

This is a repository copy of *Insights into Hunter syndrome from the structure of iduronate-2-sulfatase*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/179535/>

Version: Published Version

---

**Article:**

Demydchuk, Mykhaylo, Hill, Chris H [orcid.org/0000-0001-7037-0611](https://orcid.org/0000-0001-7037-0611), Zhou, Aiwu et al. (5 more authors) (2017) Insights into Hunter syndrome from the structure of iduronate-2-sulfatase. Nature Communications. 15786. ISSN 2041-1723

<https://doi.org/10.1038/ncomms15786>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

ARTICLE

Received 9 Nov 2016 | Accepted 27 Apr 2017 | Published 8 Jun 2017

DOI: 10.1038/ncomms15786

OPEN

# Insights into Hunter syndrome from the structure of iduronate-2-sulfatase

Mykhaylo Demydchuk<sup>1,\*</sup>, Chris H. Hill<sup>1,\*†</sup>, Aiwu Zhou<sup>2</sup>, Gábor Bunkóczi<sup>1</sup>, Penelope E. Stein<sup>3,†</sup>, Denis Marchesan<sup>3</sup>, Janet E. Deane<sup>1</sup> & Randy J. Read<sup>1</sup>

Hunter syndrome is a rare but devastating childhood disease caused by mutations in the *IDS* gene encoding iduronate-2-sulfatase, a crucial enzyme in the lysosomal degradation pathway of dermatan sulfate and heparan sulfate. These complex glycosaminoglycans have important roles in cell adhesion, growth, proliferation and repair, and their degradation and recycling in the lysosome is essential for cellular maintenance. A variety of disease-causing mutations have been identified throughout the *IDS* gene. However, understanding the molecular basis of the disease has been impaired by the lack of structural data. Here, we present the crystal structure of human IDS with a covalently bound sulfate ion in the active site. This structure provides essential insight into multiple mechanisms by which pathogenic mutations interfere with enzyme function, and a compelling explanation for severe Hunter syndrome phenotypes. Understanding the structural consequences of disease-associated mutations will facilitate the identification of patients that may benefit from specific tailored therapies.

<sup>1</sup>Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK.

<sup>2</sup>Key Laboratory of Cell Differentiation and Apoptosis of Ministry of Education of China, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China. <sup>3</sup>Department of Medicine, University of Cambridge, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0SP, UK. \* These authors contributed equally to this work. † Present address: Haematological Medicine, King's College Hospital NHS Foundation Trust, Denmark Hill, Camberwell, London SE5 5RS, UK (P.E.S); MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK (C.H.H). Correspondence and requests for materials should be addressed to R.J.R. (email: rjr27@cam.ac.uk).

**H**unter syndrome, also known as Mucopolysaccharidosis type II (MPS II, OMIM 309900)<sup>1</sup>, is an X-linked lysosomal storage disease caused by genetic deficiency of the enzyme iduronate-2-sulfatase (IDS, EC 3.1.6.13), required for the step-wise degradation and recycling of complex glycosaminoglycans (GAGs). IDS belongs to the sulfatase family of enzymes and catalyses hydrolysis of the C2-sulfate ester bond at the non-reducing end of 2-O-sulfo- $\alpha$ -L-iduronic acid residues in dermatan sulfate and heparan sulfate. Loss of IDS activity leads to abnormal accumulation of GAGs in multiple tissues and organs, resulting in progressive cellular and multi-organ dysfunction.

Signs and symptoms of Hunter syndrome include coarse facial features, stiff joints, skeletal abnormalities, hepatosplenomegaly, cardiovascular and respiratory disorders, developmental delay and deteriorating intellectual function. In general, the disease progression follows one of two forms. 'Severe' early-onset disease becomes apparent within 2–4 years of age, patients exhibit behavioural disturbances and progressive intellectual disability concomitant with neurodegeneration; death occurs before adulthood. In contrast, patients with the 'mild' late-onset form preserve normal or slightly impaired intelligence and survive into late adulthood<sup>2</sup>. The incidence of Hunter syndrome is  $\sim 1$  in 100,000 male live births, but varies considerably between different populations<sup>3,4</sup>. Sporadic cases have been reported in females.

Over 500 mutations at the *IDS* locus (Xq28) have been identified, including complex rearrangements, insertions/deletions, splicing defects and missense/nonsense point mutations. This molecular heterogeneity results in a broad spectrum of disease phenotypes, and combined with the rarity of individual mutations, the fact that most mutations are private/familial, and the lack of standardized severity reporting, genotype/phenotype correlations have been difficult to establish<sup>5,6</sup>. A notable exception to this is that patients harbouring gross structural changes in the *IDS* gene, caused by deletions or gene–pseudogene recombination events, usually develop severe forms of Hunter syndrome<sup>7</sup>. Prediction of the clinical phenotype can also be hindered by effects of other mutations in multiple genes that may contribute to the spectrum of disease<sup>8</sup>. However, differences in measured level of GAG synthesis (presumably also affected by genotype) can be harnessed to improve the prediction of disease phenotype<sup>9</sup>.

There is no curative treatment for Hunter syndrome that is currently approved, and management is usually symptomatic and supportive. Enzyme replacement therapy (ERT)<sup>10</sup> with recombinant human IDS (Elaprase, Shire Pharmaceuticals)<sup>11</sup> is effective in ameliorating peripheral symptoms associated with the 'mild' disease form, with recipients showing improved excretion of GAGs in urine, reduced liver and spleen size and some bone remodelling. However, ERT cannot prevent or reverse the characteristic cardiac and neurological deterioration of the 'severe' form, and for many patients treatment remains palliative<sup>12,13</sup>. Pharmacological chaperone therapy may be a useful alternative strategy for folding mutants with residual catalytic activity, as has been demonstrated for other lysosomal storage disorders<sup>14</sup>.

Here we report the X-ray crystal structure of human IDS with a sulfate ion bound covalently in the active site. Despite low sequence identity, the overall fold and domain organization bears a striking core resemblance to other human sulfatases. This structure generates insight into the catalytic mechanism and provides a framework for better understanding the consequences of disease-associated missense mutations.

## Results

**Protein characterization.** Clinical-grade IDS was provided by Shire Pharmaceuticals in the form of a  $\sim 76$  kDa glycoprotein,

recombinantly expressed in HT-1080 human fibroblasts, and purified following secretion into growth medium. Upon denaturation and treatment with PNGase F, IDS can be deglycosylated to a single  $\sim 58$  kDa species (Supplementary Fig. 1a). To verify that glycosylated samples for crystallization were catalytically active, assays were performed with the fluorogenic substrate 4-methylumbelliferyl- $\alpha$ -L-iduronide-2-sulfate (MU- $\alpha$ IdoA-2S, Supplementary Fig. 1b). Using this system, we report a specific activity of  $36 \mu\text{mol h}^{-1} \text{mg}^{-1}$  at room temperature (293 K). This is comparable to the commercially reported activity of Elaprase ( $46\text{--}74 \mu\text{mol h}^{-1} \text{mg}^{-1}$ ), as determined with heparin disaccharide substrate at 310 K.

## Overall structure and post-translational modifications.

The crystal structure of IDS was refined at 2.3 Å resolution. Extensive crystal optimization was required to obtain the best possible diffraction data, because loose crystal packing in the *ab* plane of the crystal (Supplementary Fig. 2) led to varying levels of anisotropy (Table 1). The enzyme is monomeric with a single-polypeptide chain in the asymmetric unit. IDS adopts a compact, globular  $\alpha/\beta$  sandwich fold, comprising residues 34–550 (Fig. 1). The first 33 residues, encoding the N-terminal signal peptide and propeptide, are cleaved during enzyme secretion and therefore are not present in the structure<sup>15</sup>. Lack of interpretable electron density prevented the modelling of one loop region (residues 444–453). A second shorter loop (residues 386–392) is very poorly ordered, but modelled with high local B-factors into weak electron density.

