse tissues. *ARSG* primers amplify a 360 bp product.

human disease.^{7–9} More recently, we have identified a cluster of sulphatase genes by positional cloning^{11,14}. Here we describe the identification of a novel sulphatase gene, *ARSG*, by a bioinformatic approach using the DBWatcher tool. This allows a periodic and systematic screening of the expressed sequence and genomic databases using sequences from genes or proteins of interest as queries. This strategy does not rely on the knowledge of the gene product or of its substrate specificity and enzymatic properties. In addition no genetic disease related to an enzymatic defect links to the region where the gene maps, consequently it is difficult to predict the function of the protein. *ARSG* protein shares a high degree of homology with all sulphatases and in particular with arylsulphatases. Sequence alignment of the putative protein product of this new gene with other sulphatases, in particular with ARSA and ARSB, has revealed a high degree of amino acid similarity. The three sequences are very similar along the entire length of the protein, particularly in the amino-terminal region, where the consensus sulphatase signature is located (Figure 1). The active site of sulphatases has been characterised and shown to display unique features.^{12,13} A modified cysteine residue and a metal ion are located at the base of a substrate-binding pocket.¹⁰ The amino-acid residues conserved throughout the sulphatase family play a role in stabilising the calcium ion and the sulphate ester in the active site.¹⁵ To try to understand the physiological role of this new gene we started a biochemical characterisation of the *ARSG* protein. We detected two molecular forms of *ARSG*, a 62 and 70 kDa protein, in cell extracts after transfection. The 62 kDa polypeptide is converted into a 70 kDa, which is likely to represent the mature enzyme, with an increase in size due to N-glycosylation. The amino acid sequence contains four potential N-glycosylation sites (asparagine residues 117, 215, 356, and 497). Assuming that the average mass of an oligosaccharide is 2 kDa, all four N-glycosylation sites are potentially utilised, accounting for the 8 kDa shift observed in SDS-PAGE. ARSA

Enzyme Assay

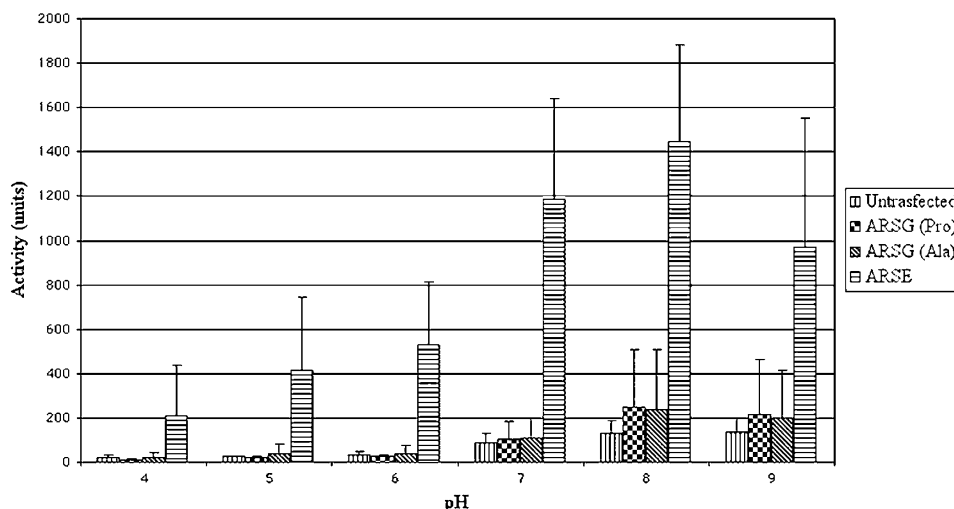


Figure 6 Biochemical study of ARSG expressed in COS7 cells. Evaluation of ARSG arylsulphatase activity on 4-MUS at different pH values.

