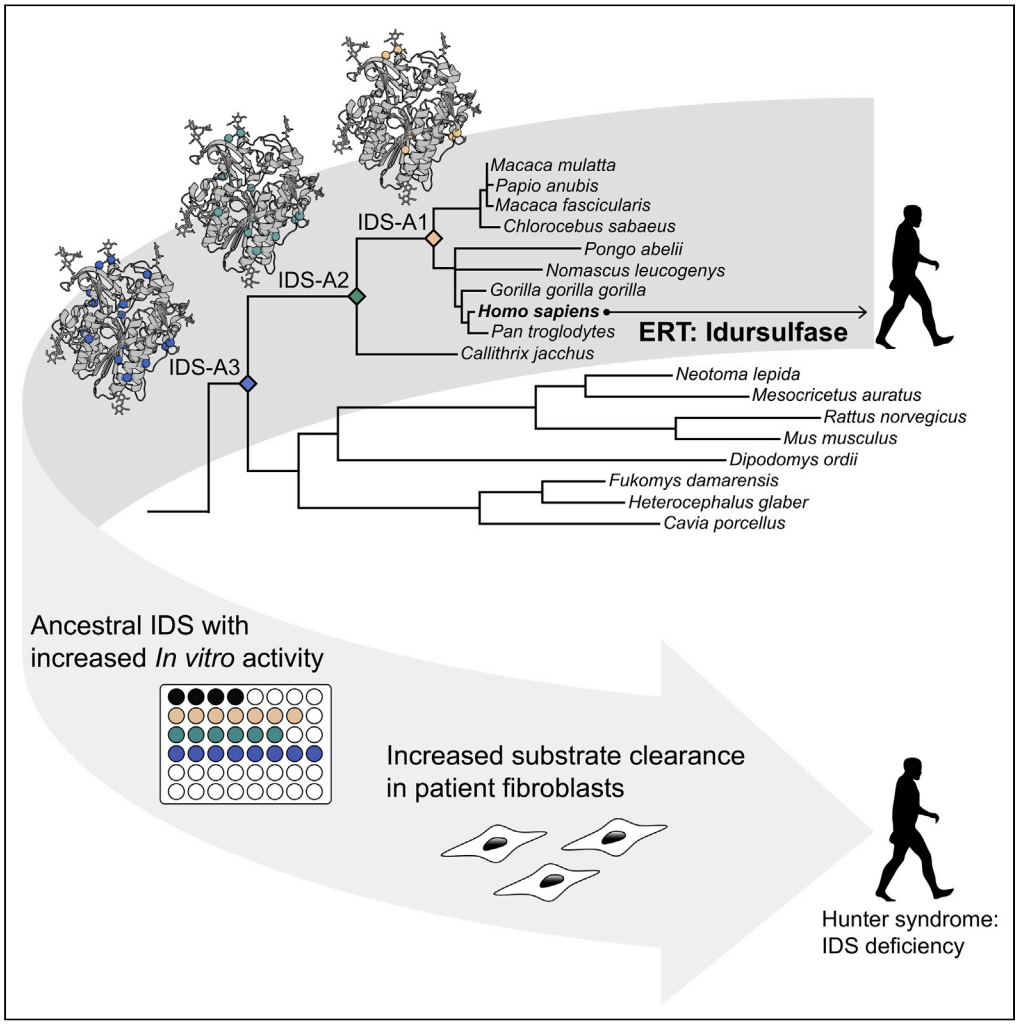
iScience

CellPress
OPEN ACCESS

Article

Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome



Natalie M. Hendrikse, Anna Sandegren, Tommy Andersson, ..., Stefan Svensson Gelius, Per-Olof Syrén, Erik Nordling

per-olof.syren@biotech.kth.se (P.-O.S.)
erik.nordling@sobi.com (E.N.)

HIGHLIGHTS

Reconstruction of ancestral lysosomal enzymes that function in complex cellular context

Ancestral iduronate-2-sulfatases with increased activity compared with the human enzyme

Increased clearance of substrate in patient fibroblasts indicates therapeutic potential

Hendrikse et al., iScience *24*, 102154
March 19, 2021
https://doi.org/10.1016/j.isci.2021.102154

iScience

CellPress
OPEN ACCESS

Article

Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome

Natalie M. Hendrikse,[1,2,3] Anna Sandegren,[1] Tommy Andersson,[1] Jenny Blomqvist,[1] Åsa Makower,[1] Dominik Possner,[1] Chao Su,[1] Niklas Thalén,[4] Agneta Tjernberg,[1] Ulrica Westermark,[1] Johan Rockberg,[4] Stefan Svensson Gelius,[1] Per-Olof Syrén,[2,3,4,*] and Erik Nordling[1,5,*]

SUMMARY

We show the successful application of ancestral sequence reconstruction to enhance the activity of iduronate-2-sulfatase (IDS), thereby increasing its therapeutic potential for the treatment of Hunter syndrome—a lysosomal storage disease caused by impaired function of IDS. Current treatment, enzyme replacement therapy with recombinant human IDS, does not alleviate all symptoms, and an unmet medical need remains. We reconstructed putative ancestral sequences of mammalian IDS and compared them with extant IDS. Some ancestral variants displayed up to 2-fold higher activity than human IDS in *in vitro* assays and cleared more substrate in *ex vivo* experiments in patient fibroblasts. This could potentially allow for lower dosage or enhanced therapeutic effect in enzyme replacement therapy, thereby improving treatment outcomes and cost efficiency, as well as reducing treatment burden. In summary, we showed that ancestral sequence reconstruction can be applied to lysosomal enzymes that function in concert with modern enzymes and receptors in cells.

INTRODUCTION

Lysosomal storage diseases (LSDs) are a group of more than 50 known hereditary disorders that are associated with impaired or depleted function of one or several lysosomal enzymes (Boustany, 2013). The most common treatment today is enzyme replacement therapy (ERT). As of 2020, ERT has been approved by the Food and Drug Administration for nine LSDs (Chen et al., 2019), but dosage and distribution remain challenging and do often not alleviate all symptoms. Required doses are generally high, resulting in lengthy intravenous administration and costly production of the drugs (Desnick and Schuchman, 2012; Grabowski and Whitley, 2017). Increasing stability and activity of therapeutic enzymes is therefore of great interest, as both properties are directly proportional to the required dose to achieve the desired therapeutic effect. To advance ERT for LSDs, the design of novel enzyme variants with improved properties is actively investigated. Previously, the catalytic efficiency of arylsulfatase A—the enzyme associated with LSD metachromatic leukodystrophy—was enhanced up to 5-fold by just a few amino acid substitutions that occurred in the evolution of non-human mammals (Simonis et al., 2019), inspired by the observation that arylsulfatase A from mouse displayed higher activity than its human homolog. Phylogenetic analyses have also been used to study the evolution of the phosphatase superfamily, which also includes arylsulfatases (Van Loo et al., 2019). These findings encouraged us to explore ancestral sequence reconstruction (ASR) (Gumulya and Gillam, 2017; Merkl and Sterner, 2016; Risso and Sanchez-Ruiz, 2017; Thornton, 2004; Wilding et al., 2017) to engineer lysosomal sulfatases. ASR is a bioinformatics method that has yielded valuable insight into enzyme evolution (Nguyen et al., 2016; Risso et al., 2013) and protein structure-function relationships (Nicoll et al., 2020; Schupfner et al., 2020; Wilson et al., 2015), but has rarely been applied in the context of biopharmaceuticals. One exception is the engineering of coagulation factor VIII with improved expression rates and decreased decay rates, allowing for lower doses in gene therapy experiments toward treatment of hemophilia A (Samelson-Jones and Arruda, 2019; Zakas et al., 2017). We have previously applied ASR toward improving therapeutic properties of phenylalanine/tyrosine ammonia lyases and obtained enzyme scaffolds with increased thermal and long-term stability as well as altered substrate specificities (Hendrikse et al., 2020).

[1]Swedish Orphan Biovitrum AB, Stockholm 112 76, Sweden

[2]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Solna 171 21, Sweden

[3]Department of Fibre and Polymer Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm 100 44, Sweden

[4]Department of Protein Science, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm 10691, Sweden

[5]Lead contact

*Correspondence: per-olof.syren@biotech.kth.se (P.-O.S.), erik.nordling@sobi.com (E.N.)

https://doi.org/10.1016/j.isci.2021.102154

We aimed to explore the application of ASR to engineer lysosomal enzymes toward treatment of LSDs—a therapeutic area with a remaining significant unmet medical need (Desnick and Schuchman, 2012). To this end we chose to apply the method to iduronate-2-sulfatase (IDS, EC 3.1.6.13), a monomeric enzyme that catalyzes the first step of the lysosomal breakdown pathway of glycosaminoglycans (GAGs) heparan sulfate and dermatan sulfate (Figure S1A). Hunter syndrome, or mucopolysaccharidosis type II (MPS II, OMIM 309900), is an X-linked genetic LSD caused by impaired function of IDS, which leads to accumulation of these GAGs in lysosomes of tissues and organs, thereby causing cellular dysfunction and organ failure (Bach et al., 1973; Hunter, 1917; Wilson et al., 1990). To be catalytically active, IDS requires N-linked glycosylation and post-translational modification of a cysteine in the active site to formylglycine (fGly). The latter is performed by the formylglycine-generating enzyme (FGE, EC 1.8.3.7) (Von Figura et al., 1998; Schmidt et al., 1995), which converts the cysteine side chain (-CH_2-SH) into an aldehyde group (-CH=O). Crystallization efforts revealed the likely presence of a calcium ion next to fGly in the active site, which is believed to stabilize the sulfate-ester formation and has also been observed in other human sulfatases (Demydchuk et al., 2017).

Over 500 pathogenic mutations in the *IDS* gene have been reported, distributed over more than 100 positions in the enzyme sequence, and structural information has provided valuable insight into the underlying disease-causing mechanisms (Demydchuk et al., 2017). Approximately two-thirds of patients with MPS II are affected by progressive deterioration of the central nervous system (CNS) associated with neuronal degradation, and patients with this severe, early-onset form of the disease seldom reach adulthood (Whiteman and Kimura, 2017). The current option for treatment is ERT with the recombinant human enzyme (Chen et al., 2019; Whiteman and Kimura, 2017), which is available in two different forms: idursulfase (Elaprase, Shire Pharmaceuticals/Takeda Pharmaceutical Company Ltd., approved for therapeutic usage in the United States in 2006 [Heartlein and Kimura, 2014; Muenzer et al., 2006]) and idursulfase beta [Hunterase, Green Cross Corporation, approved for therapeutic usage in South Korea in 2012 [Kim et al., 2017; Sohn et al., 2013]). Both enzymes are effective in treating somatic symptoms of the disease and reducing urinary excretion of GAGs (Kim et al., 2017), but have no documented impact on CNS manifestations as they do not achieve therapeutic levels of activity in the brain.

We selected human IDS (hIDS) and murine IDS (mIDS) as reference enzymes and compared them with several ancestral IDS enzymes, going back from the hIDS sequence to the last common ancestor of primates and rodents. Owing to the high degree of conservation of IDS in mammals, the oldest ancestral variant contained only 20 substitutions when compared with the extant hIDS. We found that some ancestral enzymes displayed increased activity *in vitro* when compared with hIDS, mIDS, and idursulfase, which was also evident in the fibroblasts of patients with MPS II. This increased activity may allow for lower dosage or enable enzyme levels that achieve a therapeutic effect in the brain. We showed that ASR can be successfully applied to enzymes that function intracellularly and in concert with modern enzymes and receptors, which may open up doors for various *in vivo* applications.

RESULTS

Reconstructed ancestral IDS enzymes are highly conserved and can be functionally expressed

The human IDS sequence was used as a query for constructing a phylogenetic tree (Figure 1, a complete sequence alignment and full tree with accession numbers and bootstrap values can be found in Figures S2A and S2B, respectively). We used maximum likelihood statistics implemented in PAML (Yang, 1997, 2007) to reconstruct the most likely ancestral sequences at three nodes between hIDS and the last common ancestor of primates and rodents. Owing to the high degree of conservation of IDS in mammals, the putative ancestral sequences had the same length as hIDS and only 6, 12, and 20 amino acid substitutions were predicted for the IDS-A1, IDS-A2, and IDS-A3, respectively (Figure S3). All positions at which different residues were predicted in the ancestors (referred to as *ancestral mutations*) are highlighted on the structure of hIDS in Figure 2. None of the predicted mutations occurred in one of the 128 positions where disease-causing mutations have been reported (Demydchuk et al., 2017). Inspection of the reconstructed sequences showed that cysteine 84, which is post-translationally modified to fGly, as well as all known catalytic residues were conserved (Figure S3). All glycosylation sites were also conserved, apart from a threonine to serine mutation in the motif around N144 for IDS-A2 and IDS-A3. Only two changes occurred relatively close to catalytic residues (see close-up in Figure 2): mutation A354T in IDS-A3 is located at 4.9 Å from K347 and mutation H356R in IDS-A2 and IDS-A3 is located 9.5 Å from D334. Owing to their proximity to the active site and each other we decided to investigate their individual and combined effects, and variants

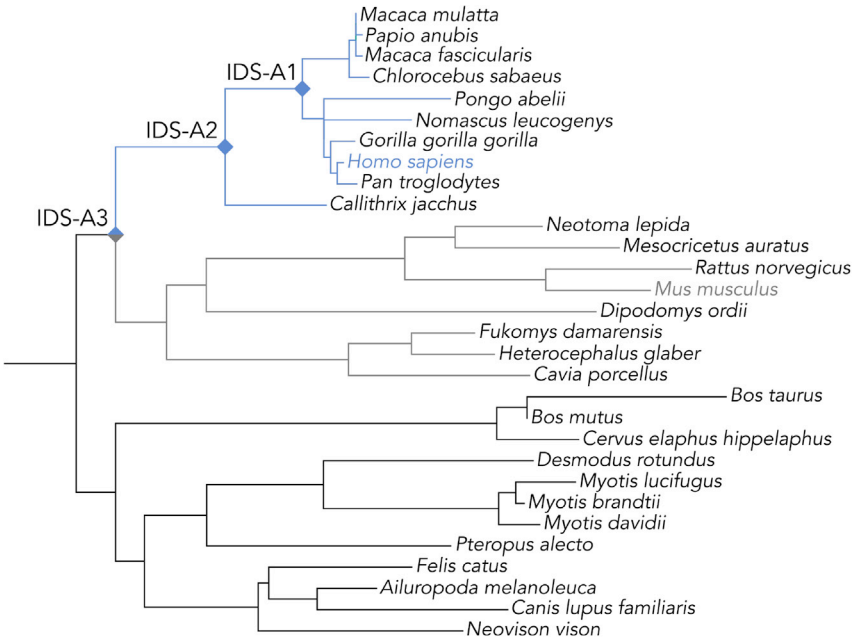


Figure 1. Phylogenetic tree of mammalian IDS homologs

Maximum likelihood tree of 30 mammalian IDS sequences. The colored clades are primates (blue) and rodents (gray), and reference enzymes hIDS and mIDS are colored accordingly. The full sequence alignment and phylogenetic tree can be found in Figure S2.

hIDS_A354T, hIDS_H356R, and hIDS_A354T_H356R were included in the study. Moreover, close to equal probabilities were predicted for valine and isoleucine at position 329 in IDS-A1 and IDS-A3, and both variants were included for each ancestor. As probabilities slightly favored valine, those sequences were denoted IDS-A1 and IDS-A3, and the corresponding variants, IDS-A1_V329I and IDS-A3_V329I. An overview of all variants included in the study and their respective number of mutations can be found in Table S1. All enzymes could be successfully expressed in ExpiCHO cells (yield of 30–140 mg/L cell culture) and were purified by affinity chromatography and size exclusion chromatography (Figure S4).

Two ancestral enzymes display increased activity compared with hIDS and idursulfase

Activity measurements were performed using a coupled assay (Figure S1B) with α-L-iduronidase and artificial substrate 4-methylumbelliferyl-α-L-idopyranosiduronic acid-2-sulfate disodium salt (4-MU-αIdoA2S) (Voznyi et al., 2001). Little difference was observed between hIDS and mIDS, and the hIDS variants displayed slightly lower activities than hIDS (Figure 3A). However, ancestral enzymes (except IDS-A1) showed higher activities than hIDS and mIDS, with an increase of more than 2-fold for IDS-A3. Despite overall low percentages of enzymes with modified Cys84 (expressed as percentage of formylglycine, or %fGly), a correlation could be seen with the specific activity for the ancestral enzymes. This suggested that the higher activities may reflect the percentage of active enzyme in the batches (Figure 3A).