Traditionally, the IDS main chain has been described as two subdomains, SD1 and SD2, corresponding to fragments resulting from lysosomal proteolytic processing events<sup>16</sup>. The N-terminal SD1 (residues 34–443) comprises the 'heavy' 42 kDa chain, and the C-terminal SD2 (residues 455–550) corresponds to the 'light' 14 kDa chain (Fig. 1a). The lack of visible electron density in the unmodelled loop between the two subdomains (residues 444–453) is consistent with a disordered region that is susceptible to proteolytic cleavage. In our structure this loop, although disordered, is likely to remain intact as the protein is purified from conditioned medium and has not trafficked to lysosomes. This is consistent with the presence of a single-band visible by SDS-PAGE following denaturation and deglycosylation (Supplementary Fig. 1a).

SD1, which contains the catalytic core, is characterized by a central 7-stranded  $\beta$ -sheet (one antiparallel and six parallel strands) surrounded by ten  $\alpha$ -helices/helical turns and five short  $\beta$ -strands. The C-terminal SD2 consists of a twisted, 4-stranded antiparallel  $\beta$ -sheet capped by a short C-terminal  $\alpha$ -helix and flanked by five helical turns. Together, SD1 and SD2 form a large hydrophobic packing interface with a buried area of  $\sim 1,850 \text{ \AA}^2$ . Hence, they remain stably associated and together form the major 58 kDa 'intermediate' form seen here which is enzymatically active (Supplementary Fig. 1b). At the C-terminus, an 18-residue loop terminates in a short  $\alpha$ -helix, which packs against the hydrophobic region of the SD2  $\beta$ -sheet.

Out of the six cysteines in the mature IDS polypeptide, four are observed to form disulfide bonds. The C171–C184 bridge links short  $\beta$ -strands near the putative substrate recognition site, while the C422–C432 linkage participates in stabilization of an extended ( $> 30$  residue) solvent-exposed loop region. Both disulfide bonds are functionally important, as mutation of any of these cysteines causes Hunter syndrome (Supplementary Table 1). One free cysteine is found at the tip of an 18-residue loop connecting strand  $\beta 1$  of the major  $\beta$ -sheet and N-terminal  $\alpha$ -helix 1. The other, C84, is post-translationally modified as a key catalytic residue in the active site.

**Table 1 | Data collection and refinement statistics.**

|   | IDS-native (5FQL)           | IDS-GdCl <sub>3</sub>       | IDS-NaI                     |
|---|-----------------------------|-----------------------------|-----------------------------|
| <i>Data collection</i>                              |                             |                             |                             |
| Wavelength (Å)                                      | 0.9763                      | 1.475                       | 0.9763                      |
| Space group   | <i>P</i> 3 <sub>1</sub> 1 2 | <i>P</i> 3 <sub>1</sub> 1 2 | <i>P</i> 3 <sub>1</sub> 1 2 |
| Cell dimensions                                     |                             |                             |                             |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                  | 71.1, 71.1, 285.9           | 68.8, 68.8, 289.0           | 72.6, 72.6, 286.4           |
| $\alpha$ , $\beta$ , $\gamma$ (°)                   | 90, 90, 120                 | 90, 90, 120                 | 90, 90, 120                 |
| Resolution (Å)                                      | 35.52–2.30 (2.39–2.30)*     | 80.0–3.28 (3.36–3.28)       | 62.90–3.50 (3.70–3.50)      |
| <i>R</i> <sub>merge</sub>                           | 0.080 (0.944)               | 0.166 (2.67)                | 0.075 (1.17)                |
| <1/ $\sigma$ >                                      | 12.4 (1.9)                  | 17.4 (1.3)                  | 18.2 (2.6)                  |
| CC(1/2)   | 0.997 (0.781)               | –                           | –                           |
| Completeness (%)                                    | 96.2 (97.7)                 | 99.0 (92.1)                 | 99.2 (99.8)                 |
| Redundancy  | 5.3 (5.2)                   | 15.8 (10.0)                 | 11.5 (11.7)                 |
| Anisotropic $\Delta B$ (Å <sup>2</sup> )            | 38                          | 117                         | 134                         |
| <i>Refinement</i>                                   |                             |                             |                             |
| Resolution (Å)                                      | 35.53–2.30 (2.38–2.30)      | –                           | –                           |
| No. reflections                                     | 35,738 (3,605)              | –                           | –                           |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> | 0.181/0.210                 | –                           | –                           |
| No. atoms (excl. H)                                 |                             |                             |                             |
| Protein   | 4,269                       | –                           | –                           |
| Water   | 213                         | –                           | –                           |
| <i>B</i> -factors                                   |                             |                             |                             |
| Protein   | 67.7                        | –                           | –                           |
| Water   | 57.3                        | –                           | –                           |
| <i>R.m.s.d.</i>                                     |                             |                             |                             |
| Bond lengths (Å)                                    | 0.004                       | –                           | –                           |
| Bond angles (°)                                     | 0.655                       | –                           | –                           |

r.m.s.d., root mean squared deviation.

\*Values in parentheses are for highest-resolution shell. One crystal was used per structure.

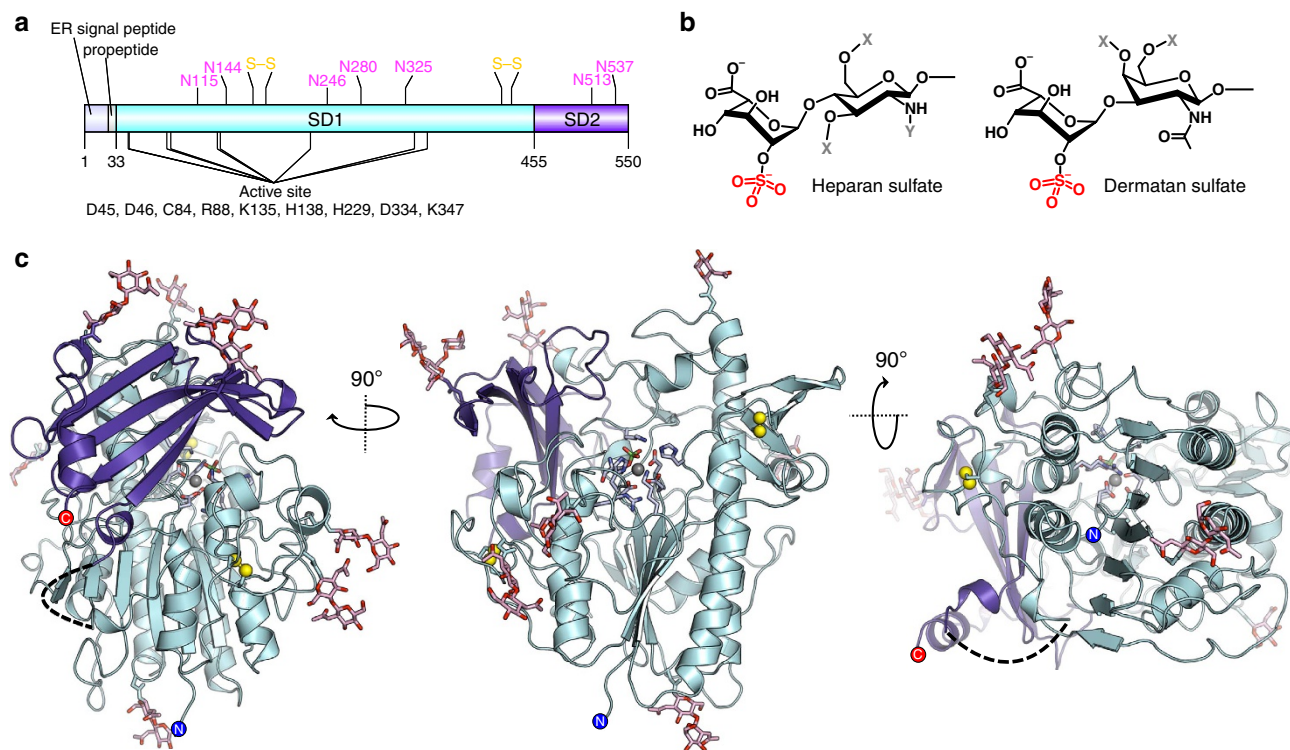
The IDS sequence contains eight putative N-linked glycosylation sites (Fig. 1a). Continuous electron densities were observed adjacent to all asparagine residues involved in N-linked glycosylation, and at least one N-acetylglucosamine moiety linked to the nitrogen of the asparagine side chain was built at each predicted site.