and ARSB are able to hydrolyse sulphated artificial substrates containing a phenolic ring (arylsulphatase activity), such as p-nitrocatechol sulphate or 4-methylumbelliferyl sulphate (4-MU sulphate) and are lysosomal in their subcellular localisation.¹ Surprisingly, ARSG does not show activity towards these substrates under the specified conditions and localises in the endoplasmic reticulum. Similarly to ARSG, a previous identified sulphatase, ARSD, is highly homologous to other arylsulphatases but does not show arylsulphatase activity.¹⁴ In conclusion, the identification of this additional sulphatase gene creates novel opportunities to study human metabolic pathways. Further studies are needed to characterise the biological role of ARSG. To this purpose we are planning to generate mice deficient for this enzyme by a gene targeting approach. This strategy will allow us to study, in the animal model, the pathological consequences of the enzyme deficiency and to clarify the metabolic role of this new sulphatase.

Acknowledgements

This work was supported by the Italian Telethon Foundation and by the European Community (grant no. QLRT-1999-00793 to A Ballabio).

References

- Parenti G, Meroni G, Ballabio A: The sulfatase gene family. *Curr Opin Genet Dev* 1997; 7: 386–391.
- Hopwood JJ, Ballabio A (ed): Multiple sulfatase deficiency and the nature of the sulfatase family; in Scriver CR, Beaudet AL (eds): *The metabolic and molecular basis of inherited disease*. New York: McGraw-Hill, 2001; 3, pp 3725–3732.
- Bergner E, Shapiro L: Metabolism of 3H-dehydroepiandrosterone sulphate by subjects with steroid sulphatase deficiency. *J Inherit Metab Dis* 1988; 11: 403–415.
- Roy AB (ed): Enzymological aspects of steroids conjugation; in Bernstein S, Solomon S (eds): *Chemical and Biological Aspects of Steroid Conjugation*. New York: McGraw-Hill, 1970, p 74.
- Peters C, Schmidt B, Rommerskirch W *et al*: Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B. *J Biol Chem* 1990; 265: 3374–3381.
- Meroni G, Franco B, Archidiacono N *et al*: Characterisation of a cluster of sulfatase genes on Xp22.3 suggests gene duplications in an ancestral pseudoautosomal region. *Hum Mol Genet* 1996; 5: 423–431.
- von Figura K, Gieselmann V, Jaeken J (ed): Metachromatic leukodystrophy; in Scriver CR, Beaudet AL (eds): *The metabolic and molecular basis of inherited disease*. New York: McGraw-Hill, 2001; 3, pp 3695–3724.
- Ballabio A, Shapiro LJ (ed): Steroid sulfatase deficiency and X-linked ichthyosis; in Scriver CR, Beaudet AL (eds): *The metabolic and molecular basis of inherited disease*. New York: McGraw-Hill, 2001; 3, pp 4241–4262.
- Neufeld EF, Muenzer J (ed): The mucopolysaccharidoses; in Scriver CR, Beaudet AL (eds): *The metabolic and molecular basis of inherited disease*. New York: McGraw-Hill, 2001; 3, pp 3421–3452.
- Schmidt B, Selmer T, Ingendoh A, von Figura K: A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* 1995; 82: 271–278.
- Franco B, Meroni G, Parenti G *et al*: A cluster of sulfatase genes on Xp22.3: mutations in chondrodysplasia punctata (CDPX) and implications for warfarin embryopathy. *Cell* 1995; 81: 15–25.
- Bond CS, Clements PR, Ashby SJ *et al*: Structure of a human lysosomal sulfatase. *Structure* 1997; 5: 277–289.
- Lukatela G, Krauss N, Theis K *et al*: Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* 1998; 37: 3654–3664.
- Puca AA, Zollo M, Repetto M *et al*: Identification by shotgun sequencing, genomic organization, and functional analysis of a fourth arylsulfatase gene (ARSF) from the Xp22.3 region. *Genomics* 1997; 42: 192–199.
- Selmer T, Hallmann A, Schmidt B, Sumper M, von Figura K: The evolutionary conservation of a novel protein modification, the conversion of cysteine to serine-semialdehyde in arylsulfatase from *Volvox carterii*. *Eur J Biochem* 1996; 238: 341–345.