To investigate the significance of increased fGly content in the ancestral enzymes, hIDS, mIDS, and the three original ancestral enzymes were co-expressed with FGE from *Cricetulus griseus* (*Cg*FGE), and an overview of the activity measurements is shown in Figure 3B. Co-expression with *Cg*FGE resulted in a 4.5-fold increase in activity for hIDS, reaching higher activity levels than idursulfase under the same conditions. Interestingly, a similar trend as in the first expression experiment was observed for activity in ancestors when compared with hIDS; activity of IDS-A3 was almost 2-fold higher. Determination of fGly content for some of the enzymes showed that all that were co-expressed with *Cg*FGE had more than 95% fGly (Figure 3B), highlighting the functional and efficient interaction between ancestral enzymes and modern *Cg*FGE. The generally high fGly content in our enzymes might explain their increased activity when compared with idursulfase (for which activity was similar to previously reported values [Demydchuk et al., 2017]), as its %fGly had been determined to be 50% (Muenzer et al., 2006) and 68% (Chung et al.,

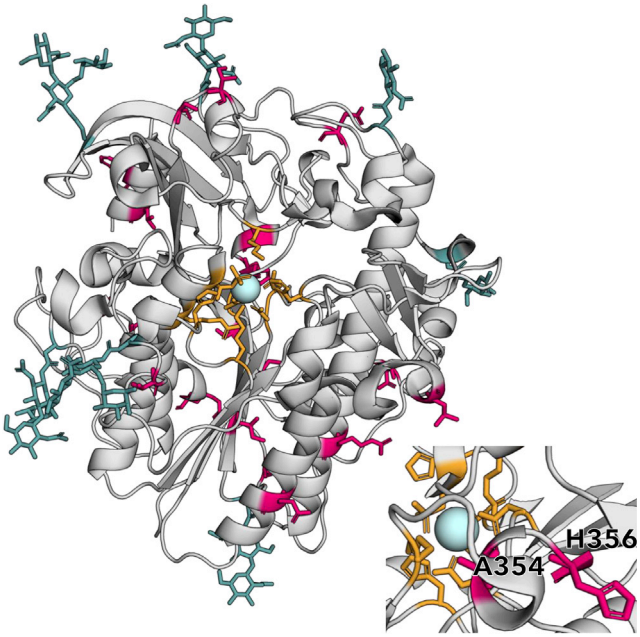


Figure 2. Positions substituted in IDS-A3 shown on the structure of hIDS

All 20 residues that are different in IDS-A3 are colored in pink in the structure of hIDS (PDB: 5FQL (Demydchuk et al., 2017)). Residues in the active site are shown as orange sticks, glycans are shown as cyan sticks, and the calcium ion in the active site is shown as a light blue sphere. The insert shows a close-up of the active site and the location of residues A354 and H356. The figure was prepared using PyMOL version 2.4 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

See also Figure S3 and Table S1.

2014). Overall, no correlation between activity and fGly content was observed in the second expression experiment as formation of fGly was essentially complete.

Thermostability of the enzymes was characterized using nanoDSF (Magnusson et al., 2019; Martin et al., 2014) (Table 1). The melting temperatures were relatively high for all enzymes, and only small differences were found—likely due to the small differences in sequence identities between the enzymes. The data showed that all enzymes with lower T_m values than hIDS contained mutation A354T, which is one of two mutations close to the active site of hIDS (Figure 2) and was predicted for IDS-A3. A new variant of IDS-A3 with reversed mutation was designed (IDS-A3_T354A), for which the melting temperature was determined to be 74.4°C (Table 1, melting curves can be found in Figure S5). Variant IDS-A3_T354A was expressed in ExpiCHO alongside hIDS and IDS-A3, both with and without human FGE (*Hs*FGE). Activities of IDS-A3 and IDS-A3_T354A expressed without *Hs*FGE were similar, and both were higher compared with hIDS (Figure S6). The IDS-A3 variants expressed with *Hs*FGE were also more active than hIDS expressed with *Hs*FGE; however, IDS-A3 displayed higher activity than IDS-A3_T354A. Overall, activities of hIDS and IDS-A3 co-expressed with *Hs*FGE were similar to those upon co-expression with *Cg*FGE.

Encouraged by the *in vitro* activity data, we aimed to verify the increased activity of ancestral enzymes *ex vivo*. Fibroblasts of patients with MPS II were incubated with IDS variants expressed with and without *Hs*FGE, followed by determination of intracellular enzyme concentration and consumed substrate (Figure 4). Figures 4A and 4B show that intracellular concentrations of IDS-A3 and IDS-A3_T354A after 24 h were substantially lower than those of the other enzymes, including IDS-A2. IDS-A3 and IDS-A3_T354A that were not expressed with *Hs*FGE clear similar amounts of substrate as idursulfase (Figure 4C), despite displaying lower activity in *in vitro* activity assays (Figure 3A). hIDS, IDS-A1 and IDS-A2 have similar intracellular activities as idursulfase and IDS-A3 variants, which is in accordance with *in vitro* activity. All IDS-A3 and IDS-A3_T354A batches that were co-expressed with *Hs*FGE clear more substrate than idursulfase and hIDS + *Hs*FGE (Figure 4D), which is also in accordance with activity measurements. These results highlight that modern *Hs*FGE is capable of catalyzing the modification of the active site cysteine in reconstructed ancestral enzymes, which are functional in a complex intracellular system.

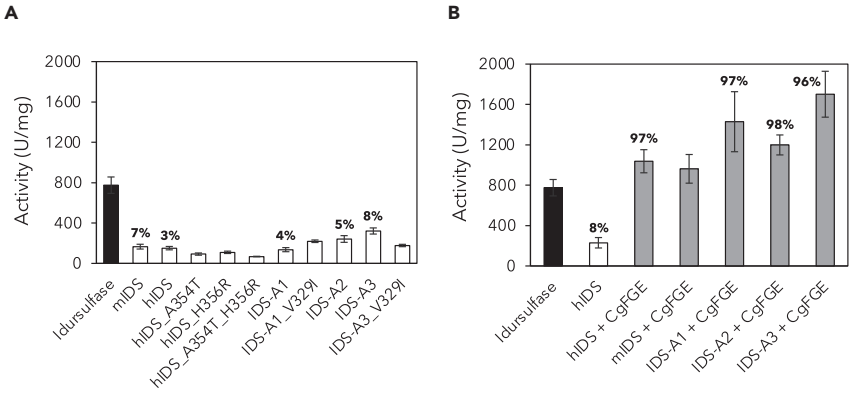


Figure 3. Activity and formylglycine content of IDS enzymes

(A) Specific activities of IDS enzymes; error bars show the standard deviations from 3 to 12 replicates. The percentage of formylglycine as determined by mass spectrometry is shown in bold above the bars for the respective variants.

(B) Specific activities of IDS enzymes co-expressed with FGE from *C. griseus* (*Cg*FGE). Average activities are shown for two independent transfection experiments for each variant, and error bars show the standard deviations from 6 to 12 replicates. The percentage of formylglycine as determined by mass spectrometry is shown in bold above the bars for the respective variants. See also Figures S1 and S6 and Table S1.

Glycosylation profiles are conserved in ancestral enzymes

With respect to the decreased intracellular concentrations observed for IDS-A3 and IDS-A3_T354A, we hypothesized that general protein shape or surface properties of the ancestral enzymes could be influenced either directly through mutations or indirectly through alterations to glycosylation caused by ancestral mutations. A series of experiments was conducted to investigate the latter (Figure 5). Figure 5A shows that hIDS, IDS-A3 and IDS-A3_T354A bind to the mannose-6-phosphate receptor (M6PR) with similar association and dissociation patterns, showing that the receptor interaction is functional and therefore likely not the reason for decreased intracellular concentrations of the ancestral enzymes. Moreover, K_D values for all three enzymes were within one order of magnitude from previously published values (Sonoda et al., 2018). A more detailed investigation of the glycan patterns was performed by size-exclusion chromatography coupled to a multiple angle light scattering detector (SEC-MALS) and *Rapi*Fluor Kit. SEC-MALS analysis showed that the molecular weights of the protein fractions of hIDS, IDS-A3 and IDS-A3_T354A were similar to previously published values for deglycosylated IDS (Wilson et al., 1990) and no significant differences in total glycan weight were found among the three variants (Figure 5B). Analysis of released N-glycans using the *Rapi*Fluor Kit showed that the IDS enzymes presented similar glycan profiles (Figure 5C, fluorescence traces can be found in Figure S7). To facilitate the comparison, glycans were attributed to one of three categories based on their retention times in liquid chromatography with a fluorescence detector coupled to mass spectrometry (LC-FLR-MS): low complexity (eluting before and including mannose-5), medium complexity (eluting after mannose-5 but before sialic acid-containing glycans), and high complexity (glycans with sialic acid). The IDS enzyme profiles were mainly dominated by low- and medium-complexity glycans, as opposed to idursulfase that contained mostly high-complexity glycans. Overall, glycan profiles and M6PR binding characteristics were highly similar among all IDS variants and did not provide any indications as to what might cause the decreased intracellular concentrations observed for IDS-A3 and IDS-A3_T354A.

To investigate other possible effects of the ancestral mutations, molecular dynamics (MD) simulations were performed with hIDS and a homology model of IDS-A3_T354A (Figure 6). Overall, the root-mean-square fluctuations (RMSF) per residue over 100-ns MD trajectories are highly similar for both enzymes and differences in dynamics only seem to occur between residues 387–392 and residues 448–454. Both stretches are situated in flexible loops on the enzyme surface, the second loop being the unresolved loop in the crystal structure (PDB: 5FQL) that is susceptible to proteolytic cleavage and separates the two subdomains of IDS (Demydchuk et al., 2017). Both stretches contain ancestral mutations: Q389E and Y452H and P454H, respectively. It should be noted that the starting conformation for the unresolved loop is predicted, resulting in an inherent uncertainty, but the same conformation is used for both hIDS and IDS-A3-T354A. Moreover, the loop is likely to be highly flexible, which may be reflected in the RMSF profile.

Table 1. Melting temperatures of all enzymes as determined by nanoDSF

	T_m (°C)
mIDS	73.3 ± 0.1
hIDS	73.3 ± 0.2
hIDS_A354T	69.8 ± 0.1
hIDS_H356R	74.6 ± 0.1
hIDS_A354T_H356R	71.9 ± 0.1
IDS-A1	73.9 ± 0.1
IDS-A1_V329I	73.9 ± 0.3
IDS-A2	75.1 ± 0.1
IDS-A3	67.8 ± 0.1
IDS-A3_V329I	69.0 ± 0.6
IDS-A3_T354A	74.4 ± 0.5

See also Figure S5.

DISCUSSION

Encouraged by evolution-inspired redesign of sulfatases through combination of human and mouse sequences (Simonis et al., 2019), we set out to explore the IDS sequence space between primates and rodents by means of ASR. We hypothesized that ASR could be a way to increase the activity or stability of lysosomal enzymes and thereby enhance their potential for enzyme replacement therapy, as we had previously obtained interesting enzyme scaffold using this method (Hendrikse et al., 2018, 2020). Three putative ancestors and several variants were designed and compared with extant IDS enzymes from *Homo sapiens* and *Mus musculus*. Apart from little difference in stability, we found two ancestral variants that displayed higher activity than hIDS and idursulfase, both in *in vitro* assays as well as in *ex vivo* experiments using patient fibroblasts.

The phylogenetic tree that was used for ancestral reconstruction is mostly in accordance with a species tree that was created with the same set of species (Upham et al., 2019), apart from small differences in the clades wherein the ancestral nodes were chosen—in particular at divergence points that are not resolved in the IDS tree. These unresolved nodes are likely due to a too high degree of conservation between the extant sequences, which is also reflected in the low number of mutations in the ancestors (20 for IDS-A3). The high sequence similarity between the ancestors and the modern human enzyme could be beneficial in a therapeutic setting, where immunogenicity is a common challenge. This advantage had previously been described for ancestral uricases (Kratzer et al., 2014). None of the 20 predicted ancestral mutations occur in any of the 128 positions in the 550-residue protein that harbor reported disease-causing mutations (Demydchuk et al., 2017). We take this as a sign of the robustness of the reconstruction process, as important sites are more likely to be conserved throughout evolution and are also more prone to disease-causing mutations. Analysis of evolutionary conservation throughout the IDS alignment with ConSurf (Armon et al., 2001) shows that most ancestral mutations occur in highly variable positions (Figure S8), which could be expected. However, mutations T146S (close to glycosylation site N144), H356R (that appears to be stabilizing), I360M, and S470A occur in relatively conserved positions. Despite its destabilizing effect, mutation A354T occurs at a site that is highly variable in the alignment.

Based on the initial activity measurements and determination of %fGly (Figure 3A) we hypothesized that the ancestral IDS enzymes may be more prone to modification by FGE. The FGE binds to newly synthesized, unfolded sulfatases in the endoplasmic reticulum and catalyzes the generation of fGly from a cysteine residue through a multistep redox process (Roeser et al., 2006). The sequence motif that is required for FGE binding (C-[TSAC]-PSR) is conserved in all human sulfatases, suggesting a general binding mechanism, but conversion of the cysteine to fGly in the different sulfatases depends on the flanking sequences of approximately 10 residues on either side (Preusser-Kunze et al., 2005). However, no ancestral mutations occur in the vicinity of cysteine 84, and it is thus not obvious that binding to the FGE is altered in the ancestral enzymes. This was supported by activity and %fGly data upon co-expression with *Cs*FGE (Figure 3B), and we concluded that the observed increased activity of especially IDS-A3 when compared with hIDS in the *in vitro* assays was likely due to an increase in catalytic rate. Overall the activity data show that ancestral and modern IDS from primates and rodents are successfully modified by modern FGE from primates and rodents as well.

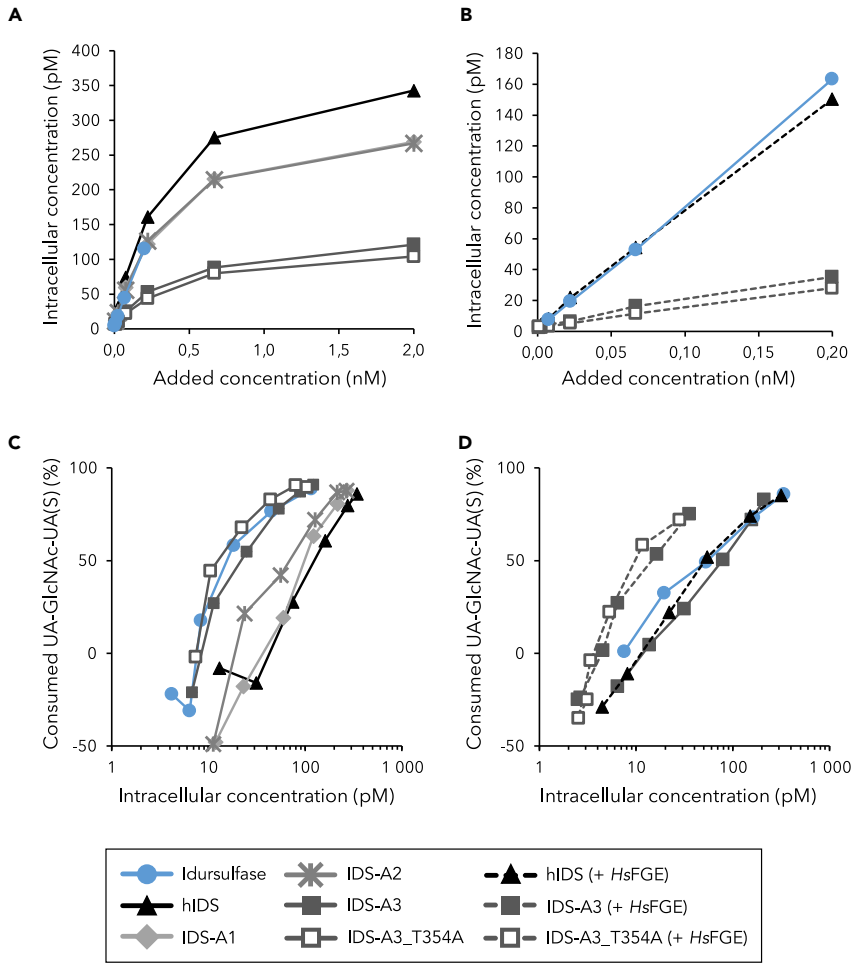


Figure 4. Intracellular concentrations of IDS variants and degradation of heparan sulfate in fibroblasts of patients with MPS II

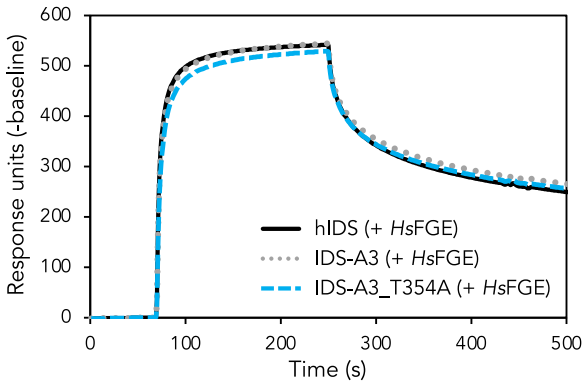
(A) Intracellular concentrations of enzymes expressed without *Hs*FGE 24 h after treatment. The saturation indicates receptor-mediated uptake of all IDS variants.

(B) Intracellular concentrations of enzymes expressed with *Hs*FGE 24 h after treatment. Different concentrations were used due to differences in activity of one order of magnitude. (C) Consumed substrate UA-GlcNAc-UA(S) (when compared with internal standard chondroitin disaccharide Δdi-4S sodium) as a function of intracellular enzyme concentration for enzymes expressed without *Hs*FGE.

(D) Consumed substrate UA-GlcNAc-UA(S) (when compared with the internal standard) as a function of intracellular enzyme concentration for enzymes expressed with *Hs*FGE. For clarity, the intracellular enzyme concentration is shown on a logarithmic axis in (C and D). Based on biological duplicates the average errors of intracellular concentration and consumed substrate are estimated to be 8% and 33%, respectively. Errors were smaller for high enzyme concentrations and larger for low enzyme concentrations.