**Comparison to other sulfatase enzymes.** To date, 17 genes encoding human sulfatases have been identified. Amino acid sequences vary significantly amongst members of the sulfatase family, and pairwise alignment of IDS with other human sulfatases reveals only ~20% sequence identity, with greater conservation in N-terminal regions (Supplementary Fig. 3). X-ray crystal structures have been determined of human sulfatases arylsulfatase A (ARSA)<sup>17</sup>, arylsulfatase B (ARSB)<sup>18</sup> the membrane-associated estrone or steroid sulfatase (STS, also known as arylsulfatase C)<sup>19</sup>, galactosamine-6 sulfatase (GALNS)<sup>20</sup> and N-sulfoglucosamine sulfohydrolase (SGSH)<sup>21</sup>. Deficiencies of these enzymes also lead to rare inherited disorders: metachromatic leukodystrophy, Maroteaux–Lamy syndrome (MPS VI), X-linked ichthyosis, Morquio A syndrome (MPS IV A) and Sanfilippo A syndrome (MPS III A), respectively.

Structural alignments reveal significant topological similarity between N-terminal regions of IDS and other sulfatases, concentrated around the central  $\beta$ -sheet in SD1 (Fig. 2a, 174 residues with backbone C $\alpha$  RMSD < 2.0). Beyond this, poor conservation of surface loops and divergent C-terminal sequences prevent main-chain superposition, but the 4-stranded  $\beta$ -sheet in SD2 is a feature shared by all human sulfatases. Nevertheless, several large differences are observed. The central  $\beta$ -sheet at the core of IDS is formed by seven strands in contrast to ten for O-sulfatases ARSA, ARSB, STS and GALNS, more closely resembling the 8-stranded core of SGSH, an N-sulfatase. This

has consequences for the shape of the putative substrate-binding region. In IDS, the cleft is relatively wide and shallow, while in other sulfatases it forms a narrower cavity, deepening towards the catalytic site (Fig. 2b–e). In arylsulfatases A and B this cavity is extended by three additional strands of the central  $\beta$ -sheet; in STS, the two transmembrane helices emerge from the globular domain directly adjacent to the substrate-binding cleft, forming a deep-binding pocket (Fig. 2e). Differences are also observed in the C-terminal regions; most notably the very extensive C-terminal loops or  $\beta$ -meander motifs present in other sulfatases are absent from IDS. In ARSA, ARSB and GALNS these C-terminal extensions help to define the substrate-binding pockets. In IDS, the four antiparallel strands comprising the SD2  $\beta$ -sheet are considerably longer than those of other sulfatases, and hence a greater contribution to the shape of the substrate-binding cleft comes directly from SD2.

**Substrate-binding cleft and active site.** Despite highly divergent sequences and the diversity of substrates processed by these enzymes, the active-site residues (red stars in Supplementary Fig. 3) are highly conserved and are almost completely superimposable in all six sulfatase structures (Fig. 2f), providing strong support for a common catalytic mechanism. The active site of IDS is located in a cleft comprising many basic residues, conferring a positive electrostatic surface potential consistent with binding negatively charged polymeric substrates (Figs 3a and 1b). The catalytic residue C84 is situated at the N-terminus of helix 2, opposite the central  $\beta$ -sheet. In all eukaryotic sulfatases, the thiol group of this residue is oxidized to C $\alpha$ -formylglycine (FGly, 2-amino-3-oxopropionic acid), an essential modification required for catalytic activity<sup>22</sup>. This modification is performed by FGly-generating enzyme (FGE), which recognizes the highly conserved sequence motif CxPSR during translocation of the



**Figure 1 | X-ray crystal structure of IDS.** (a) Annotated diagram of the IDS sequence, showing the ER signal peptide and propeptide (grey), subdomains (SD) 1 and 2 (light blue and purple, respectively), active-site residues (black), disulfide bridges (yellow) and N-linked glycosylation sites (pink). (b) Natural IDS substrates heparan sulfate (with  $\alpha$ 1-4 linkage) and dermatan sulfate ( $\alpha$ 1-3 linkage). The sulfate group removed by IDS is indicated (red). The residue at the +1 position can be either D-glucosamine or D-galactosamine, with variable substituents at C2<sub>N</sub>, C3, C4 and C6 positions (X = H/SO<sub>3</sub><sup>-</sup>; Y = H/SO<sub>3</sub><sup>-</sup> / COCH<sub>3</sub>) (c) Ribbon diagram of IDS coloured by subdomain, shown in three orthogonal views. Disulfide bonds (yellow spheres), calcium ion (grey sphere) and glycans (pink sticks) are shown. The active site is indicated by grey sticks.

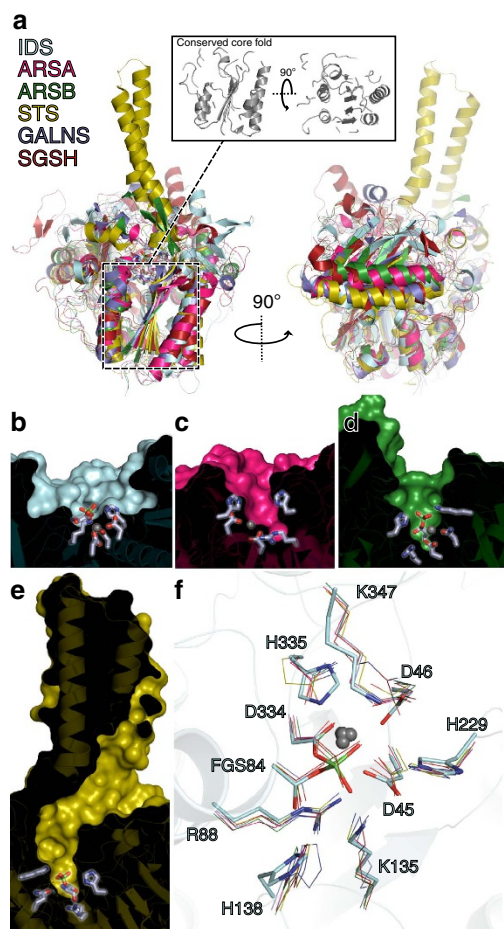
nascent (and unfolded) polypeptide chain into the endoplasmic reticulum (ER) lumen<sup>23,24</sup>. FGly functions catalytically as an aldehyde hydrate (FGly aldehyde-hydrate, FGH) with two geminal hydroxyl groups; here we observe what we have interpreted as a covalent sulfate-ester intermediate (FGly sulfate, FGS) in the IDS active site (Fig. 3b), which could originate from natural substrate encountered in cells during expression. An alternative interpretation that this is a phosphate picked up from the storage buffer, as proposed for SGSH<sup>21</sup>, cannot be excluded on the crystallographic evidence. However, the higher stability of sulfate esters makes this the more likely interpretation.

Extra electron density for a metal ion was observed next to the modified FGS residue. This was modelled as Ca<sup>2+</sup>, coordinated by one nitrogen and five oxygen atoms with approximate octahedral geometry: one oxygen each from the side chains of D45, D46 and D334; two sulfate oxygens from FGS84 and one nitrogen from the side chain of H335. (Fig. 3c). The metal–oxygen bond distances range from 2.3 Å to 2.5 Å, corresponding to the expected values for calcium<sup>25</sup>. The second carboxylate oxygen of D334 forms an additional longer interaction of 3.3 Å. Successive temperature factor refinements for bivalent cations Ca<sup>2+</sup>, Mn<sup>2+</sup> and Mg<sup>2+</sup> resulted in values of 55.2 Å<sup>2</sup>, 65.4 Å<sup>2</sup> and 39.7 Å<sup>2</sup>, respectively; the B-factor for a Ca<sup>2+</sup> ion correlates best with the average B-factor (52 Å<sup>2</sup>) of the coordinating side chain atoms<sup>25</sup>. Calcium was also identified in the metal-binding pocket of human ARSB<sup>18</sup>, STS<sup>19</sup>, GALNS<sup>20</sup>, SGSH<sup>21</sup> and *Pseudomonas aeruginosa* arylsulfatase (PAS)<sup>26</sup> whilst in human ARSA, both magnesium and calcium have been described<sup>17,27</sup>. With its lower Lewis acidity, Ca<sup>2+</sup> is more consistent than Mg<sup>2+</sup> with the stability of the observed

sulfate-ester modification of formylglycine: in related structures, the identification of either a sulfate ester (ARSB<sup>18</sup>, STS<sup>19</sup>) or phosphate ester (ARSA<sup>17,27</sup>, SGSH<sup>21</sup>) modification correlates with the identification of the metal ion as Ca<sup>2+</sup>. The presence of divalent metal ions in sulfatases is thought to be essential for active-site stabilization and sulfate-ester formation<sup>28</sup>.

Nine out of ten IDS active-site residues are highly conserved among sulfatases: D45, D46, C84, R88, K135, H138, H229, D334 and K347. The histidine residue at position 335 is replaced in most sulfatases by asparagine, and by glutamine in STS (Supplementary Fig. 3). Side chains of conserved H138 and H229 likely play important roles in catalysis, making direct contacts with the free geminal hydroxyl of FGS and a sulfate oxygen (OS3) of FGS, respectively. The positively charged groups of R88, K135 and K347 make several structural interactions within the catalytic core. R88 stabilizes the orientation of metal-binding residues D45 and D334 and interacts with the free geminal hydroxyl of FGS; K135 also interacts with D45 to correctly orient the side chain for metal-binding. D46 is positioned for metal-binding via interactions with K347 and W337. The sulfate moiety of FGS is stabilized by hydrogen bonds with the side chains of D46, K135, D334 and K347 and, via a water molecule, to the side chains of Y348 and K479 (Fig. 3d).

Attempts to obtain an enzyme–substrate complex structure by co-crystallization or soaking with a variety of ligands were not successful. Nevertheless, to gain insight into possible modes of substrate binding, we performed docking experiments with 2-sulfo-iduronic acid, heparan sulfate disaccharides and a dermatan sulfate disaccharide, restraining the position of the scissile sulfate group adjacent to the catalytic FGH residue with appropriate geometry for attack (Fig. 3e). This demonstrates



**Figure 2 | Structural alignment of IDS with other human sulfatase enzymes.** (a) Secondary structure-based superposition of IDS (blue), ARSA (pink), ARSB (green), STS (yellow), GALNS (slate) and SGSH (dark red) in two orthogonal views. Active-site residues are shown as grey sticks. Inset: core structural elements shared by all human sulfatases, showing only conserved regions with backbone  $C\alpha$  RMSD  $< 2.0 \text{ \AA}$ . (b–e) Cutaway diagrams of the substrate-binding cleft of IDS, ARSA, ARSB and STS, coloured as above. Active-site residues are shown as grey sticks. (f) Least squares superposition of conserved active-site residues, coloured as above. The metal ions are shown as grey spheres.

that it is possible to place the substrate in a favourable orientation consistent with our mechanistic understanding. The 2S-IdoA monosaccharide binds in a low energy  ${}^4C_1$  conformation, making additional hydrogen bonds to R273, N167 and Y348 from SD1, and K479 from SD2. The substrate-binding cleft is large enough to accommodate both the heparan sulfate ( $\alpha 1-4$  linkage) and dermatan sulfate ( $\alpha 1-3$  linkage) disaccharides with different substituents at the +1 position: H226, K479 and K135 can make polar contacts with additional OH,  $\text{SO}_3^-$  or  $\text{COCH}_3$  groups. Different ring puckers of 2S-IdoA (ref. 29) may also be tolerated by the binding site. An analysis of mean B-factors per residue reveals that the residues proximal to the scissile sulfate are well ordered, indicating that the active site is likely preorganised. However, K479 and D269 are likely flexible and may change conformation to accommodate the diverse and heterogeneous substrates processed by this enzyme.

## Discussion

Two mechanisms for sulfate-ester hydrolysis have been proposed based on crystal structures of human ARSA, human ARSB and

*Pseudomonas* arylsulfatase (PAS), and studies on the enzyme kinetics of active-site mutants<sup>30</sup>. Most recent experimental evidence supports the transesterification-elimination (TE) mechanism, in which the catalytically relevant species is an FGly aldehyde-hydrate (FGH) with a second geminal hydroxyl. The arrangement of the IDS active site is compatible with the consensus TE mechanism, hence we propose a similar mechanism for IDS in which FGH84 performs an  $\text{S}_{\text{N}}2$  nucleophilic attack on the sulfur atom to generate the covalent FGS adduct, followed by elimination of bound sulfate and rehydration of the resultant FGly aldehyde. In particular, H229, H138, D334 and K347 are likely to play active roles in proton transfer or activation of water during the reaction cycle, as proposed for other sulfatases.