Increased clearance of substrate by IDS-A3 and its variant IDS-A3_T354A when compared with hIDS and idursulfase was also observed in fibroblasts of patients with MPS II (Figure 4). Both IDS-A3 variants that were not co-expressed with *Hs*FGE cleared similar percentages of substrate as idursulfase in patient fibroblasts, despite lower activity in the *in vitro* assay (Figure S6). The same variants but co-expressed with *Hs*FGE consume more substrate compared with idursulfase and hIDS + *Hs*FGE, which is mostly in accordance with *in vitro* activity measurements. It should be noted that the observed increased clearance of substrate after 24 h could be due to several factors, such as increased catalytic rate, and also increased stability of the enzymes in the lysosome. The experiments in patient fibroblast also revealed a decreased intracellular concentration for IDS-A3 and IDS-A3_T354A when compared with all other enzymes. This could be due to a decreased cellular uptake, which is receptor-mediated through interaction of the M6PR with the glycans on the surface of IDS. Studies have shown that all putative glycosylation sites in IDS are

A



B

	$MW_{protein}$ (kDa)	MW_{glycan} (kDa)
hIDS (+ *Hs*FGE)	61 ± 2	16 ± 3
IDS-A3 (+ *Hs*FGE)	58 ± 2	15 ± 2
IDS-A3_T354A (+ *Hs*FGE)	60 ± 3	15 ± 4

C

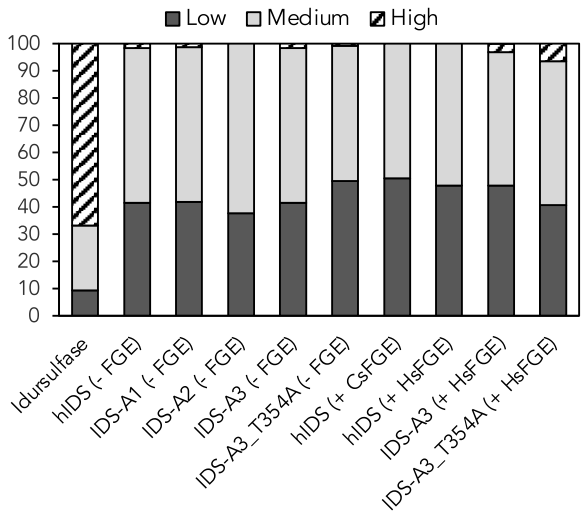


Figure 5. Investigation of glycan profiles

(A) Binding of hIDS, IDS-A3 and IDS-A3_T354A co-expressed with *Hs*FGE to the recombinant human M6P receptor was analyzed by surface plasmon resonance. Measurements were performed on a Biacore T200 system using the single cycle kinetics mode. K_D values were determined to be 32 nM (hIDS), 26 nM (IDS-A3), and 46 nM (IDS-A3_T354A).

(B) Molecular weights of protein and glycan fractions as determined by SEC-MALS.

(C) Profiles of released N-glycans as determined by GlycoWorks RapiFluor-MS N-Glycan Kit. Glycans were divided into three categories based on retention time in LC-FLR-MS: low, medium, and high complexity (relative abundancy shown). The low-complexity N-glycans elute earlier than more complex glycans. Fluorescence traces can be found in Figure S7.

used, but none of them is crucial for activity (Millat et al., 1997). Even though none of the glycosylation sites was found to be essential for cellular uptake and lysosomal targeting of the enzyme, N280 appeared to be the most important site for this purpose. All eight glycosylation sites were conserved in the ancestral IDS enzymes, apart from T146S in the motif around N144 for IDS-A2 and IDS-A3. Moreover, no mutations occur in the direct vicinity of N280, making it unlikely that ancestral mutations disturb the overall glycosylation in

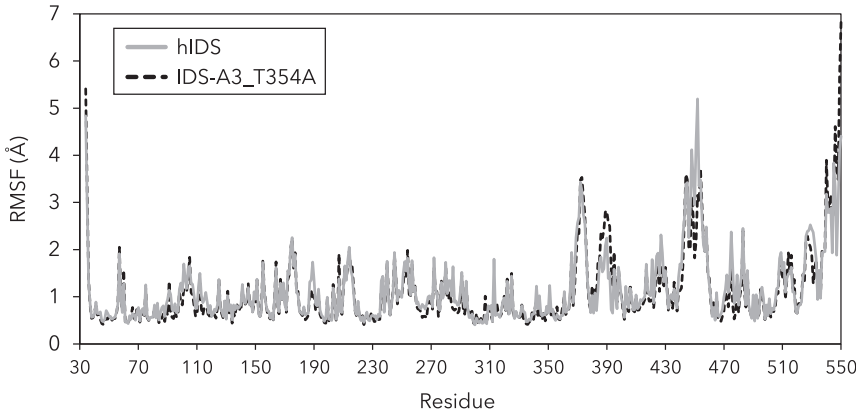


Figure 6. Molecular dynamics simulations of hIDS and IDS-A3_T354A

Root-mean-square fluctuation (RMSF) of hIDS and IDS-A3_T354A during molecular dynamics (MD) simulations. The RMSF values are the average of three independent 100-ns MD trajectories for each enzyme.

ancestral IDS, as supported by the data in Figure 5. No sialic acid-containing glycans were detected in any of the IDS samples that had been produced in ExpiCHO (Figure S7), which we had previously experienced with expression of other lysosomal enzymes in this strain (unpublished data). It has been posed that sialylation is important for antibody recognition/immunogenicity and circulating half-life of IDS (Chung et al., 2014).

In case that cellular uptake is influenced directly by ancestral mutations on the surface, the difference between IDS-A2 and IDS-A3 is of particular interest. Only 10 positions differ between the two sequences, of which 7 are located on the surface of the enzyme: L189V(IDS-A2)/A(IDS-A3), E254Q, S369P, S397T, P454H, I472F(IDS-A2)/S(IDS-A3), and F545P (and likely Y452H, which is located in the unresolved loop in the structure [Demydchuk et al., 2017]). Changes in surface charge occur due to E254Q and possibly P454H and Y452H, which are located in the unresolved loop in the structure (PDB: 5FQL). E254 is spatially close to flexible region 270–280, which includes glycosylated residue N280 that is hypothesized to be of importance for cellular uptake. The decreased intracellular concentration of IDS-A3 variants could be a possible advantage as it implies a prolonged systemic circulation, or plasma half-life, for these enzymes. Studies with other sulfatases have shown that prolonged systemic circulation may result in increased brain uptake (Gustavsson et al., 2019). The obstacle of treating CNS manifestations of MPS II remains one of the biggest challenges to be met at the present time, and various approaches to improve distribution are actively investigated, such as the use of other receptors and endocytic mechanisms (Chen et al., 2019; Sonoda et al., 2018).

We have shown that ASR can successfully be applied to enzymes in complex systems that function in concert with extant enzymes and receptors in cells. Two ancestral lysosomal enzymes were found to have increased activity *in vitro* and *ex vivo* when compared with idursulfase, the enzyme that is currently available for ERT. This increase in activity would have been difficult to achieve by other protein engineering strategies, as the identification and selection of positions to mutate as well as the resulting ancestral residues is not an obvious process. Most importantly, the activity increase allows for a decreased dose and shorter intravenous administration times in ERT, which are both major advantages for potential therapeutic enzymes. We believe these results are highly encouraging as they demonstrate the potential of ASR for improving therapeutic properties of lysosomal enzymes for ERT, a therapeutic area with a significant unmet medical need.

Limitations of the study

Ancestral sequence reconstruction is a relatively young bioinformatics method: updated methods, models, and software are continuously developed. Owing to the time that is needed to carry out the experimental work to verify the computational results, some of the methods or programs may have been further developed. We acknowledge that even though our model searches in two different software programs suggested a gamma distribution as optimal for our data, there may be advanced models that more accurately describe among-site rate heterogeneity in our dataset.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Erik Nordling (Erik.Nordling@sobi.com).

Materials availability

All protein sequences can be found in the Supplemental Data Items and the corresponding proteins can be obtained as described in the Transparent methods section.

Data and code availability

The published article includes all datasets (sequences) generated and analyzed during this study.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2021.102154.

ACKNOWLEDGMENTS

We acknowledge the financial support of the Swedish Foundation for Strategic Research (SSF), grant number ID16-0036 and the Swedish Innovation Agency Vinnova, grant numbers: 2019-00103 and 2017-02105. We thank Johanna Rodhe, Monica Andersson, and Anna Johansson for assistance with protein expression.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S., S.S.G., E.N., J.R., P.-O.S., and N.M.H.; Methodology, N.M.H., C.S., Å.M., A.T., T.A., U.W., and J.B.; Investigation, N.M.H., N.T., D.P., C.S., Å.M., A.T., T.A., U.W., J.B.; Writing, N.M.H. wrote most of the manuscript. All authors contributed and have approved the manuscript.

DECLARATION OF INTERESTS

Anna Sandegren, Tommy Andersson, Jenny Blomqvist, Åsa Makower, Dominik Possner, Chao Su, Agneta Tjernberg, Ulrica Westermark, and Stefan Svensson Gelius were employed by Swedish Orphan Biovitrum AB at the time the study was conducted. Natalie Hendrikse was funded by the Swedish Foundation for Strategic Research and was jointly employed by the Royal Institute of Technology and Swedish Orphan Biovitrum AB. Erik Nordling is currently employed by Swedish Orphan Biovitrum AB. All authors declare no competing interests.

Received: July 1, 2020
Revised: November 11, 2020
Accepted: February 2, 2021
Published: March 19, 2021

REFERENCES

Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. *307*, 447–463.

Bach, G., Eisenberg, F., Jr., Cantz, M., and Neufeld, E.F. (1973). The defect in the Hunter syndrome: deficiency of sulfoiduronate sulfatase. Proc. Natl. Acad. Sci. U S A *70*, 2134–2138.

Boustany, R.-M.N. (2013). Lysosomal storage diseases—the horizon expands. Nat. Rev. Neurol. *9*, 583–598.

Chen, H.H., Sawamoto, K., Mason, R.W., Kobayashi, H., Yamaguchi, S., Suzuki, Y., Orii, K., Orii, T., and Tomatsu, S. (2019). Enzyme replacement therapy for mucopolysaccharidoses; past, present, and future. J. Hum. Genet. *64*, 1153–1171.

Chung, Y.K., Sohn, Y.B., Sohn, J.M., Lee, J., Chang, M.S., Kwun, Y., Kim, C.H., Lee, J.Y., Yook, Y.J., Ko, A.R., et al. (2014). A biochemical and physicochemical comparison of two recombinant enzymes used for enzyme replacement therapies of hunter syndrome. Glycoconj. J. *31*, 309–315.

Demydchuk, M., Hill, C.H., Zhou, A., Bunkóczi, G., Stein, P.E., Marchesan, D., Deane, J.E., and Read, R.J. (2017). Insights into Hunter syndrome from the structure of iduronate-2-sulfatase. Nat. Commun. *8*, 15786.

Desnick, R.J., and Schuchman, E.H. (2012). Enzyme replacement therapy for lysosomal diseases: lessons from 20 Years of experience and remaining challenges. Annu. Rev. Genomics Hum. Genet. *13*, 307–335.

Von Figura, K., Schmidt, B., Selmer, T., and Dierks, T. (1998). A novel protein modification generating an aldehyde group in sulfatases: its role in catalysis and disease. BioEssays *20*, 505–510.

Grabowski, G.A., and Whitley, C. (2017). Ten plus one challenges in diseases of the lysosomal system. Mol. Genet. Metab. *120*, 38–46.

Gumulya, Y., and Gillam, E.M.J. (2017). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro approach to protein engineering. Biochem. J. *474*, 1–19.

Gustavsson, S., Ohlin Sjöström, E., Tjernberg, A., Janson, J., Westermark, U., Andersson, T., Makower, Å., Arnelöf, E., Andersson, G., Svartengren, J., et al. (2019). Intravenous delivery of a chemically modified sulfamidase efficiently reduces heparan sulfate storage and brain pathology in mucopolysaccharidosis IIIA mice. Mol. Genet. Metab. Rep. *21*, 100510.

Heartlein, M., and Kimura, A. (2014). Discovery and clinical development of idursulfase (Elaprase®) for the treatment of mucopolysaccharidosis II (Hunter syndrome). RSC Drug Discov. Ser. *38*, 164–182.

Hendrikse, N.M., Charpentier, G., Nordling, E., and Syrén, P.-O. (2018). Ancestral diterpene cyclases show increased thermostability and substrate acceptance. FEBS J. *285*, 4660–4673.

Hendrikse, N.M., Holmberg Larsson, A., Svensson Gelius, S., Kuprin, S., Nordling, E., and Syrén, P.O. (2020). Exploring the therapeutic potential of modern and ancestral phenylalanine/tyrosine ammonia-lyases as supplementary treatment of hereditary tyrosinemia. Sci. Rep. *10*, 1315.

Hunter, C. (1917). A rare disease in two brothers. Proc. R. Soc. Med. *10*, 104–116.

Kim, C., Seo, J., Chung, Y., Ji, H.J., Lee, J., Sohn, J., Lee, B., and Jo, E.C. (2017). Comparative study of idursulfase beta and idursulfase in vitro and in vivo. J. Hum. Genet. *62*, 167–174.

Kratzer, J.T., Lanaspa, M.A., Murphy, M.N., Cicerchi, C., Graves, C.L., Tipton, P.A., Ortlund, E.A., Johnson, R.J., and Gaucher, E.A. (2014). Evolutionary history and metabolic insights of ancient mammalian uricases. Proc. Natl. Acad. Sci. U S A *111*, 3763–3768.

Van Loo, B., Bayer, C.D., Fischer, G., Jonas, S., Valkov, E., Mohamed, M.F., Vorobieva, A., Dutruel, C., Hyvönen, M., and Hollfelder, F. (2019). Balancing specificity and promiscuity in enzyme evolution: multidimensional activity transitions in the alkaline phosphatase superfamily. J. Am. Chem. Soc. *141*, 370–387.

Magnusson, A.O., Szekrenyi, A., Joosten, H.J., Finnigan, J., Charnock, S., and Fessner, W.D. (2019). nanoDSF as screening tool for enzyme libraries and biotechnology development. FEBS J. *286*, 184–204.

Martin, L., Schwarz, S., and Breitsprecher, D. (2014). Thermal unfolding analyzing thermal unfolding of Proteins : nanotemper technol. Application Note NT-PR-001, 1–8.

Merkl, R., and Sterner, R. (2016). Ancestral protein reconstruction: techniques and applications. Biol. Chem. *397*, 1–21.

Millat, G., Froissart, R., Maire, I., and Bozon, D. (1997). Characterization of iduronate sulphatase mutants affecting N-glycosylation sites and the cysteine-84 residue. Biochem. J. *326*, 243–247.

Muenzer, J., Wraith, J.E., Beck, M., Giugliani, R., Harmatz, P., Eng, C.M., Vellodi, A., Martin, R., Ramaswami, U., Gucsavas-Calikoglu, M., et al. (2006). A phase II/III clinical study of enzyme replacement therapy with idursulfase in mucopolysaccharidosis II (Hunter syndrome). Genet. Med. *8*, 465–473.

Nguyen, V., Wilson, C., Hoemberger, M., Stiller, J.B., Agafonov, R.V., Kutter, S., English, J., Theobald, D.L., and Kern, D. (2016). Evolutionary drivers of thermoadaptation in enzyme catalysis. Science *355*, 289–294.

Nicoll, C.R., Bailleul, G., Fiorentini, F., Mascotti, M.L., Fraaije, M.W., and Mattevi, A. (2020). Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. Nat. Struct. Mol. Biol. *27*, 14–24.

Preusser-Kunze, A., Mariappan, M., Schmidt, B., Gande, S.L., Mutenda, K., Wenzel, D., Von Figura, K., and Dierksl, T. (2005). Molecular characterization of the human Cα-formylglycine-generating enzyme. J. Biol. Chem. *280*, 14900–14910.

Risso, V.A., and Sanchez-Ruiz, J.M. (2017). Resurrected ancestral proteins as scaffolds for protein engineering. In Directed Enzyme Evolution: Advances and Applications, M. Alcalde, ed. (Springer International Publishing AG), pp. 229–255.

Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013). Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases. J. Am. Chem. Soc. *135*, 2899–2902.

Roeser, D., Preusser-Kunze, A., Schmidt, B., Gasow, K., Wittmann, J.G., Dierks, T., Von Figura, K., and Rudolph, M.G. (2006). A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. Proc. Natl. Acad. Sci. U S A *103*, 81–86.

Samelson-Jones, B.J., and Arruda, V.R. (2019). Protein-engineered coagulation factors for hemophilia gene therapy. Mol. Ther. Methods Clin. Dev. *12*, 184–201.

Schmidt, B., Selmer, T., Ingendoh, A., and von Figurat, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. Cell *82*, 271–278.

Schrödinger, LLC, 2020. The PyMOL Molecular Graphics System. Schrödinger, LLC. Version 2.0..

Schupfner, M., Straub, K., Busch, F., Merkl, R., and Sterner, R. (2020). Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. Proc. Natl. Acad. Sci. U S A *117*, 346–354.

Simonis, H., Yaghootfam, C., Sylvester, M., Gieselmann, V., and Matzner, U. (2019). Evolutionary redesign of the lysosomal enzyme arylsulfatase A increases efficacy of enzyme replacement therapy for metachromatic leukodystrophy. Hum. Mol. Genet. *28*, 1810–1821.

Sohn, Y.B., Cho, S.Y., Park, S.W., Kim, S.J., Ko, A.R., Kwon, E.K., Han, S.J., and Jin, D.K. (2013). Phase I/II clinical trial of enzyme replacement therapy with idursulfase beta in patients with mucopolysaccharidosis II (Hunter Syndrome). Orphanet J. Rare Dis. *8*, 2–9.

Sonoda, H., Morimoto, H., Yoden, E., Koshimura, Y., Kinoshita, M., Golovina, G., Takagi, H., Yamamoto, R., Minami, K., Mizoguchi, A., et al. (2018). A blood-brain-barrier-penetrating anti-human transferrin receptor antibody fusion protein for neuronopathic mucopolysaccharidosis II. Mol. Ther. *26*, 1366–1374.

Thornton, J.W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. *5*, 366–375.

Upham, N.S., Esselstyn, J.A., and Jetz, W. (2019). Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. *17*, e3000494.

Voznyi, Y.V., Keulemans, J.L.M., and van Diggelen, O.P. (2001). A fluorimetric enzyme assay for the diagnosis of MPS II (Hunter disease). J. Inherit. Metab. Dis. *24*, 675–680.