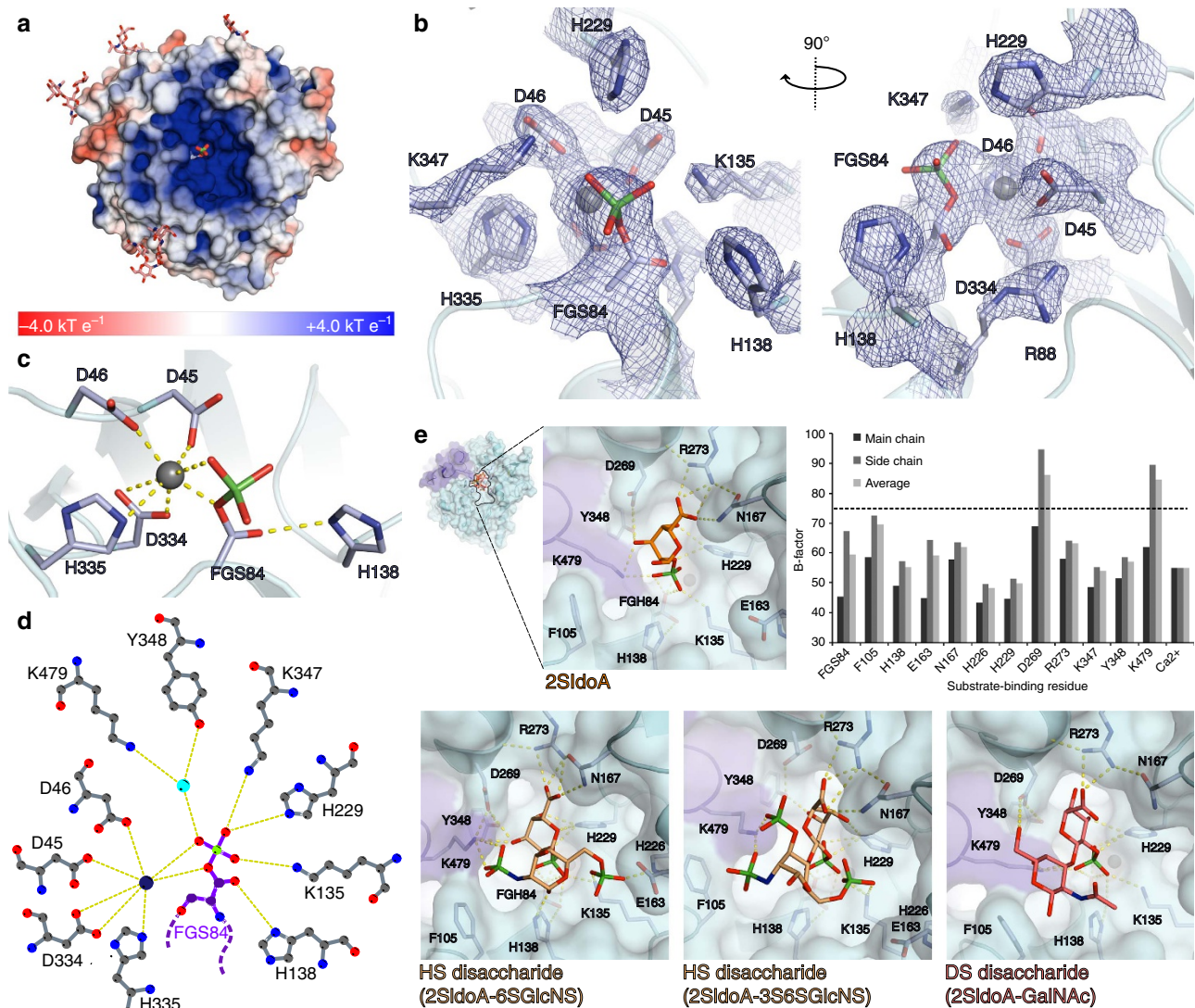
Interestingly, we observe a sulfate ion bound in the active site, similar to the observed ‘resting state’ of ARSB and STS. The covalent linkage is unambiguous, with a distance of  $1.5 \text{ \AA}$  between the sulfur atom and the FGS ester oxygen. Attempts to refine a non-covalently bound sulfate ion adjacent to an FGH residue, as seen in the high-resolution PAS structure, did not agree with the electron density present in the IDS structure (Supplementary Fig. 4). Due to the dependency on general acid/base catalysis, it is likely that sulfate elimination will only occur efficiently at the optimum lysosomal pH of  $\sim 4.8$ , and not in the storage buffer (pH 7.5) or crystallization buffer (pH 6.1). Indeed, despite this observed sulfate ion, IDS samples were enzymatically active at pH 5.0 (Supplementary Fig. 1b).

To date, 542 pathogenic mutations in the *IDS* gene have been identified in patients with Hunter syndrome (Supplementary Table 1). The mutation spectrum is diverse and ranges from small point mutations, deletions and insertions to truncations, large *IDS* gene deletions and complex rearrangements. A large proportion of mutations associated with Hunter syndrome are missense mutations randomly distributed throughout the protein sequence, with 217 amino acid substitutions described at 128 positions (Fig. 4). These are likely to result in protein misfolding, catalytic inactivation, premature degradation or failure of lysosomal targeting. However, at least five of these are likely to be false positives, as they are found in the Exome Aggregation Consortium database<sup>31</sup> at a frequency too high to be consistent with a role in this rare disease (Supplementary Table 1).

The structure of IDS provides detailed insight into the consequences of many disease-associated mutations. Several mutations directly affect the catalytic core of the enzyme, either by direct substitution of key active-site residues, interference with metal ion coordination, indirect active-site destabilization/distortion or corruption of the CxPSR motif necessary for the essential FGly modification (Fig. 4). Such changes are likely to eliminate or severely reduce catalytic activity with grave phenotypic consequences; for example R88H/C/P/L/G/S mutations and K347I/Q/T/E mutations often lead to severe, early-onset disease.

Outside the active site, a small number of mutations are located on the solvent-accessible surface of the enzyme. These include substitutions that alter local surface charge (E125V) and others that interfere with N-linked glycosylation by either removal of existing sites (N115I/Y, S117Y), or aberrant introduction of new glycans (T130N) (Fig. 4). As well as causing misfolding, incorrect glycosylation may interfere with lysosomal targeting by impaired binding to the mannose-6-phosphate receptor.

In contrast to those affecting the enzyme surface, nearly 90% of pathogenic missense mutations involve amino acids that are buried within the fold of IDS (Supplementary Table 1). The expected severity of these mutations is variable and consequences range from local destabilization and misfolding to global unfolding, leading to premature degradation. In the tightly