Whiteman, D.A.H., and Kimura, A. (2017). Development of idursulfase therapy for mucopolysaccharidosis type II(Hunter syndrome): the past, the present and the future. Drug Des. Devel. Ther. *11*, 2467–2480.

Wilding, M., Peat, T.S., Kalyaanamoorthy, S., Newman, J., Scott, C., and Jermiin, L.S. (2017). Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. Green Chem *19*, 5375–5380.

Wilson, C., Agafonov, R.V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D.L., et al. (2015). Using ancient protein kinases to unravel a modern cancer drug's mechanism. Science *347*, 882–886.

Wilson, P.J., Morris, C.P., Anson, D.S., Occhiodoro, T., Bielicki, J., Clements, P.R., and Hopwood, J.J. (1990). Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA. Proc. Natl. Acad. Sci. U S A *87*, 8531–8535.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

Zakas, P.M., Brown, H.C., Knight, K., Meeks, S.L., Spencer, H.T., Gaucher, E.A., and Doering, C.B. (2017). Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. Nat. Biotechnol. *35*, 35–37.

iScience, Volume *24*

Supplemental information

Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome

Natalie M. Hendrikse, Anna Sandegren, Tommy Andersson, Jenny Blomqvist, Åsa Makower, Dominik Possner, Chao Su, Niklas Thalén, Agneta Tjernberg, Ulrica Westermark, Johan Rockberg, Stefan Svensson Gelius, Per-Olof Syrén, and Erik Nordling

Supplemental Data Items

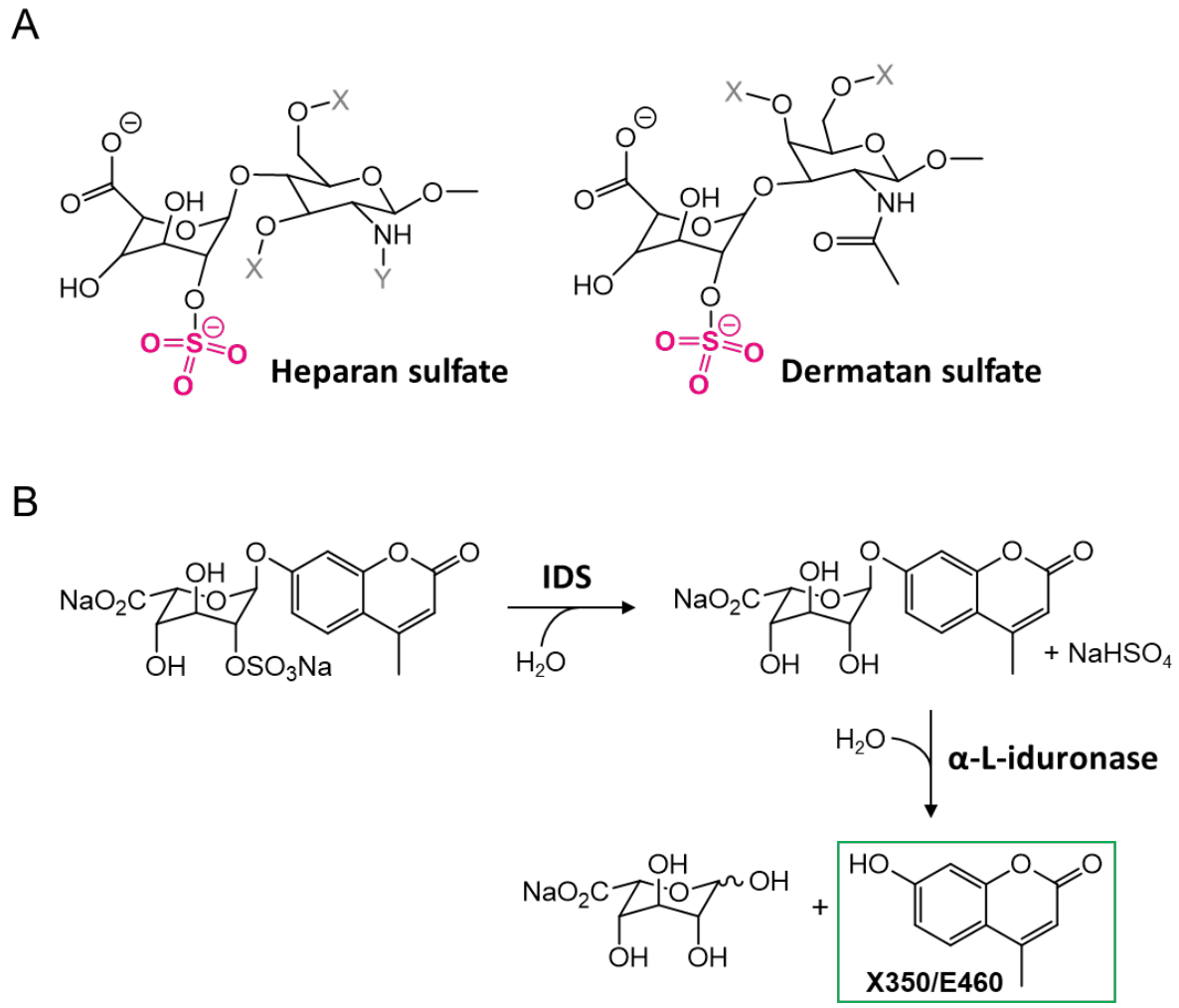
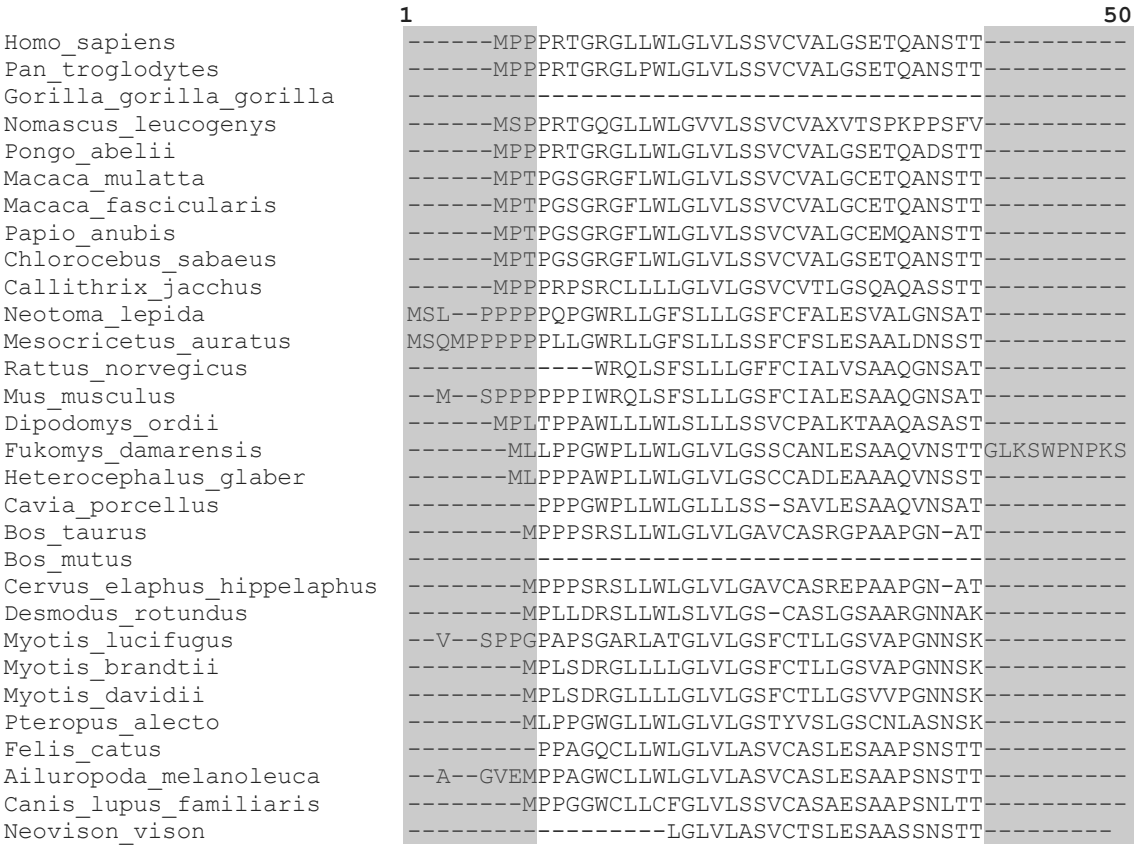


Figure S1: Natural and artificial substrates of IDS, related to Figure 3. A) Natural substrates heparan sulfate (**α**1-4 linkage) and dermatan sulfate (**α**1-3 linkage). The sulfate group that is removed by IDS is indicated in pink. Variable substituents are indicated by X and Y, where X can be either H or SO_3^- and Y can be H, SO_3^- or $COCH_3$. **B)** Artificial substrate 4-methylumbelliferyl-**α**-L-idopyranosiduronic acid-2-sulphate disodium salt (4-MU-**α**IdoA2S) is converted to 4-MU- **α**-iduronide by IDS, followed by hydrolysis by **α**-L-iduronidase to release 4-methylumbelliferone. 4-methylumbelliferone (4-MU) can be detected in a fluorometric assay through excitation at 350 nm and measuring emission at 460 nm.

A


1
50
Homo_sapiens ------MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTT----------
Pan_troglodytes ------MPPPRTGRGLPWLGLVLSSVCVALGSETQANSTT----------
Gorilla_gorilla_gorilla ----------
Nomascus_leucogenys ------MSPPRTGQGLLWLGVVLSSVCVAXVTSPKPPSFV----------
Pongo_abelii ------MPPPRTGRGLLWLGLVLSSVCVALGSETQADSTT----------
Macaca_mulatta ------MPTPGSGRGFLWLGLVLSSVCVALGCETQANSTT----------
Macaca_fascicularis ------MPTPGSGRGFLWLGLVLSSVCVALGCETQANSTT----------
Papio_anubis ------MPTPGSGRGFLWLGLVLSSVCVALGCEMQANSTT----------
Chlorocebus_sabaeus ------MPTPGSGRGFLWLGLVLSSVCVALGSETQANSTT----------
Callithrix_jacchus ------MPPPRPSRCLLLLGLVLGSVCVTLGSQAQASSTT----------
Neotoma_lepida MSL--PPPPPQPGWRLLGFSLLLGSFCFALESVALGNSAT----------
Mesocricetus_auratus MSQMPPPPPPLLGWRLLGFSLLLSSFCFSLESAALDNSST----------
Rattus_norvegicus ----------WRQLSFSLLLGFFCIALVSAAQGNSAT----------
Mus_musculus --M--SPPPPPPIWRQLSFSLLLGSFCIALESAAQGNSAT----------
Dipodomys_ordii ------MPLTPPAWLLLWLSLLLSSVCPALKTAAQASAST----------
Fukomys_damarensis -------MLLPPGWPLLWLGLVLGSSCANLESAAQVNSTTGLKSWPNPKS
Heterocephalus_glaber -------MLPPPAWPLLWLGLVLGSCCADLEAAAQVNSST----------
Cavia_porcellus ---------PPPGWPLLWLGLLLSS-SAVLESAAQVNSAT----------
Bos_taurus --------MPPPSRSLLWLGLVLGAVCASRGPAAPGN-AT----------
Bos_mutus ----------
Cervus_elaphus_hippelaphus --------MPPPSRSLLWLGLVLGAVCASREPAAPGN-AT----------
Desmodus_rotundus --------MPLLDRSLLWLSLVLGS-CASLGSAARGNNAK----------
Myotis_lucifugus --V--SPPGPAPSGARLATGLVLGSFCTLLGSVAPGNNSK----------
Myotis_brandtii --------MPLSDRGLLLLGLVLGSFCTLLGSVAPGNNSK----------
Myotis_davidii --------MPLSDRGLLLLGLVLGSFCTLLGSVVPGNNSK----------
Pteropus_alecto --------MLPPGWGLLWLGLVLGSTYVSLGSCNLASNSK----------
Felis_catus ---------PPAGQCLLWLGLVLASVCASLESAAPSNSTT----------
Ailuropoda_melanoleuca --A--GVEMPPAGWCLLWLGLVLASVCASLESAAPSNSTT----------
Canis_lupus_familiaris --------MPPGGWCLLCFGLVLSSVCASAESAAPSNLTT----------
Neovison_vison ----------LGLVLASVCTSLESAASSNSTT---------
51
100
Homo_sapiens ----------DALNVLLIIVDDLRPSLGCY
Pan_troglodytes ----------DALNVLLIIVDDLRPSLGCY
Gorilla_gorilla_gorilla ----------
Nomascus_leucogenys ----------DALNVLLIIVDDLRPSLGCY
Pongo_abelii ----------DGLNVLLIIVDDLRPSLGCY
Macaca_mulatta ----------DALNILLIIVDDLRPSLGCY
Macaca_fascicularis ----------DALNILLIIVDDLRPSLGCY
Papio_anubis ----------DALNILLIIVDDLRPSLGCY
Chlorocebus_sabaeus ----------DALNILLIIVDDLRPSLGCY
Callithrix_jacchus ----------DALNVLLIIVDDLRPSLGCY
Neotoma_lepida ----------DALNILLIIVDDLRPSLGCY
Mesocricetus_auratus ----------DALNVLLIIVDDLRPSLGCY
Rattus_norvegicus ----------DALNILLIIVDDLRPSLGCY
Mus_musculus ----------DALNILLIIVDDLRPSLGCY
Dipodomys_ordii ----------GALNVLLIIVDDLRPSLGCY
Fukomys_damarensis ATHTSGFLCILAPSVVTHIPGGFPLWLGCQDAPNVLLIIVDDLRPSLGCY
Heterocephalus_glaber ----------DARNVLLIIVDDLRPSLGCY
Cavia_porcellus ----------DALNVLLIIVDDLRPSLGCY
Bos_taurus ----------DPLNVLLIIVDDLRPSLGCY
Bos_mutus ----------DPLNVLLIIVDDLRPSLGCY
Cervus_elaphus_hippelaphus ----------DPLNVLLIIVDDLRPSLGCY
Desmodus_rotundus ----------DALNVLLIIVDDLRLSLGCY
Myotis_lucifugus ----------DALNVLLIIVDDLRLSLGCY
Myotis_brandtii ----------DALNVLLIIVDDLRLSLGCY
Myotis_davidii ----------DALNVLLIIVDDLRLSLGCY
Pteropus_alecto ----------DALNVLLIIVDDLRPSLGCY
Felis_catus ----------GPLNVLLIIVDDLRPSLGCY
Ailuropoda_melanoleuca ----------APLNVLLIIVDDLRPSLGCY
Canis_lupus_familiaris ----------APLNVLLIIVDDLRPSLGCY
Neovison_vison ----------APLNVLLIIVDDLRPSLGCY

```
                                101                                 150
Homo_sapiens                    GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Pan_troglodytes                 GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Gorilla_gorilla_gorilla         -DELVKAPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Nomascus_leucogenys             GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Pongo_abelii                    GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Macaca_mulatta                  GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Macaca_fascicularis             GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Papio_anubis                    GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Chlorocebus_sabaeus             GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLHN
Callithrix_jacchus              GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Neotoma_lepida                  GDKLVRSPNIDQLASHSILFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Mesocricetus_auratus            GEKVVRSPNIDQLASHSVLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Rattus_norvegicus               GDKLVRSPNIDQLASHSIVFENAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Mus_musculus                    GDKLVRSPNIDQLASHSVLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Dipodomys_ordii                 GDKIVRSPNIDQLAAHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Fukomys_damarensis              GDKAVKSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Heterocephalus_glaber           GDKLVKSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Cavia_porcellus                 GDKLVKSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Bos_taurus                      GNKLIRSPNIDQLASRSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Bos_mutus                       GNKLIRSPNIDQLASRSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Cervus_elaphus_hippelaphus      GNKLIRSPNIDQLASRSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Desmodus_rotundus               GDKLVKSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Myotis_lucifugus                GDKLIRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Myotis_brandtii                 GDKLIRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Myotis_davidii                  GDKLIRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Pteropus_alecto                 GDKLIRSPNIDQLASHSLLFWNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Felis_catus                     GDKAIRSPNIDQLASRSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Ailuropoda_melanoleuca          GDKLIRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Canis_lupus_familiaris          GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYD
Neovison_vison                  GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLPGRRPDTTRLYD
                                 :: :::*******::*::* ****************.*********::

                                151                                 200
Homo_sapiens                    FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHTDDSPYSW
Pan_troglodytes                 FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHTDDSPYSW
Gorilla_gorilla_gorilla         FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGGISSNHTDDSPYSW
Nomascus_leucogenys             FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHTDDSPYSW
Pongo_abelii                    FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHTDDSPYSW
Macaca_mulatta                  FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GITSNHTDDSPYSW
Macaca_fascicularis             FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GITSNHTDDSPYSW
Papio_anubis                    FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHPEGITSNHTDDSPYSW
Chlorocebus_sabaeus             FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GITSNHTDDSPYSW
Callithrix_jacchus              FNSYWRVHAGNFSTIPQYFKDNGYVTMSVGKVFHP-GISSNHSDDSPYSW
Neotoma_lepida                  FNSYWRVHSGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Mesocricetus_auratus            FNSYWRVHSGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Rattus_norvegicus               FNSYWRVHSGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDYPYSW
Mus_musculus                    FNSYWRVHSGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDYPYSW
Dipodomys_ordii                 FNSYWRSHSGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNYSDDSPYSW
Fukomys_damarensis              FNSYWRVHSGNFSTLPQYFKENGYVTLSVGKVFHP-GISSNHTDDSPYSW
Heterocephalus_glaber           FNSYWRVHSGNFSTLPQYFKENGYVTLSVGKVFHP-GISSNHSDDSPYSW
Cavia_porcellus                 FGSYWRVHAGNFSTLPQHFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Bos_taurus                      FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Bos_mutus                       FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Cervus_elaphus_hippelaphus      FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Desmodus_rotundus               FNSYWRVQAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Myotis_lucifugus                FNSYWRVQAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Myotis_brandtii                 FNSYWRVQAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDFPYSW
Myotis_davidii                  FNSYWRVQAGNFSTIPQYFKENGYMTMSVGKVFHP-GISSNHSDDSPYSW
Pteropus_alecto                 FNSYWRAHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNHSDDSPYSW
Felis_catus                     FNSYWRIHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNYSDDSPYSW
Ailuropoda_melanoleuca          FNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHP-GISSNYSDDSPYSW
Canis_lupus_familiaris          FNSYWRVHAGNFSTLPQYFKENGYVTMSVGKVFHP-GISSNYSDDSPYSW
Neovison_vison                  FNSYWRVHAGNFSTIPQHFKENGYVTMSVGKVFHP-GISSNYSDDSPFSW
                                *.**** ::*****:**:**:***:*:******** **:**::** *:**
```

```
                                201                                 250
Homo_sapiens                    SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTE
Pan_troglodytes                 SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTE
Gorilla_gorilla_gorilla         SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTE
Nomascus_leucogenys             SFPPYHPSSXXXXXXKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTE
Pongo_abelii                    SFPPYHPSSEKYENTKTCRGPDGELHANLIAKKMCWMFPRAPCCDKQSTE
Macaca_mulatta                  SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Macaca_fascicularis             SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Papio_anubis                    SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Chlorocebus_sabaeus             SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Callithrix_jacchus              SFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Neotoma_lepida                  SFPPYHPSTEKYENTKTCRGQDGKLHANLLCPVDVADVPEGTLPDKQSTE
Mesocricetus_auratus            SFPPYHPSSEKYENTKTCRGQDGKLHANLLCPVDVADVPEGTLPDKQSTE
Rattus_norvegicus               SFPPYHPSSEKYENTKTCKGQDGKLHTNLLCPVDVADVPEGTLPDKQSTE
Mus_musculus                    SFPPYHPSSEKYENTKTCKGQDGKLHANLLCPVDVADVPEGTLPDKQSTE
Dipodomys_ordii                 SFLPYHPSSEKYENTKTCRGQDGELHANLLCPVDVKDVPEGTLPDIQSTE
Fukomys_damarensis              SFPPYHPSSEKYENTKTCRGQDGELHANLICPVDMADVPEGTLPDKQSTE
Heterocephalus_glaber           SFPPYHPSSEKYENTKTCRGQDGELHANLICPVDMADVPEGTLPDKQSTE
Cavia_porcellus                 SLPPYHPSSEKYENTKTCRGQDGELHANLICPVDMADVPEGTLPDKQSTE
Bos_taurus                      SVPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Bos_mutus                       SVPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDKQSTE
Cervus_elaphus_hippelaphus      SVPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVVDVPEGTLPDMQSTE
Desmodus_rotundus               SVPPYHPSSEKYENTKTCRGPDGELHANLICPVDMADIPEGTLPDKQSTK
Myotis_lucifugus                SVLPYHPSSEKYENTKTCKGPDGELHANLICPVDMADIPEGTLPDKQSTK
Myotis_brandtii                 SVLPYHPSSEKYENTKTCKGPDGELHANLICPVDMADIPEGTLPDKQSTK
Myotis_davidii                  SVPPYHPSSEKYENTKTCKGPDGELHANLICPVDMADIPEGTLPDKQSTK
Pteropus_alecto                 SIPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVADVPEGTLPDKQSTE
Felis_catus                     SFPPYHPSSEKYENTKTCRGPDGQLHANLLCPVDVADVPEGTLPDKQSTE
Ailuropoda_melanoleuca          SIPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVADVPEGTLPDKQSTE
Canis_lupus_familiaris          SIPPYHPSSEKYENTKTCRGPDGELHANLLCPVDIADVPEGTLPDKQSTE
Neovison_vison                  SVPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVADVPEGTLPDKQSTE
                                *. *****:      ***:* **:**:**:.      .*...  * ***:

                                251                                 300
Homo_sapiens                    QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Pan_troglodytes                 QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Gorilla_gorilla_gorilla         QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Nomascus_leucogenys             QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Pongo_abelii                    QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Macaca_mulatta                  QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDS
Macaca_fascicularis             QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDS
Papio_anubis                    QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDS
Chlorocebus_sabaeus             QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDS
Callithrix_jacchus              EAIRLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Neotoma_lepida                  EAIRLLEKMKTSSSPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Mesocricetus_auratus            EAIRLLEKMKTSASPFFLVVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Rattus_norvegicus               EAIRLLEKMKTSVSPFFLAVGYHKPHIPFRYPKEFQKLYPLENMTLAPDP
Mus_musculus                    EAIRLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Dipodomys_ordii                 EAIRLLEKMKTLSSPFFLAIGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Fukomys_damarensis              EAIRLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Heterocephalus_glaber           EAIRLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Cavia_porcellus                 EAICLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Bos_taurus                      QAIQLLGKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENVSLAPDP
Bos_mutus                       QAIQLLGKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENVSLAPDP
Cervus_elaphus_hippelaphus      QAIQLLGKMKTLASPFFLAVGYHKPHIPFRYPKEFQKLYPLENVSLAPDP
Desmodus_rotundus               QAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Myotis_lucifugus                KAIQLLEKMKTSSSPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Myotis_brandtii                 KAIQLLEKMKTSSSPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Myotis_davidii                  KAIQLLEKMKTSSSPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Pteropus_alecto                 QAIRLLEKMTSSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Felis_catus                     QAIRLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Ailuropoda_melanoleuca          QAIRLLEKTKTSARPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Canis_lupus_familiaris          QAIRLLEKTKTSTRPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
Neovison_vison                  QAIRLLEKMKTSAHPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDP
                                :** ** * .:   ****.:**********************::****.
```

```
                                301                                 350
Homo_sapiens                    EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Pan_troglodytes                 EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Gorilla_gorilla_gorilla         EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDFQQRKIRQSYFA
Nomascus_leucogenys             EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Pongo_abelii                    EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QQKIRQSYFA
Macaca_mulatta                  EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVEF-QRKIRQSYFA
Macaca_fascicularis             EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVEF-QRKIRQSYFA
Papio_anubis                    EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVEF-QRKIRQSYFA
Chlorocebus_sabaeus             EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVEF-QRKIRQSYFA
Callithrix_jacchus              EVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Neotoma_lepida                  QVPDGLPPVAYNPWMDIREREDVQALNISVPYGPISVDF-QRKIRQSYFA
Mesocricetus_auratus            QVPDGLPPVAYNPWMDIREREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Rattus_norvegicus               HVPDSLPPVAYNPWMDIREREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Mus_musculus                    HVPDSLPPVAYNPWMDIREREDVQALNISVPYGPIPEDF-QRKIRQSYFA
Dipodomys_ordii                 QVPGGLPPVAYNPWMDIRQREDVKALNISVPYGPIPVDF-QQKIRQSYFA
Fukomys_damarensis              EVPDGLPAVAYNPWMDIRQREDVQALNISIPYGPIPVDF-QRKIRQSYFA
Heterocephalus_glaber           EVPDSLPAVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Cavia_porcellus                 EVPDSLPAVAYNPWMDIRQREDVQALNISVPYGPIPVDF-QRKIRQSYFA
Bos_taurus                      QVPTGLPLVAYNPWMDLRQREDIQALNLSVPYGPIPADF-QRKIRQSYFA
Bos_mutus                       QVPTGLPLVAYNPWMDLRQREDIQALNLSVPYGPIP--F-QRKIRQSYFA
Cervus_elaphus_hippelaphus      QVPAGLPLVAYNPWMDLRQREDVQALNLSVPYGPIPADF-QRKIRQSYFA
Desmodus_rotundus               QVPAGLPPVAYNPWMDIRLREDIQALNLSVPYGPIPVDF-QRKIRQSYFA
Myotis_lucifugus                QVPPGLPPVAYNPWMDIRQREDVEALNLSVPYGSIPVDF-QRKIRQSYFA
Myotis_brandtii                 QVPPGLPPVAYNPWMDIRQREDVEALNLSVPYGSIPVDF-QRKIRQSYFA
Myotis_davidii                  QVPPGLPPVAYNPWMDIRQREDVEALNLSVPYGSIPVDF-QRKIRQSYFA
Pteropus_alecto                 QVPAGLPPVAYNPWMDIRQREDVKALNLSVPYGPIPVDF-QRKIRQSYFA
Felis_catus                     QVPAGLPPVAYNPWMDIRQREDVQALNLSVPYGPIPVDF-QRKIRQSYFA
Ailuropoda_melanoleuca          QVPAGLPPVAYNPWMDIRQREDVQALNLSVPYGPIPVDF-QRKIRQSYFA
Canis_lupus_familiaris          EVPAGLPPVAYNPWMDIRQREDVQALNLSVPYGPIPVDF-QRKIRQSYFA
Neovison_vison                  QVPAGLPPVAYNPWMDIRQREDVQALNLSVPYGPIPVDF-QRKIRQSYFA
                                .** .** ********:* ***::***:*:***.*.  * *:********

                                351                                 400
Homo_sapiens                    SVSYLDTQVGRLLSALDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Pan_troglodytes                 SVSYLDTQVGRLLSALDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Gorilla_gorilla_gorilla         SVSYLDTQVGRLLSALDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Nomascus_leucogenys             SVSYLDTQVGRLLSALDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Pongo_abelii                    SVSYLDTQVGRLLSTLDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Macaca_mulatta                  SVSYLDTQVGRLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Macaca_fascicularis             SVSYLDTQVGHLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Papio_anubis                    SVSYLDTQVGRLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Chlorocebus_sabaeus             SVSYLDTQVGRLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Callithrix_jacchus              SVSYLDTQVGHLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Neotoma_lepida                  SVSYLDTQVGHLLSALDDLHLANNTIIAFTSDH--GWALGEHGEWAKYSN
Mesocricetus_auratus            SVSYLDTEVGHLLSALDDLHLANNTIIAFTSDH--GWALGEHGEWAKYSN
Rattus_norvegicus               SVSYLDTQVGHLLSALDDLRLAHNTIIAFMSDH--GWALGEHGEWAKYSN
Mus_musculus                    SVSYLDTQVGHVLSALDDLRLAHNTIIAFTSDH--GWALGEHGEWAKYSN
Dipodomys_ordii                 SVSYLDTQVGHLLGALDDLELANRTIIAFMSDH--GWSLGEHGEWAKYSN
Fukomys_damarensis              SVSYLDTQVGRLLSALDDLQLANSTIIAFTSDH--GWALGEHGEWAKYSN
Heterocephalus_glaber           SVSYLDTQVGRLLSALDDLQLANSTIIALTSDH--GWALGEHGEWAKYSN
Cavia_porcellus                 SVSYLDTQVGHLLSALDDLQLANNTIIAFTSDH--GWALGEHGEWAKYSN
Bos_taurus                      CVSYLDTQVGRLLSALDDLQLASSTIVAFTSDH--GWALGEHGEWAKYSN
Bos_mutus                       CVSYLDTQVGRLLSALDDLQLASSTIVAFTSDH--GWALGEHGEWAKYSN
Cervus_elaphus_hippelaphus      CVSYLDTQVGHLLSALDDLQLASSTIVALTSDH--GWALGEHGEWAKYSN
Desmodus_rotundus               SVSYLDTQVGHLLSALDDFHLANSTIVAFTSDH--GWSLGEHGEWAKYSN
Myotis_lucifugus                SVSYMDTQVGHLLSALDDLQLANSTIVAFTSDHEPGWALGEHGEWAKYSN
Myotis_brandtii                 SVSYMDTQVGHLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Myotis_davidii                  SVSYMDSQVGHLLSALDDLQLANSTIVAFTSDH--GWALGEHGEWAKYSN
Pteropus_alecto                 SVSYLDTQVGHLLSSLDDLHLANNTIVAFTSDH--GWALGEHGEWAKYSN
Felis_catus                     SVSYLDTQVGHLLSALDDLQLTNSTIIAFASDH--GWALGEHGEWAKYSN
Ailuropoda_melanoleuca          SISYLDTQVGHLLSALDDLELANSTIIVFVSDH--GWALGEHGEWAKYSN
Canis_lupus_familiaris          SISYLDTQVGHLLSALDDLQLANSTIIVFASDH--GWALGEHGEWAKYSN
Neovison_vison                  SVSYLDTQVGHLLSALDDLQLANSTIIAFVSDH--GWALGEHGEWAKYSN
                                .:**:*::**::*.:***:.*:  **:.: ***  **:************
```

```
                                401                                 450
Homo_sapiens                    FDVATHVPLIFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDL
Pan_troglodytes                 FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDL
Gorilla_gorilla_gorilla         FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDL
Nomascus_leucogenys             FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Pongo_abelii                    FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Macaca_mulatta                  FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Macaca_fascicularis             FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Papio_anubis                    FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Chlorocebus_sabaeus             FDVATHVPLMFYVPGRTASLPEAGEKLFPYLDPFDSASELMEPGRQSMDL
Callithrix_jacchus              FDVATRVPLMFYVPGRTASLPEADEKLFPYVDPFHSASELMEPGRQSMDL
Neotoma_lepida                  FDVTTRVPLMFYVPGKTAPLPGVDEKLFPYQDPFNLTSKLMDPGRHIVDL
Mesocricetus_auratus            FDVTTRVPLMFYVPGKTAPLPGPGEKLFPYQDPFDPTSKSMDSGRHVVDL
Rattus_norvegicus               FDVTTRVPLMFYVPGKTALLPAAGEKLFPYQDPFDPASELMGAGRRTEDL
Mus_musculus                    FDVATRVPLMLYVPGRTAPLPAAGQKLFPYRDPFDPASDWMDAGRHTEDL
Dipodomys_ordii                 FDVTTRVPLMFYVPGKTAPLPEAGQKLFPYRDPFDLASQWMAPGRQSWDL
Fukomys_damarensis              FDVTTRVPLMFYVPGRTTSFPEAGEMLFPYLDPFDSASQLMEPGRQATDL
Heterocephalus_glaber           FDVTTRVPLMFYVPGRTTSFLEAGEMLFPYLDPFDSASELMEPGRQTTDL
Cavia_porcellus                 FDVTTRVPLMFYVPGRTTSFPEAGKMLFPYLDPFDSASQSMEPGMQTTDL
Bos_taurus                      FDVTTRVPLMFYVPGRTAPLPVEGEKFFPYLDPFDSVSEVPEPGMEMQNL
Bos_mutus                       FDVTTRVPLMFYVPGRTAPLPVEGEKFFPYLDPFDSVSEVPEPGRQTGDL
Cervus_elaphus_hippelaphus      FDVTTRVPLMFYVPGRTAPLPVEGEKFFPYLDPFDSISEAPEPGRQTEDL
Desmodus_rotundus               FDVATHVPLMFYVPGRTAVLPEAGEKLFPYINPFDPASELMEPGQQAVDL
Myotis_lucifugus                FDVTTHVPLMFYVPGRTAPLPEAGEKLFPYIDPFDSISELMEPGRQVQDL
Myotis_brandtii                 FDVTTHVPLMFYVPGRTAPLPEAGEKLFPYIDPFDSISELMEPGRQVQDL
Myotis_davidii                  FDVTTHVPLMFYVPGRTALLPEAGEKLFPYIDPFDSISELMEPGRQVEDL
Pteropus_alecto                 FDITTRVPLMFYVPGRTAPLPEAGKKLFPYIDPFDS-SELMEPGRRAEDL
Felis_catus                     FDVATRVPLMFYVPGRTAPLPEAGEELFPYIDPFDSVLESMEPGRRVMDL
Ailuropoda_melanoleuca          FDVTTRVPLMFYVPGRTAPLPEAGEKLFPFVDPFDSVLERMEPGRQVMDL
Canis_lupus_familiaris          FDITTRVPLMFYVPGRTAPLPEAGEKLFPYIDPFSSVQELMEPGRQVTDL
Neovison_vison                  FDVATRVPLMFYVPGRTAPLLEAGEKLFPYIDPFDSIVELMEPGRQVVDL
                                **::*:***::****:*: :   .: :**: :**    .   .* .  :*

                                451                                 500
Homo_sapiens                    VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Pan_troglodytes                 VELVSLFPTLAGLAGLQAPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Gorilla_gorilla_gorilla         VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Nomascus_leucogenys             VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Pongo_abelii                    VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Macaca_mulatta                  VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Macaca_fascicularis             VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Papio_anubis                    VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Chlorocebus_sabaeus             VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEED
Callithrix_jacchus              VELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKSLLKHFRFHGLEED
Neotoma_lepida                  VELVSLFPTLAGLAGLQVPPRCPIPSFHVELCREGQNLQKHLQLRDLEED
Mesocricetus_auratus            VELVSLFPTLAELSGLQVPPRCPIPSFHVELCREGQSLQKHLQLHDFEEE
Rattus_norvegicus               VELVSLFPTLAGLAGMPVPPWCPIPSFHVELCREGHNLQRYLQLHDLEED
Mus_musculus                    VELVSLFPTLAGLAGLPVPPRCPIPSFHVELCREGQNLQKHLQLHDLEEE
Dipodomys_ordii                 VELMSLFPTLAGLAGLQVPPRCPIPSFHIELCREGQSLLNHFRFHDWEED
Fukomys_damarensis              VELVSLFPTLTGLAGLHVPPQCPVPSFDVELCREGEDLLKHFRSRDLEED
Heterocephalus_glaber           VELVSLFPTLTGLAGLHVPPQCPVPSFDVEVCREGDNLLKHFRSRDLEED
Cavia_porcellus                 VELVSLFPTLASLAGLPILPPCPVPSFHVELCREGENLLKHFGSHDLEED
Bos_taurus                      LHALSSFMSLIGLSFLLRPRHCKDPK-RLELWPRGRNLLKLFEDHGEEEE
Bos_mutus                       VELLSLFPTLAGLAGLRVPPRCPIPSFHMELCREGRNLLKLFEDHGEEEE
Cervus_elaphus_hippelaphus      VELLSLFPTLTGLAGLRVPPRCPIPSFHMELCREGRNLLKLFEGRGEEEE
Desmodus_rotundus               VELVSLFPTLTGLAGLHVPPRCPIPSFHVELCREGQSLLKHFRFHDLEED
Myotis_lucifugus                VELVSLFPTLAGLAGLHVPPRCPSPSFHVELCREGHSLLNHFQFHDLEED
Myotis_brandtii                 VELVSLFPTLAGLAGLHVPPRCPSPSFHVELCREGHSLLNHFQFHDLEED
Myotis_davidii                  VELVSLFPTLAGLAGLHVPPRCPSPSFHVELCREGQSLLNHFQFHDLEED
Pteropus_alecto                 VELVSLFPTLAGLAGLRVPPRCPVPSFHVELCREGQSLGKHFRPHDLEED
Felis_catus                     VELLSLSPTLAGLAGLHVPPRCPVPSFHVQLCREGQNLGKHFQLQDLGED
Ailuropoda_melanoleuca          VELLSLSPTLAGLAGLHVPPRCPVPSFHVKLCREGQSLVKHFRFQDLEDD
Canis_lupus_familiaris          VELLSLSPTLAGLAGLHVPPRCPVPSFHVELCREGQNLMKHFQVEDVEGD
Neovison_vison                  VELLSLFPTLAGLAGLRVPPRCPVPSFRVAQCREGQSLVKYFRFQDLDED
                                :. :*   :*  *: :     *  *.  :    .* .* . :  ..   :
```

```
                              501                                                 550
Homo_sapiens                  PYLPGNPRELIAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Pan_troglodytes               PYLPGNPRELIAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Gorilla_gorilla_gorilla       PYLPGNPRELIAYSQYPRPADIPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Nomascus_leucogenys           PYLPGNPRELIAYSQYPRPADIPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Pongo_abelii                  PYLPGNPRELIAYSQYPRPADIPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Macaca_mulatta                PYLPGNPRELIAYSQYPRPADFPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Macaca_fascicularis           PYLPGNPRELIAYSQYPRPADFPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Papio_anubis                  PYLPGNPRELIAYSQYPRPADFPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Chlorocebus_sabaeus           PYLPGNPRELIAYSQYPRPADFPQWNSDKPNLKDIKIMGYSIRTIDYRYT
Callithrix_jacchus            PYLPGNPRELIAYSQYPRPADFPQWNSDKPSLKYIKIMGYSIRTVDYRYT
Neotoma_lepida                PHFLGNPRELIAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTVDYRYT
Mesocricetus_auratus          LHFLGNPRESIAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Rattus_norvegicus             PYLFGNPRELIAYSQYPRPADFPQWNSDKPSLKDIKVMGYSIRTIDYRYT
Mus_musculus                  PDLFGNPRELIAYSQYPRPADFPQWNSDKPSLNDIKVMGYSIRTVDYRYT
Dipodomys_ordii               PHLHGNPHESVAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Fukomys_damarensis            PVLHDNPRELVAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Heterocephalus_glaber         PDLHDNPRESVAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTVDYRYT
Cavia_porcellus               PELRGNPRELVAYSQYPRPADTPQWNSDNPSLKDIKVMGYSIRTIDYRYT
Bos_taurus                    PHVRSNPRELVAYSQYPRPADSPQWNSDKPSLEDIKVMGYSIRTVDYRYT
Bos_mutus                     PHVRSNPRELVAYSQYPRPADSPQWNSDKPSLEDIKVMGYSIRTVDYRYT
Cervus_elaphus_hippelaphus    PHVRGNPRELVAYSQYPRPADSPQWNSDKPSLEDIKVMGYSIRTVDYRYT
Desmodus_rotundus             LTLHDNPHESIAYSQYPRPANYPQWNSDKPSLKDIKVMGYSIRTTDYRYT
Myotis_lucifugus              LTLRGNPRESIAYSQYPRPADYPQWNSDKPSLKDIKFMGYSIRTIDYRYT
Myotis_brandtii               LTLRGNPRESIAYSQYPRPADYPQWNSDKPSLKDIKFMGYSIRTIDYRYT
Myotis_davidii                LTHRGNPRESIAYSQYPRPADYPQWNSDKPSLKDIKFMGYSIRTIDYRYT
Pteropus_alecto               PALHGDPRESIAYSQYPRPADSPQWNSDKPSLKDIKVMGYSIRTIDYRFT
Felis_catus                   PHLRGNPRESIAYSQYPRPADYPQWNSDKPSLKDIKVMGYSIRTIDYRYT
Ailuropoda_melanoleuca        PSLRGNPRESIAYSQYPRPADSPQWNSDKPSLKDIKIMGYSIRTIDYRYT
Canis_lupus_familiaris        PHLRGNPRESIAYSQYPRPADSPQWNSDKPSLKDIKVMGYSIRTIDYRYT
Neovison_vison                QYLRGNPRESVAYSQYPRPADSPQWNSDKPSLKDIKVMGYSIRTVDYRYT
                                  .:*:* :********:: ******:*.*: **.******* ***:*

                              551                                                 606
Homo_sapiens                  VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Pan_troglodytes               VWIGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Gorilla_gorilla_gorilla       VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Nomascus_leucogenys           VWVGFSPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Pongo_abelii                  VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Macaca_mulatta                VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Macaca_fascicularis           VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Papio_anubis                  VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Chlorocebus_sabaeus           VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP------
Callithrix_jacchus            VWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQGGELFQSLMP------
Neotoma_lepida                VWVGFNPAEFLANFSDIHAGELYFVDSDPLQDHNVYNDSQHGGLPYSLKP------
Mesocricetus_auratu           VWVGFNPAEFLANFSDIHAGELYFVDSDPLQDHNVYNDSQHGGLPHSLNP------
Rattus_norvegicus             VWVGFSPTEFLANFSDIHAGELYFVDSDPLQDHNVYNDSQHGGLLLSLRP------
Mus_musculus                  VWVGFDPSEFLANFSDIHAGELYFVDSDPLQDHNVYNDSQHGGLLHSLRP------
Dipodomys_ordii               VWVGFNPDEFLANFTDIHAGELYFVDSDPLQDHNVYNDSQHREFLQSLIP------
Fukomys_damarensis            VWVGFDPDEFLANFSDIHAGELYFVDSDPLQDHNIYNYSQHGEIPPSLMP------
Heterocephalus_glaber         VWVGFDPDEFLANFSDIHAGELYFVDSDPLQDHNIYNYSQHGEIPPSLMP------
Cavia_porcellus               VWVGFDPHEFLANFSDIHAGELYFVDSDPLQDHNIYNYSQHGEIPPSLMP------
Bos_taurus                    VWVGFNPQEFLANFSDIHAGELYFVASDPLQDHNVYNDSQGGTLPQSFMP------
Bos_mutus                     VWVGFNPQEFLANFSDIHAGELYFVASDPLQDHNVYNDSQGGTLPQSFMP------
Cervus_elaphus_hippelaphus    VWVGFNPQEFLANFSDIHAGELYFVASDPLQDHNVYNDSLGETLPQSFIP------
Desmodus_rotundus             VWVGFNPHEFLANFSDVHAGELYFVDTDPLQDHNVYNDFQGGDLLQSLIP------
Myotis_lucifugus              VWVSFNPHEFLANFSDVHAGELYFVDSDPLQDYNMYNETQGGDLPQLLTP------
Myotis_brandtii               VWVSFNPHEFLANFSDVHAGELYFVDSDPLQDHNMYNETQGGDLPQLLTP------
Myotis_davidii                VWVSFNPHEFLANFSDVHAGELYFVDSDPLQDYNMYNETQGGDLPQLLTP------
Pteropus_alecto               VWVGFSPHEFLANFSDVHAGELYFVDSDPLQDHNMYNDTQGEDLPRLLAP------
Felis_catus                   VWVGFDPREFLANFSDVHAGELYFVDSDPLQDHNMYNDSQGRDLLQSLMP------
Ailuropoda_melanoleuca        VWVGFSPHEFLANFSDVHAGELYFVDSDPLQDHNMYNDSQGRDLLRSLMP------
Canis_lupus_familiaris        VWVGFSPHEFLANFSDVHAGELYFVDSDPLQDHNMYNDSQGKDLLRAFDAIFLANH
Neovison_vison                VWVGFNPREFLANFSDVHAGELYFIDSDPLQDHNMYNGSWGRDLLQSLLP------
                              **:.*.* ******:*:*******: :*****:*:**      :   : .
```