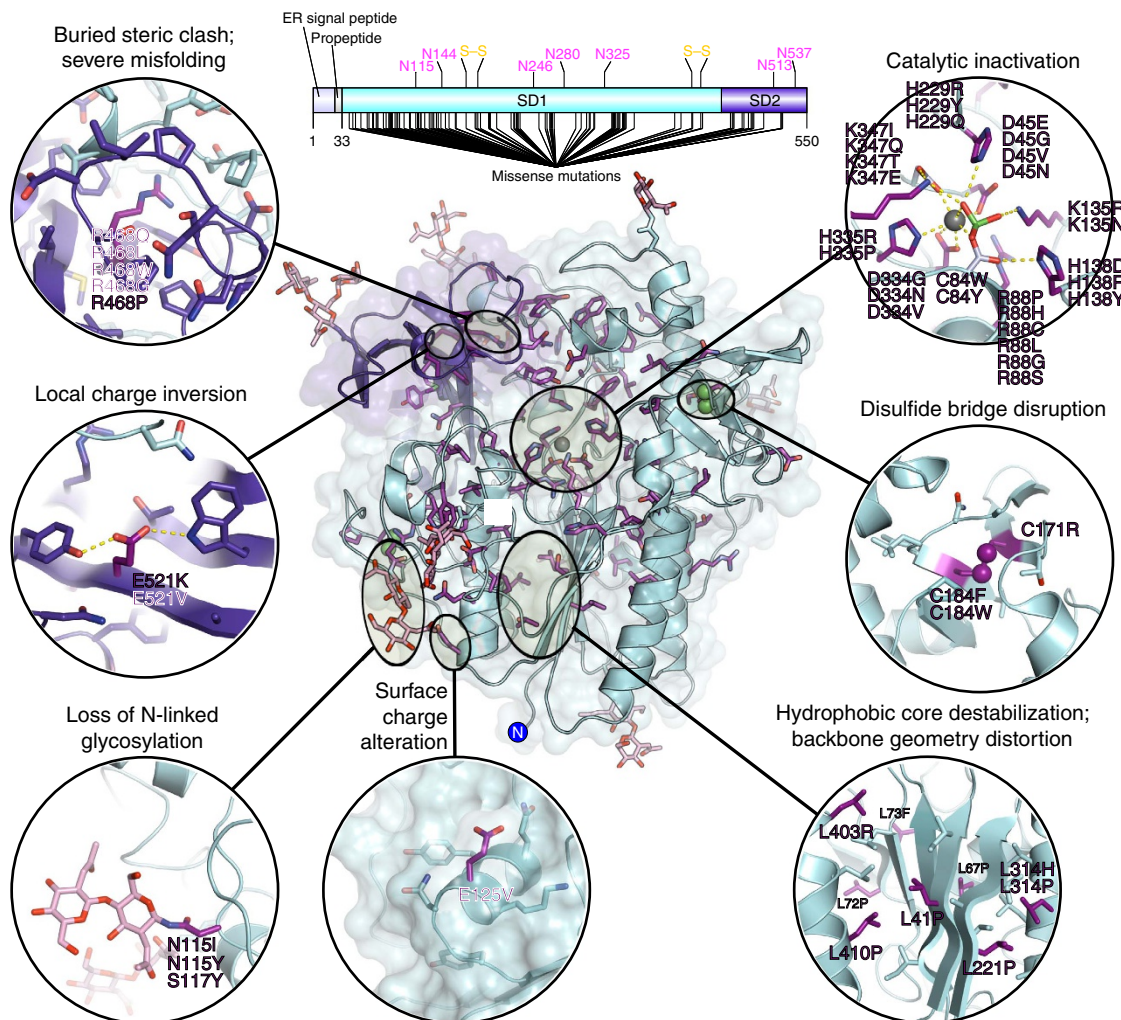


**Figure 3 | Details of IDS active site and substrate-binding cleft.** (a) The substrate-binding cleft is coloured by electrostatic potential at the solvent-accessible surface from red (negative,  $-4.0 \text{ kT e}^{-1}$ ) to blue (positive,  $+4.0 \text{ kT e}^{-1}$ ). Electrostatic potential was calculated using a pH value of 4.8 when assigning side chain protonation. The catalytic, modified FGS84 residue is shown as sticks. (b) Two orthogonal views of electron density (blue mesh,  $2mF_o - DF_c$ , contoured at  $0.35 \text{ e}^{-} \text{ \AA}^{-3}$ ) for IDS active-site residues (grey sticks) and the  $\text{Ca}^{2+}$  ion (grey sphere). (c) Details of the  $\text{Ca}^{2+}$  ion coordination (yellow dashed lines) in the IDS active site. (d) Schematic representation of amino acid side chain interactions with the covalently-modified FGS84 residue. Hydrogen bonds (yellow dashed lines) are shown. (e) Docked structures of 2-sulfo-iduronic acid, two variably sulfated heparan sulfate disaccharides and a dermatan sulfate disaccharide, showing details of the putative substrate-binding region. Hydrogen bonds (dashed yellow lines) are indicated. For solvent-accessible residues situated in the substrate-binding cleft, a graph of average B-factor per residue is shown. The mean B-factor for all IDS non-hydrogen atoms is indicated with a dotted line.

packed hydrophobic core surrounding the central  $\beta$ -sheet, several such mutations (L41P, L67P, L72P, L73F, L221P, L314P/H, L403R and L410P) would introduce main-chain geometry distortion or buried steric clashes, disrupting the local secondary structure in the core of SD1 (Fig. 4). In SD2, the guanidinium group of R468 participates in a buried hydrogen bonding network and hydrophobic stacking interactions stabilizing several nearby loops, helical turns and  $\beta$ -strands (Fig. 4, Supplementary Fig. 5). Mutation of this residue (R468Q/L/W/G/P) is associated with a range of Hunter disease phenotypes from mild to severe.

In general, proteins with folding defects will be retained by the ER quality control machinery, preventing trafficking through the Golgi apparatus and final delivery to lysosomes<sup>32</sup>. In experiments with COS cells, transiently expressed P86L/R and P480L/Q

mutants did not efficiently traffic to lysosomes<sup>33</sup>. However, misfolding does not always lead to catalytic inactivation, and several IDS missense mutants with residual enzyme activity have been identified<sup>33–37</sup>. Such mutants could be promising candidates for treatment by pharmacological chaperone therapy (PCT). PCT relies on use of small molecules which selectively bind and stabilize target proteins, shifting the folding equilibrium towards functional material, thus alleviating ER retention and increasing lysosomal delivery. This therapeutic approach is under active development for several other lysosomal storage disorders including Pompe<sup>38</sup>, Fabry<sup>39</sup>, Gaucher<sup>40,41</sup> and Krabbe disease<sup>42,43</sup>. As well as the IDS active site, several other surface pockets with different electrostatic potentials are present which may be exploited for future chaperone development (Supplementary Fig. 6).



**Figure 4 | Disease-associated mutations of IDS.** (Top) Annotated diagram of the IDS protein, indicating positions of all known Hunter syndrome missense mutations. All residues that are mutated are shown as sticks (purple) on the structure of IDS, coloured as above. Different mutations have different consequences; for a selection of these a more detailed view is provided (Insets) showing the surrounding region of the structure that is perturbed.

IDS variants with disruptive mutations such as complete deletions or large rearrangements will not be amenable to PCT, and patients carrying such mutations are also more susceptible to the development of neutralizing antibodies that reduce the efficacy of ERT<sup>44</sup>. Knowledge of the structure of IDS could also facilitate structure-based de-immunization<sup>45,46</sup> of the protein used in ERT, to remove T-cell epitopes while preserving a stable structure.

The structure of IDS described here provides a much-needed snapshot of the enzyme defective in Hunter syndrome, delivering a detailed atomic framework for the rationalization and further understanding of disease-causing mutations, and supplying ‘blueprints’ for the future development of new PCT agents or less immunogenic protein for ERT. In addition, details of the active site support a likely catalytic mechanism shared by other sulfatases, adding to our understanding of this family of enzymes.

## Methods

**Deglycosylation experiments.** IDS protein in 100 mM NaCl, 10 mM Tris · HCl pH 7.5 was denatured by the addition of urea up to 2.5 M final concentration. Samples were treated with PNGase F (Sigma) as directed by the manufacturer (5 h, 310 K) before analysis by reducing SDS-PAGE.

**Enzyme activity assay.** Fluorogenic assays were performed based on the methods of Voznyi *et al.*<sup>47</sup> and Tolun *et al.*<sup>48</sup>. In brief, the sulfate group on substrate

4-methylumbelliferyl- $\alpha$ -L-iduronide-2-sulfate (MU- $\alpha$ IdoA-2S) is hydrolysed by IDS, upon which the desulfated product 4-methylumbelliferyl- $\alpha$ -L-iduronide (MU- $\alpha$ IdoA) becomes a substrate for reporter enzyme  $\alpha$ -iduronidase, which cleaves MU- $\alpha$ IdoA to generate fluorescent product 4-methylumbelliferone.