B

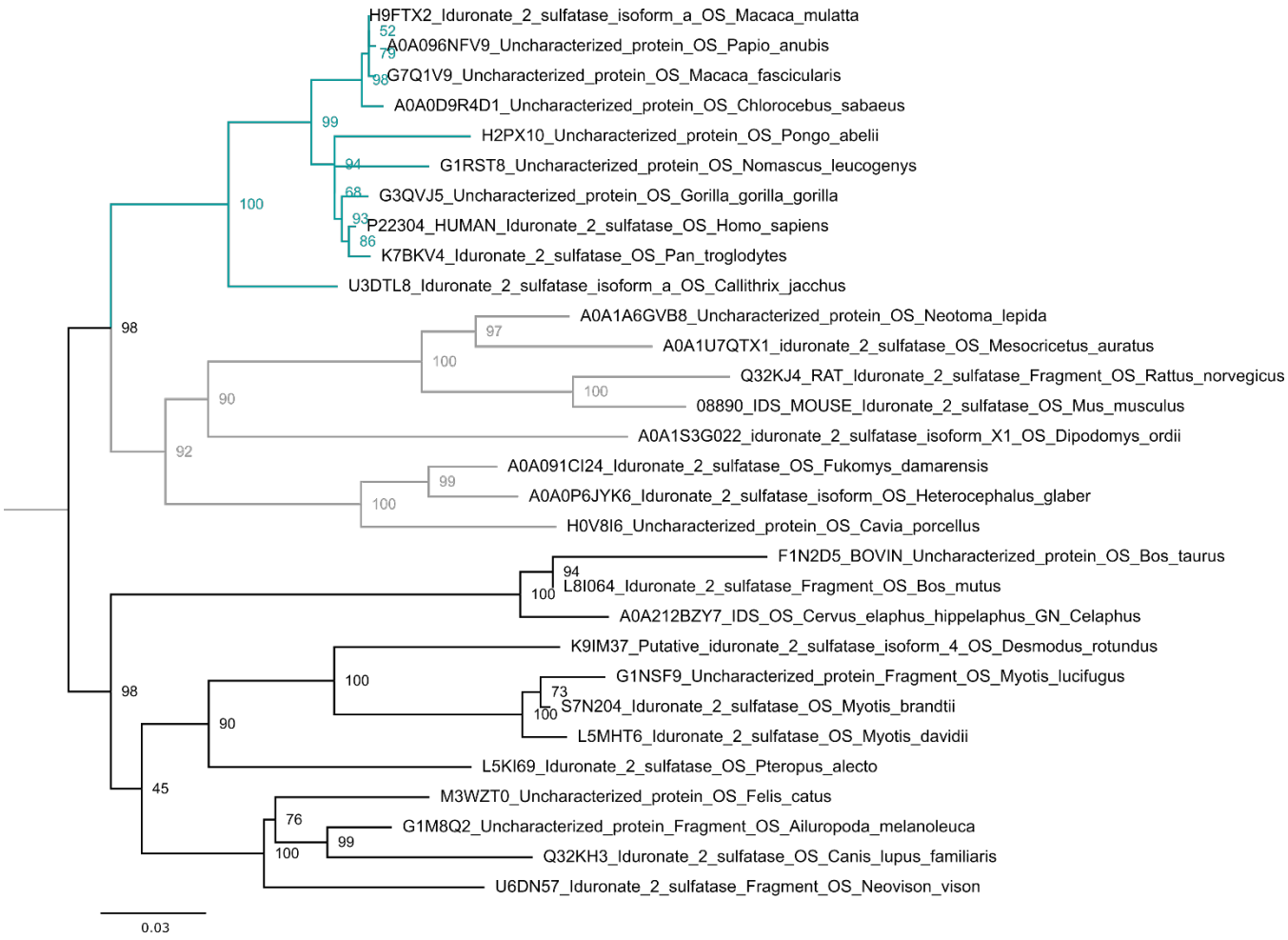


Figure S2: Multiple sequence alignment and phylogenetic tree of mammalian IDS homologues, related to Figure 1. A) Full-length multiple sequence alignment of 30 mammalian IDS homologues that was used for ancestral sequence reconstruction. The positions that were excluded by trimming (as described in the Transparent Methods section) are shaded in grey. **B)** Maximum Likelihood tree of the same set of sequences created with IQTREE (Nguyen et al., 2015). Bootstrap values are shown for 1000 replicates and the colored clades are primates (green) and rodents (grey).

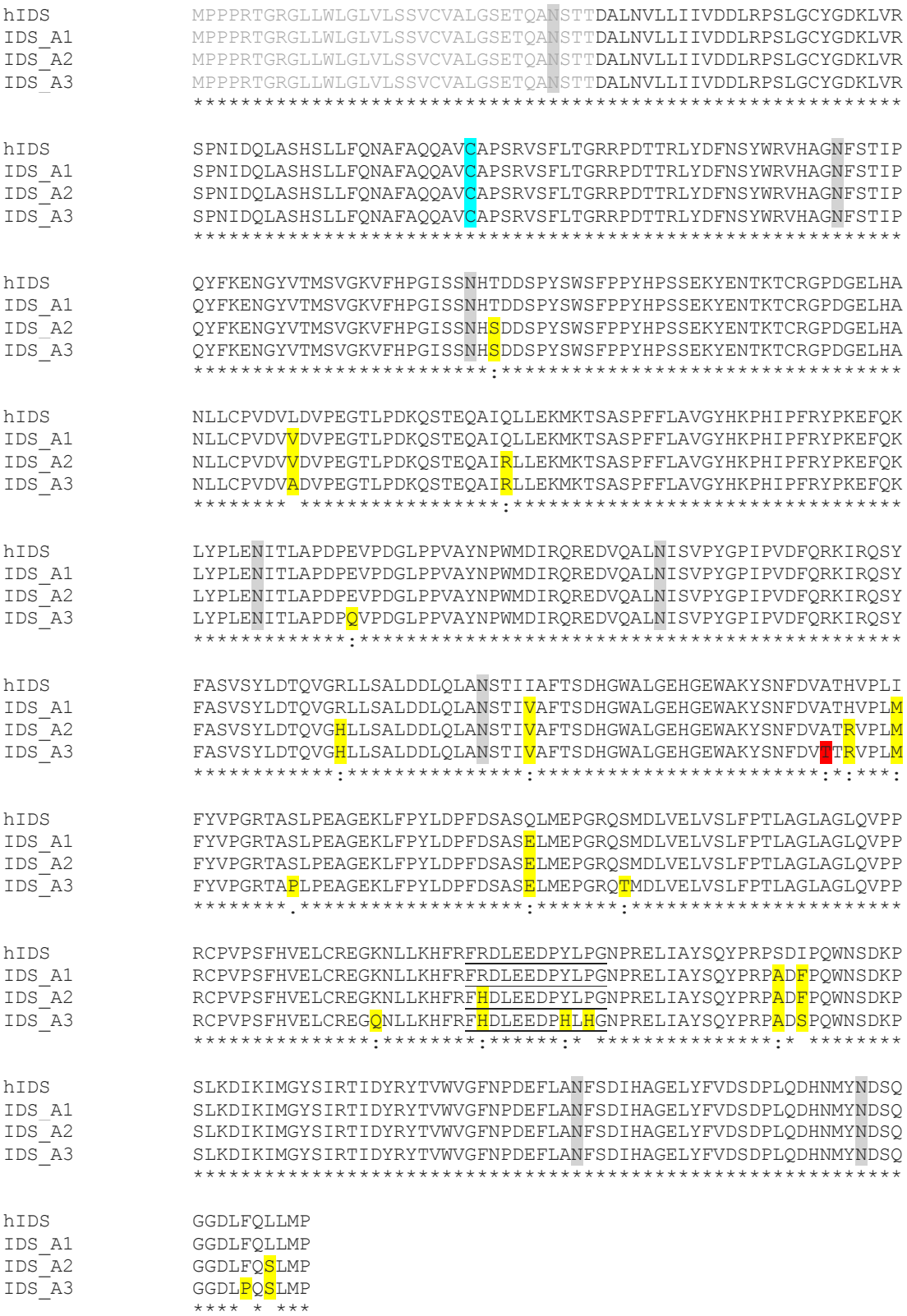


Figure S3: Alignment of hIDS and ancestors IDS-A1, IDS-A2 and IDS-A3, related to Figure 2. All ancestral mutations are marked in yellow, except for mutation A354T that is marked in red. Cysteine 84 that is post-translationally modified to formylglycine is marked in cyan and all known glycosylation sites are market in grey. The signal sequence is shown in grey and the sequence that is susceptible to proteolytic cleavage in the lysosome is underlined. The alignment was created using MAFFT.

Table S1: IDS variants that were evaluated in this study, related to Figure 2 and 3.