Substrates and reporter enzymes were obtained from Moscerdam Substrates. Reactions of 50  $\mu$ l comprised 0.3 mM MU- $\alpha$ IdoA-2S, 1.3 nM IDS (100 ng ml<sup>-1</sup>) and excess  $\alpha$ -iduronidase reporter enzyme in 0.1 M sodium acetate pH 5.0, 10 mM lead acetate, 0.01% Tween-20. Reactions were incubated at room temperature (293 K) for 60 min before termination with 200  $\mu$ l stopping buffer (0.5 M Na<sub>2</sub>CO<sub>3</sub> pH 10.7, 0.025% Triton X-100). Liberated 4-methylumbelliferone was quantified by measuring fluorescence ( $\lambda_{\text{ex}} = 365$  nm;  $\lambda_{\text{em}} = 445$  nm) with a SpectraMax M2 plate reader (Molecular Devices).

**Protein preparation and crystallization.** Purified recombinant human IDS (Elapraxe) was kindly provided by Shire Pharmaceuticals. Crystallization trials were carried out with both glycosylated and deglycosylated protein, but considering that no improvement in diffraction quality was observed after deglycosylation and that the two crystal forms were isomorphous, the glycosylated material was used for structure determination. IDS was buffer-exchanged by dialysis (10-kDa molecular mass cut-off; Thermo Scientific) into 100 mM NaCl, 10 mM Tris HCl pH 7.5 and concentrated to 10.0 mg ml<sup>-1</sup> using Amicon Ultra 10-kDa molecular mass cut-off centrifugal concentrators (Millipore). Crystals were grown at 20 °C by hanging drop vapour diffusion. In total, 1.0  $\mu$ l of protein was mixed with 1.0  $\mu$ l reservoir solution (0.1 M MES-NaOH pH 6.1, 1.0 M LiCl and 20% wt/vol PEG 6 K) and equilibrated against 1.0 ml reservoir solution. Crystals were cryoprotected by transferring to a drop of cryo solution (equilibrated reservoir solution increased to 30% wt/vol PEG 6 K and supplemented with 20% wt/vol ethylene glycol) for 30 s before flash-cooling in liquid nitrogen.



**Crystal soaks for experimental phasing.** For the NaI-derivative data set, crystals were soaked sequentially in cryo solution (above) supplemented with 0.4 M NaI (1 min) and 1.0 M NaI (5 min). For the GdCl<sub>3</sub>-derivative data set, crystals were prepared by soaking in cryo solution supplemented with 50 mM GdCl<sub>3</sub> (30 min). Crystals were then flash-cooled as above.

**Crystallographic structure determination and analysis.** Data collection and processing statistics are detailed in Table 1. All data sets were collected at a temperature of 100 K. Diffraction data were recorded at Diamond Light Source beamline I03 on a Pilatus 6M detector (Dectris). The native data set was collected at  $\lambda = 0.9763 \text{ \AA}$  over a 360° oscillation range with  $t = 0.15 \text{ s}$  and  $\Delta\Phi = 0.2^\circ$  per image. Diffraction data were processed using the XIA2 pipeline<sup>49</sup> implementing MOSFLM<sup>50</sup> for indexing and integration, POINTLESS<sup>51</sup> for point group determination and AIMLESS<sup>52</sup> for scaling and merging. iMOSFLM<sup>50</sup> was also used for manual indexing and integration. Resolution cut-off was decided by  $CC_{1/2} > 0.5$  and  $\langle I/\sigma I \rangle \geq 1.8$  in the outer resolution shell<sup>53</sup>. The NaI-derivative data set was collected as above; an initial attempt to collect data at the iodide peak wavelength suffered from excessive radiation damage. The GdCl<sub>3</sub>-derivative data set was collected at  $\lambda = 1.475 \text{ \AA}$ .

SAD phases useful to about 3.8 Å resolution were determined for the GdCl<sub>3</sub>-derivative crystal with phenix.autosol<sup>54</sup>, which found 4 Gd sites. The resulting electron density was not interpretable but allowed docking of known sulfatase structures to aid further interpretation; the best agreement was obtained with PDB entry 3B5Q, a putative sulfatase from *Bacterioides thetaiotaomicron* VPI-5482. Electron density cut out from the SAD map, using the docked sulfatase structure as a mask, was used to solve the native and iodide-soak data by molecular replacement in Phaser<sup>55</sup>. A substructure of 11 iodide sites was then determined using SAD log-likelihood-gradient maps<sup>56</sup>, starting from the electron density model. At this point, initial phase information, a molecular mask, and operators relating the positions of IDS in the three crystals were available, allowing averaging to be carried out with phenix.multi\_crystal\_average<sup>57</sup>. An iterative process of cycles of automated model-building with phenix.autobuild<sup>58</sup>, followed by determination of updated substructures (ultimately 8 sites for the GdCl<sub>3</sub> derivative and 13 sites for the iodide soak) and multi-crystal averaging, eventually led to a model comprising 405 of 517 residues, with  $R_{\text{free}} = 0.42$ .

All further model building and refinement was performed iteratively using COOT<sup>59</sup> and phenix.refine<sup>60</sup> (Table 1). The Protein Data Bank file and CIF restraints for the non-standard FGS84 residue were generated using eLBOW<sup>61</sup>. Hydrogens were refined in the riding position. MolProbity<sup>62</sup> was consulted throughout the refinement process, at the end of which 95.4% of residues were in the favoured Ramachandran region and none were outliers. Estimation of buried surface area in the IDS structure was carried out using PISA<sup>63</sup>. For the electrostatic potential calculations, partial charges were assigned using the PDB2PQR server<sup>64</sup>, implementing PROPKA<sup>65</sup> to estimate side chain pKa values. Electrostatic surfaces were calculated using APBS<sup>66</sup> and structural figures were rendered using PyMOL (Schrodinger LLC). Two-dimensional hydrogen bonding diagrams were created using LigPlot+ (ref. 67).

**Docking.** To prepare the protein model for docking, the sulfated FGS84 residue was replaced with FGH (Protein Data Bank code DDZ: 3,3-dihydroxy L-alanine) and waters and glycans were removed. Ligand PDB files for 2-sulfo-iduronic acid and various heparan sulfate and dermatan sulfate disaccharides were generated with the GLYCAM-Web GAG builder (<http://glycam.org>) and minimized with PRODRG<sup>68</sup>. All docking was performed with HADDOCK 2.2 (ref. 69) using the WeNMR grid service<sup>70</sup>. Residues FGH84, Ca<sup>2+</sup>, K135, K347 and H229 were specified as active, with passive residues defined automatically within a 6.5 Å radius of active residues. Unambiguous distance restraints were enforced to maintain the correct orientation of the nucleophile (via interaction with H138) and preserve Ca<sup>2+</sup> coordination geometry during docking. For ligands with multiple sulfate groups, an additional restraint was provided to juxtapose the relevant iduronate-2-sulfate group close to the active site.

**Amino acid sequence alignment.** Primary sequences for known human sulfatases were retrieved from the UniProt database<sup>71</sup> and annotated ER signal peptides were removed. Sequences were aligned by using ClustalW<sup>72</sup>, and graphical representations of the alignment were prepared using ALINE<sup>73</sup>.

**Data availability.** The atomic coordinates and structure factors for IDS have been deposited in the wwPDB under accession code 5FQL. All other data are available from the corresponding author upon reasonable request.