	# Mutations (compared to hIDS)[a]	% Identity (compared to hIDS)[a]	% Identity (compared to mIDS)[a]
hIDS_A354T	1	99.8	88.2
hIDS_H456R	1	99.8	88.2
hIDS_A354T_H354R	2	99.6	88.4
IDS-A1	6	98.8	88.8
IDS-A1_V329I	5	99.0	89.0
IDS-A2	12	97.7	89.9
IDS-A3	20	96.1	90.3
IDS-A3_V329I	19	96.3	90.5
IDS-A3_T354A	19	96.3	90.5
mIDS	60	88.4	100

[a] Signal sequences are excluded

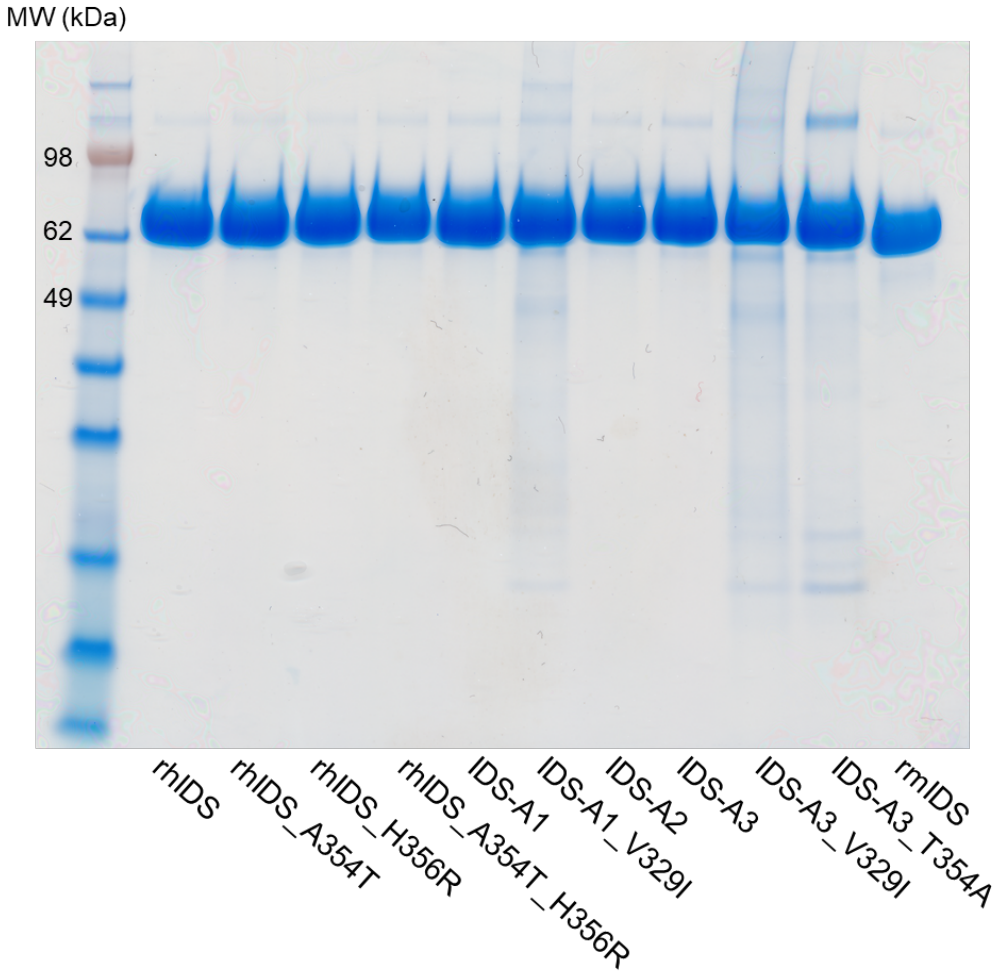


Figure S4: IDS enzymes can be functionally expressed in ExpiCHO, related to Figure 3. SDS-PAGE showing batches of purified IDS enzymes; the molecular weight of glycosylated IDS enzymes is around 76 kDa. The SeeBlue™ Plus2 pre-stained protein standard is shown for reference and 10 µg of protein was loaded for each enzyme.

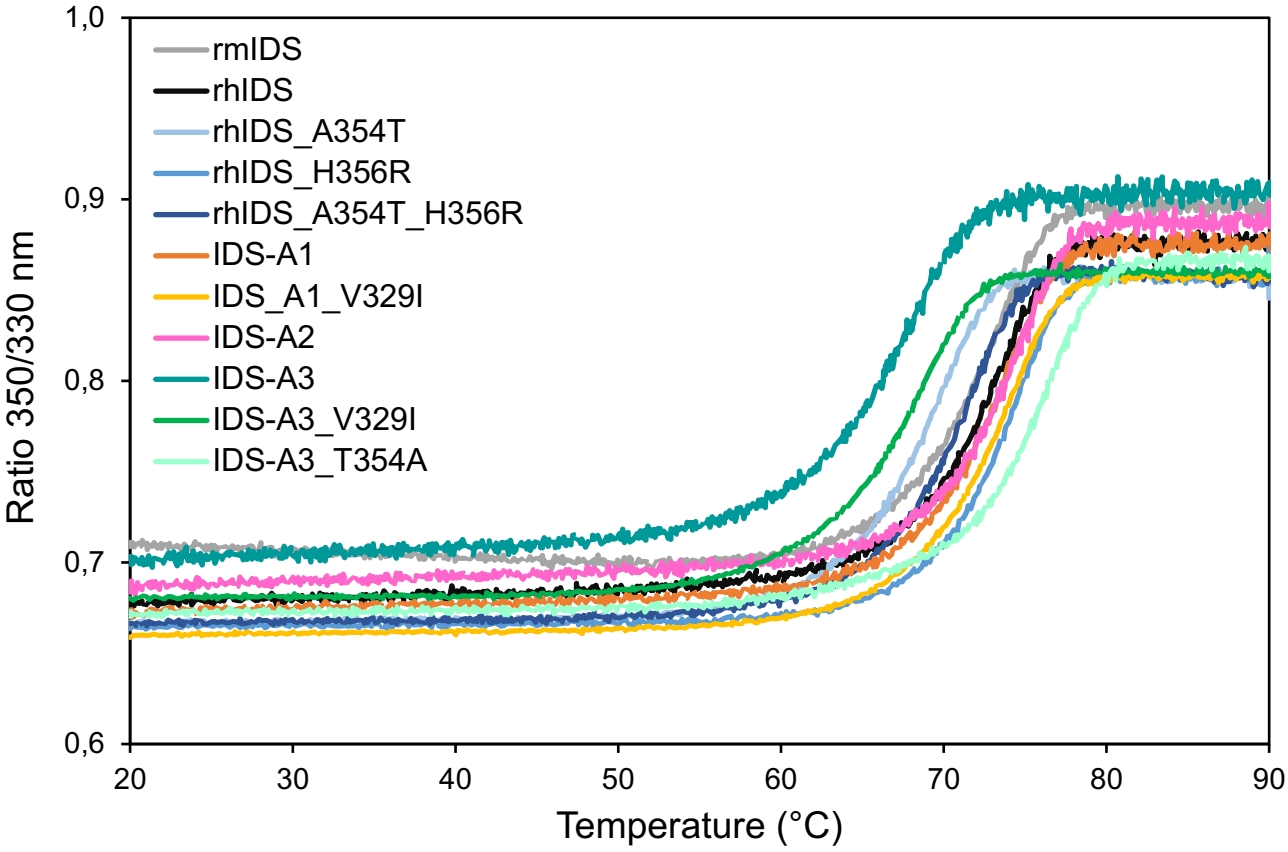


Figure S5: Melting curves of IDS enzymes established by nanoDSF, related to Table 1. The ratio of intrinsic tryptophan fluorescence at 350 over 330 nm was monitored while heating the samples with 1 degree per minute from 20 °C to 90 °C. The melting temperature is determined as the maximum of the first derivative of the ratio.

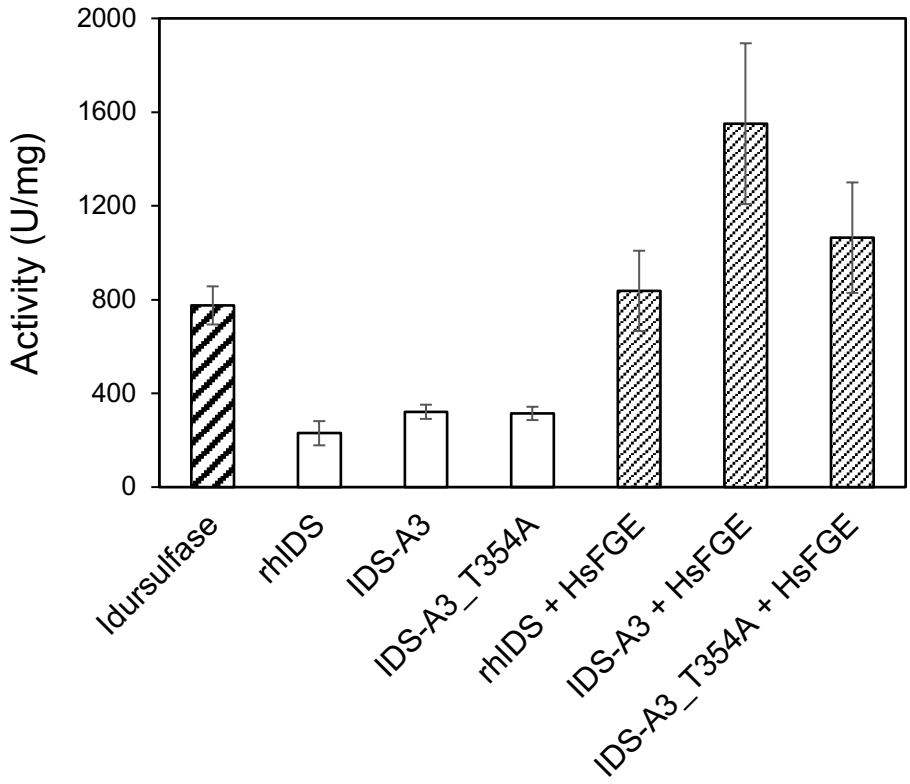


Figure S6: Activity of hIDS, IDS-A3 and variants, related to Figure 3. Comparison of activities of hIDS, IDS-A3 and IDS-A3_T354A with and without co-expression of *Hs*FGE, Idursulfase is included for reference. Average activities are shown for two independent transfection experiments for each variant and error bars show the standard deviations from 6-12 replicates.

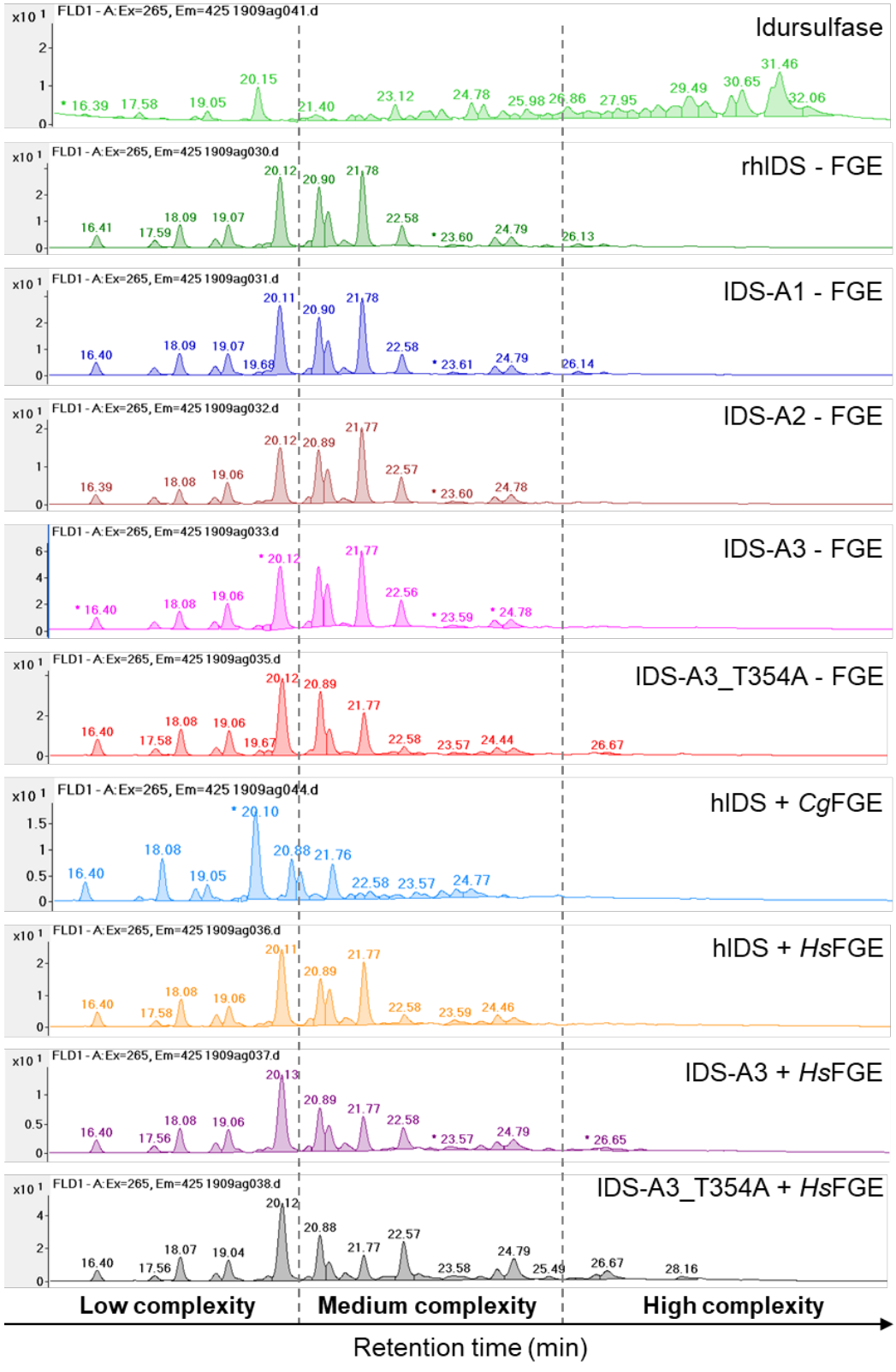


Figure S7: Glycan profiles of IDS enzymes, related to Figure 5. Profiles of released N-glycans as determined by the GlycoWorks *Rapi*Fluor-MS N-Glycan Kit. Glycans were divided into three categories for comparison, based on retention time in LC-FLR-MS: simple, semi-mature and mature glycans (shown on bottom).

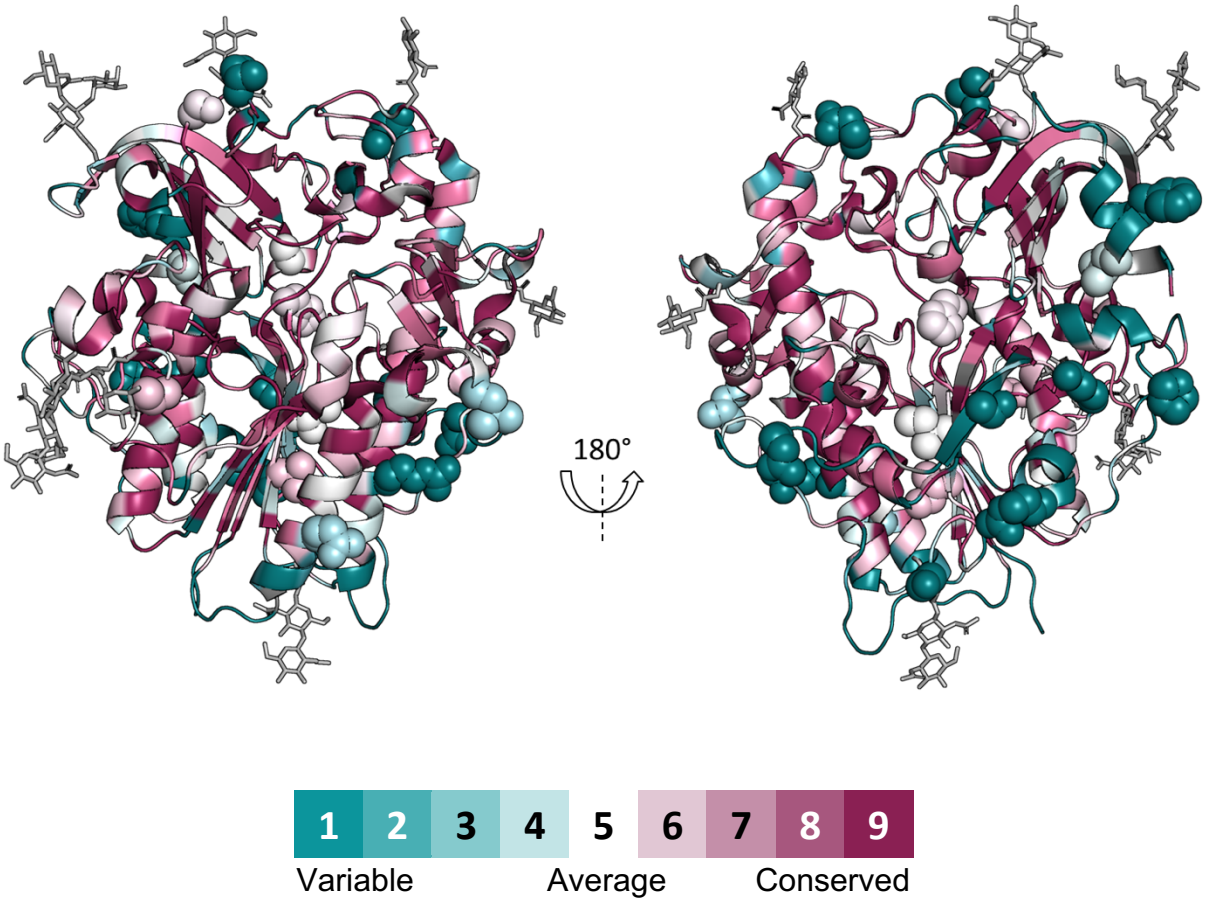


Figure S8: Analysis of evolutionary conservation in IDS alignment using ConSurf, related to Figure 1 and Figure 2. Conservation for each site in the IDS alignment that was used for ancestral reconstruction was analyzed using ConSurf. The structure of human IDS (PDB code: 5FQL (Demydchuk et al., 2017)) is shown (front view and back view) and colored according to conservation on a scale from 1-9, where 1 represents highly variable positions and 9 represents highly conserved positions. All 20 residues that are mutated in the ancestors are shown as spheres in their respective colors. Almost all ancestral mutations occur in positions with conservation score 1, apart from H356R (score: 4), T146S (score: 5) and I360 and S470 (score: 6). Glycans are shown as grey sticks.

Transparent Methods

Ancestral sequence reconstruction

A protein BLAST search was performed via NCBI in the non-redundant database using the sequence of hIDS (GenBank accession no. AAA16877.1) as a query. The 100 closest mammalian homologues were selected and duplicates, incomplete sequences, recombinant proteins and mutants were removed. The remaining 30 sequences were aligned in MAFFT version 7 (Katoh, 2002; Katoh and Standley, 2013) using the L-INS-i algorithm and were trimmed using trimAl (Capella-Gutiérrez et al., 2009) and the implemented "gappyout" method, which resulted in removal of approximately 10% of the positions. Model testing was performed in IQ-TREE (Nguyen et al., 2015) and MEGA7 (Kumar et al., 2016) and both found the Jones-Taylor-Thornton (JTT) model (Jones et al., 1992) to be the best evolutionary model for the dataset, including a gamma distribution (4 categories) to model rate variation across sites. The same model was found when performing the test with the full-length sequence alignment. A maximum likelihood tree was constructed in IQTREE with 1000 bootstrap replicates and the settings from the model test. We inferred the most probable ancestral sequences in PAML version 4 (Yang, 2007) at three different nodes, going back from hIDS to the last common ancestor of primates and rodents.

Homology modeling & MD simulations

A full-length model of hIDS was constructed using the homology modelling feature in ICM (Molsoft LLC) (Cardozo et al., 1995). The missing loop in the crystal structure (PDB: 5FQL (Demydchuk et al., 2017)) was added by sampling loop conformations from the PDB, followed by energy minimization. To allow for efficient MD analysis the fGly residue in position 84 was mutated to a serine, which has been shown to perform the same function as fGly in the coordination of the calcium ion in the related enzyme Arylsulfatase A, but rendering the enzyme inactive (von Bülow et al., 2001). A homology model for IDS-A3_T354A was built based on the full-length hIDS model by manually replacing the differing residues followed by energy minimization. Both models were subjected to MD simulations in the GROMACS software version 2018.4 (Pronk et al., 2013) using the Amber ff99SB-ILDN force field (Lindorff-Larsen et al., 2010). A cubic box was used for both systems with edges 1 nm from the protein, which was filled with SPC/E modeled water molecules (Kusalik and Svishchev, 1994). The systems were neutralized by adding 19 sodium ions, followed by energy minimization and equilibration with position restraints, a time step of 1 fs and divided in two phases; the first with an NVT ensemble for 100 ps, followed by an NPT ensemble for another 100 ps. Triplicate MD trajectories were run for each system for 100 ns at 300 K, taking the starting velocities from the NPT simulation and using no position restraints. Time steps of 2 fs were used and all run associated data were stored every 2 ps. The simulation used Particle-Mesh Ewald electrostatics (Darden et al., 1999), velocity-rescaling temperature coupling (Bussi et al., 2007) and Parrinello-Rahman pressure coupling (Parrinello and Rahman, 1981). The results were analyzed using GROMACS analytical tools.

Protein expression and purification

Sequences were codon-optimized for expression in *C. griseus*, synthesized by GeneArt (ThermoFisher Scientific) and subcloned into mammalian expression vector pcDNA3.4-TOPO. A C-terminal GGGSEPEA tag was included for affinity purification. ExpiCHO cells for transient expression (ThermoFisher Scientific) were transfected with vectors containing sequences of hIDS, mIDS, the hIDS variants and all ancestral enzymes (total plasmid concentrations of 1 µg/mL). The 50 mL cell cultures were treated according to the high titer protocol from the manufacturers manual and all reagents were from ThermoFisher Scientific. Cells were harvested 8 days after transfection by centrifugation for 30 min (4°C, 2264 x g). The supernatant was filtered through a 0.22 µm filter before being loaded onto a 1 ml CaptureSelect™ C-tagXL column (ThermoFisher Scientific) equilibrated with 20 mM Tris-HCl, 150 mM NaCl at pH 7.5, using an ÄKTA Explorer (GE Healthcare). Variants expressed without SUMF1 were

eluted with a 50 mM citrate, 150 mM NaCl buffer at pH 3.0 and elution fractions were analyzed by SDS-PAGE. Fractions containing the correctly sized protein were loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) which was equilibrated with a 20 mM Tris-HCl, 125 mM NaCl buffer at pH 7.5 and eluted with the same buffer. Elution fractions were analyzed on SDS-PAGE and fractions containing protein of the correct size were pooled, concentrated and stored at -80°C. Variants co-expressed with SUMF1 were eluted with a 20 mM Tris-HCl, 2 M $MgCl_2$ buffer at pH 7.1 and elution fractions were analyzed on SDS-PAGE. Fractions containing the correctly sized protein were loaded onto three 5 mL HiTrap Sephadex G-25 desalting columns (GE Healthcare) connected in tandem, which were equilibrated with a 20 mM Tris-HCl, 125 mM NaCl buffer at pH 7.5 and eluted with the same buffer. Elution fractions were analyzed by SDS-PAGE, and fractions containing protein of the correct size were pooled, concentrated and stored at -80°C.

Enzyme assays

IDS enzyme activity was measured in triplicates in a microplate format through a coupled reaction to **α**-L-iduronidase (Figure S1b) (Voznyi et al., 2001). All solutions and plates were pre-incubated at 37 °C and plates were sealed immediately after mixing the solutions. Reactions were initiated by mixing 10 µL IDS solution (1-2 nM, diluted with a diluent containing 0.2 % BSA, 3 mM NaN_3 and 0.05% Triton X-100) with 10 µL substrate solution (200 µM 4-MU-**α**IdoA2S and 0.1 mU **α**-L-iduronidase in a pH 4.5 buffer containing 0.1 M NaOAc in the diluent) in an incubation plate (PS96U) and were left shaking at 600 rpm at 37 °C in a microplate incubator (Thermostar, BMG Labtech). Blanks showed that 4-MU-**α**IdoA2S is not a substrate for **α**-L-iduronidase and they could be mixed before addition of IDS. The reaction was typically terminated after 35 min by addition of 100 µL stop buffer (40 mM $NaHCO_3$, 460 mM Na_2CO_3, 0.025% Triton X-100, pH 10.7). 100 µL of the terminated solution was then transferred to a reading plate (Black PS96F) and covered by a black lid before reading. The fluorescence (excitation: 355 nm, emission: 460 nm) of the solution in the well was measured by a multilabel plate reader (EnVision, PerkinElmer). The coupled reaction lag was determined to be 5 min under the assay condition. Product formation was calculated against a standard calibration curve of 4-MU, which was made in parallel with the assay wells at eight concentrations of 5 pmol interval from 0 to 35 pmol. One unit of IDS activity will hydrolyze 1.0 nmol of 4-MU-**α**IdoA2S per minute at pH 4.5, 37 °C under the defined condition.

Determination of formylglycine content

All reagents were purchased from Sigma-Aldrich unless stated otherwise and all mixtures were freshly prepared before the experiment. Enzyme samples (20 µg) were reduced, alkylated and digested with trypsin. Reduction was done by incubation in 5 µL DL-dithiothreitol (10 mM in 50 mM NH_4HCO_3) at 50 °C for 60 min. Subsequent alkylation with 5 µL iodoacetamide (55 mM in 50 mM NH_4HCO_3) was performed at room temperature in the dark for 45 min. For tryptic digestion 20 µL of 50 mM NH_4HCO_3, 5 mM $CaCl_2$ (pH 8) and trypsin (0.2 µg/µL sequencing grade, Promega) were added followed by incubation at 37 °C over night. The digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.5%. Prior to analysis 50 µL of 5% acetonitrile, 0.1% propionic acid and 0.02% TFA were added. Samples were transferred to HPLC vials and analyzed directly with liquid chromatography-mass spectrometry (LC-MS) (injection volume: 15 µL). LC separation was performed by use of a XSELECT CSH 130Å C18 column (100 x 2.1 mm, Waters) with a column temperature of 35 ^{0}C and a flow rate of 0.2 mL/min. Mobile phase A consisted of 5% acetonitrile, 0.1% propionic acid and 0.02% TFA and mobile phase B consisted of 95% acetonitrile, 0.1% propionic acid and 0.02% TFA. The gradient used was 0-5 min: 0-10% B, 5-30 min: 10-70% B, 30-35 min: 70-90% B, 35.1 min: 0% B. An Agilent 1200 HPLC system coupled to an Agilent 6545A Q-TOF-MS was used for the LC-MS analysis. During the course of data acquisition, the fragmentor voltage, skimmer voltage, and octopole RF were set to 175 V, 65 V, and 750 V, respectively. Scan range was set between 300 and 2800 m/z. The Q-TOF instrument was operated in positive electrospray ion mode and was controlled by MassHunter Workstation. Relative amounts of tryptic peptide SPNIDQLASHSLLFQNAFAQQAVCAPSR,

with the various modifications on Cys, were calculated by measuring the peak areas from reconstructed ion chromatograms of the triply charged ions. The following Cys variants were searched for: Cys (alkylated), FGly, Ser, hydroxylated Cys (sulfenic, sulfinic, sulfonic acid), Cys-SO_2-SH. No correction for ionization efficiency was done.

Thermostability measurements

Thermostability measurements were performed on a Prometheus NT.Plex nanoDSF instrument (NanoTemper Technologies) in a 20 mM Tris-HCl, 150 mM NaCl buffer at pH 7.5. Protein unfolding was monitored by following the ratio of intrinsic protein fluorescence at 350 nm to 330 nm over a temperature gradient from 20 °C to 90 °C with an increase of 1 °C per minute. The melting temperature corresponds to the maximum of the first derivative of the 350/330 ratio vs. temperature.

Analysis of cell uptake and intracellular activity

MPSII patient fibroblasts (GM00615, Coriell Institute) were seeded on 6-well plates at a density of 16000 cells/cm^2. One day after seeding, growth media was removed and treatment was added. The enzymes were diluted in growth media (+PEST) to a concentration of 2 nM and were further diluted to a series of six concentrations (2-0.01 nM for enzymes expressed without FGE and 0.2-0.001 for enzymes expressed with FGE), of which 2 mL was added to each well. The day after, cells were washed with cold PBS, trypsinized and resuspended in growth media. The mixtures were spun down and the cell pellets were washed in PBS. Cell pellets were resuspended in 100 µL 1 mM Tris (pH 7.6) for lysis, which was achieved by four freeze/thaw cycles. Debris was spun down and supernatants were saved for analysis of intracellular IDS concentration and substrate levels. The IDS concentration in the cell lysates were determined by an electrochemiluminiscence immunoassay using the Meso Scale Discovery (MSD) platform. The wells of a 96-well streptavidin gold MSD plate (#L155A-1) were blocked with 1% fish gelatin in PBS, washed with wash buffer (PBS + 0.05% Tween-20) and incubated with a biotinylated, affinity purified goat-anti-hIDS polyclonal antibody (BAF2449, R&D Systems) overnight in 4°C. After washing, different dilutions of standard and cell lysates in sample diluent (1% fish gelatin in wash buffer) were incubated in the plate shaking at 700 rpm in RT for 2 h. The plate was washed and an IDS-specific Rutenium (SULFO-TAG, MSD) tagged goat polyclonal antibody (AF2449, R&D) was added and allowed to bind to the captured IDS. The plate was washed and 2× Read Buffer (MSD) was added. The plate content was analyzed using an MSD Sector 2400 Imager Instrument.

In order to determine intracellular activity of the IDS enzymes the amount of sulfated trisaccharide substrate uronic acid – *N*-acetylglucosamine – uronic acid (UA-GlcNAc-UA(S)) was measured, which has previously been shown to be a marker for MPS (Fuller et al., 2004). Final concentrations of UA-GlcNAc-UA(S) were determined as follows: 0.1 ml 3-methyl-1-phenyl-2-pyrazolin-5-one (35 mg/ml in 0.4 M NH_3 pH 9.1 containing 47.5 % ethanol) and 20 µL of the internal standard chondroitin disaccharide Δdi-4S sodium (ΔUA-GalNAc4S, 10 µg/ml, Mw 459 Da) were added to the samples, followed by derivatization for 90 min at 70 °C. Samples were centrifuged for 6 min at 7000 rpm. The µ-WAX columns were preconditioned with 100 µl acetonitrile:10 M acetic acid (HOAc) (9:1 v:v) using low vacuum and the samples were transferred to the columns. The columns were washed with 0.2 mL methanol and eluted with 0.2 mL 0.4 M NH_3 in 60 % methanol into an LoBind 96-well plate (Eppendorf) using low vacuum. 20 µL of concentrated HOAc was added to all samples prior to injection (5 µl) into the LC-MS system. The samples were analyzed with LC–MS/MS using a QTrap 6500+ mass spectrometer equipped with an ESI interface. LC separation was performed on a Poroshell 120 EC-C18 column (2.7 µm, 50 × 2.1 mm, Agilent) with a flow rate of 0.45 ml/min. Mobile phase A consisted of 5% acetonitrile and 2% formic acid and mobile phase B consisted of 95% acetonitrile and 2% formic acid. The gradient program was equilibration of 95% A, followed by a linear decrease to 80% A (0.5 min), followed by another linear decrease to 75% A (1.7 min). The column was washed with 1% A (0.8 min) prior to re-equilibration with 95% A (3 min). Detection was carried out in negative ion mode using multiple reaction monitoring. The ionization optimization gave the following ESI settings: ion spray voltage 4.5 kV, curtain

gas 30 l/h, ion source gas1 60 l/h, ion source gas 2 60 l/h, and source temperature 500 °C. The precursor/product ions used for detection were internal standard Chondroitin disaccharide Δdi-4S (788.2/534.1) and UA-GlcNAc-UA(S) (490.5/476.0). The analytes elution times were between 1 and 3 min, the declustering potential used was -80 V and the collision energy was -18 eV. The data were acquired and processed using Analyst software (Sciex).

Analytical size exclusion chromatography

All purified enzymes were diluted to final concentrations of ca. 0.5-1.5 mg/ml in 20 mM Tris-HCl, 150 mM NaCl buffer (pH 7.5). Samples of 50 µL were loaded onto a Superdex 200 Increase 10/300 GL column equilibrated with phosphate-buffered saline (PBS, pH 7.4) and coupled to an ÄKTAmicro (GE Healthcare). Proteins were eluted using PBS (pH 7.4), and molecular weight of the proteins in the elution peaks was determined through online MALS detection, using a miniDAWN TriStar detector and ASTRA 5.3.4.20 software (Wyatt Technology).

Analysis of binding to mannose-6-phosphate receptor

Binding of IDS and its variants to the M6PR was compared by surface plasmon resonance using a Biacore T200 system (GE Healthcare). Recombinant human M6PR (R&D systems) was immobilized on a CM5 sensor chip (GE Healthcare) using an amine coupling kit (GE Healthcare) to approximately 8000 RU. IDS enzymes were diluted to a final concentration of 1 µM in running buffer (10 mM HEPES, 150 mM NaCl, 0.05 % P20, pH 7.4) and injected at 30 µl/min. Measurements were performed using the single cycle kinetics mode with an injection time of 180 s and a dissociation time of 180 s. The sensor surface was regenerated for 90 s + 30 s with regeneration buffer (10 mM NaOAc, 300 mM NaCl, pH 5.0). Results were evaluated using the Biacore T200 evaluation software v3.0 (GE Healthcare).

Analysis of glycans

An analysis of the glycans of different IDS batches was performed using the GlycoWorks *Rapi*Fluor-MS N-Glycan Kit (Waters). Enzymes were concentrated to final concentrations of 1-2 mg/mL and 15 µg of enzyme was used for the assay. The enzymes were denaturated at 90 °C for 5 min followed by deglycosylation using PNGase F (Promega, #V4831) at 37°C over night. Finally, the enzymes were labelled according to the manufacturer's instructions. Samples were analyzed on a LC-FLR-MS QToF system (Agilent), using a Waters Glycan BEH Amide column (2.1x150 mm, 130 Å 1.7 µm) with a column temperature of 52 °C and using the gradient and flow recommended in the GlycoWorks *Rapi*Fluor-MS N-Glycan protocol.

Supplemental References

von Bülow, R., Schmidt, B., Dierks, T., von Figura, K., and Usón, I. (2001). Crystal Structure of an Enzyme-Substrate Complex Provides Insight into the Interaction between Human Arylsulfatase A and its Substrates During Catalysis. J Mol Biol *305*, 269–277.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J Chem Phys *126*, 14101–14107.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Cardozo, T., Totrov, M., and Abagyan, R. (1995). Homology modeling by the ICM method. Proteins Struct. Funct. Bioinforma. *23*, 403–414.

Darden, T., Perera, L., Li, L., and Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure *7*, 55–60.

Fuller, M., Rozaklis, T., Ramsay, S.L., Hopwood, J.J., and Meikle, P.J. (2004). Disease-specific markers for the mucopolysaccharidoses. Pediatr. Res. *56*, 733–738.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Bioinformatics *8*, 275–282.

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

Kusalik, P.G., and Svishchev, I.M. (1994). The Spatial Structure in Liquid Water. Science *265*, 1219–1221.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–1958.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics *29*, 845–854.