## References

- Hunter, C. A rare disease in two brothers. *Proc. R. Soc. Med.* **10**, 104–116 (1917).
- Li, P., Bellows, A. B. & Thompson, J. N. Molecular basis of iduronate-2-sulfatase gene mutations in patients with mucopolysaccharidosis type II (Hunter syndrome). *J. Med. Genet.* **36**, 21–27 (1999).
- Lin, H. Y. *et al.* Incidence of the mucopolysaccharidoses in Taiwan, 1984–2004. *Am. J. Med. Genet. A* **149A**, 960–964 (2009).
- Baehner, F. *et al.* Cumulative incidence rates of the mucopolysaccharidoses in Germany. *J. Inher. Metab. Dis.* **28**, 1011–1017 (2005).
- Froissart, R. *et al.* Identification of iduronate sulfatase gene alterations in 70 unrelated Hunter patients. *Clin. Genet.* **53**, 362–368 (1998).
- Muenzer, J. *et al.* Multidisciplinary management of Hunter syndrome. *Pediatrics* **124**, e1228–e1239 (2009).
- Froissart, R., Da Silva, I. M. & Maire, I. Mucopolysaccharidosis type II: an update on mutation spectrum. *Acta Paediatr. Suppl.* **96**, 71–77 (2007).
- Lehner, B. Modelling genotype-phenotype relationships and human disease with genetic interaction networks. *J. Exp. Biol.* **210**, 1559–1566 (2007).
- Piotrowska, E. *et al.* Correlation between severity of mucopolysaccharidoses and combination of the residual enzyme activity and efficiency of glycosaminoglycan synthesis. *Acta Paediatr.* **98**, 743–749 (2009).
- Wraith, J. E. *et al.* Mucopolysaccharidosis type II (Hunter syndrome): a clinical review and recommendations for treatment in the era of enzyme replacement therapy. *Eur. J. Pediatr.* **167**, 267–277 (2008).
- Muenzer, J. *et al.* A phase II/III clinical study of enzyme replacement therapy with idursulfase in mucopolysaccharidosis II (Hunter syndrome). *Genet. Med.* **8**, 465–473 (2006).
- Alcalde-Martin, C. *et al.* First experience of enzyme replacement therapy with idursulfase in Spanish patients with Hunter syndrome under 5 years of age: case observations from the Hunter Outcome Survey (HOS). *Eur. J. Med. Genet.* **53**, 371–377 (2010).
- da Silva, E. M., Strufaldi, M. W., Andriolo, R. B. & Silva, L. A. Enzyme replacement therapy with idursulfase for mucopolysaccharidosis type II (Hunter syndrome). *Cochrane Database Syst. Rev.* **2**, CD008185 (2016).
- Parenti, G. Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics. *EMBO Mol. Med.* **1**, 268–279 (2009).
- Wilson, P. J. *et al.* Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA. *Proc. Natl Acad. Sci. USA* **87**, 8531–8535 (1990).
- Froissart, R., Millat, G., Mathieu, M., Bozon, D. & Maire, I. Processing of iduronate 2-sulphatase in human fibroblasts. *Biochem. J.* **309**, 425–430 (1995).
- Lukatela, G. *et al.* Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **37**, 3654–3664 (1998).
- Bond, C. S. *et al.* Structure of a human lysosomal sulfatase. *Structure* **5**, 277–289 (1997).
- Hernandez-Guzman, F. G., Higashiyama, T., Pangborn, W., Osawa, Y. & Ghosh, D. Structure of human estrone sulfatase suggests functional roles of membrane association. *J. Biol. Chem.* **278**, 22989–22997 (2003).
- Rivera-Colon, Y., Schutsky, E. K., Kita, A. Z. & Garman, S. C. The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. *J. Mol. Biol.* **423**, 736–751 (2012).
- Sidhu, N. S. *et al.* Structure of sulfamidase provides insight into the molecular pathology of mucopolysaccharidosis IIIA. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1321–1335 (2014).
- Schmidt, B., Selmer, T., Ingendoh, A. & von Figura, K. A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271–278 (1995).
- Preusser-Kunze, A. *et al.* Molecular characterization of the human Alpha-formylglycine-generating enzyme. *J. Biol. Chem.* **280**, 14900–14910 (2005).
- Dierks, T., Lecca, M. R., Schlotterhose, P., Schmidt, B. & von Figura, K. Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J.* **18**, 2084–2091 (1999).
- Zheng, H. *et al.* Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat. Protoc.* **9**, 156–170 (2014).
- Boltes, I. *et al.* 1.3A structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure* **9**, 483–491 (2001).
- Chruszcz, M. *et al.* Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J. Inorg. Biochem.* **96**, 386–392 (2003).
- Ghosh, D. Three-dimensional structures of sulfatases. *Methods Enzymol.* **400**, 273–293 (2005).
- Ferro, D. R. *et al.* Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **195**, 157–167 (1990).
- Hanson, S. R., Best, M. D. & Wong, C. H. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed. Engl.* **43**, 5736–5763 (2004).
- Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Hampton, R. Y. ER-associated degradation in protein quality control and cellular regulation. *Curr. Opin. Cell Biol.* **14**, 476–482 (2002).
- Millat, G. *et al.* COS cell expression studies of P86L, P86R, P480L and P480Q Hunter's disease-causing mutations. *Biochim. Biophys. Acta* **1406**, 214–218 (1998).

34. Villani, G. R., Daniele, A., Balzano, N. & Di Natale, P. Expression of five iduronate-2-sulfatase site-directed mutations. *Biochim. Biophys. Acta* **1501**, 71–80 (2000).
35. Bonuccelli, G. *et al.* The effect of four mutations on the expression of iduronate-2-sulfatase in mucopolysaccharidosis type II. *Biochim. Biophys. Acta* **1537**, 233–238 (2001).
36. Sukegawa-Hayasaka, K. *et al.* Effect of Hunter disease (mucopolysaccharidosis type II) mutations on molecular phenotypes of iduronate-2-sulfatase: enzymatic activity, protein processing and structural analysis. *J. Inherit. Metab. Dis.* **29**, 755–761 (2006).
37. Lau, K. C. & Lam, C. W. Molecular investigations of a novel iduronate-2-sulfatase mutant in a Chinese patient. *Clin. Chim. Acta* **392**, 8–10 (2008).
38. Parent, G. *et al.* A chaperone enhances blood alpha-glucosidase activity in pompe disease patients treated with enzyme replacement therapy. *Mol. Ther.* **22**, 2004–2012 (2014).
39. Benjamin, E. R. *et al.* The pharmacological chaperone 1-deoxygalactonojirimycin increases alpha-galactosidase A levels in Fabry patient cell lines. *J. Inherit. Metab. Dis.* **32**, 424–440 (2009).
40. Sanchez-Olle, G. *et al.* Promising results of the chaperone effect caused by imino sugars and aminocyclitol derivatives on mutant glucocerebrosidases causing Gaucher disease. *Blood Cells Mol. Dis.* **42**, 159–166 (2009).
41. Lieberman, R. L. *et al.* Structure of acid beta-glucosidase with pharmacological chaperone provides insight into Gaucher disease. *Nat. Chem. Biol.* **3**, 101–107 (2007).
42. Hill, C. H. *et al.* Azasugar inhibitors as pharmacological chaperones for Krabbe disease. *Chem. Sci.* **6**, 3075–3086 (2015).
43. Lee, W. C. *et al.* Molecular characterization of mutations that cause globoid cell leukodystrophy and pharmacological rescue using small molecule chemical chaperones. *J. Neurosci.* **30**, 5489–5497 (2010).
44. Pano, A., Barbier, A. J., Bielefeld, B., Whiteman, D. A. & Amato, D. A. Immunogenicity of idursulfase and clinical outcomes in very young patients (16 months to 7.5 years) with mucopolysaccharidosis II (Hunter syndrome). *Orphanet. J. Rare Dis.* **10**, 50 (2015).
45. Parker, A. S., Choi, Y., Griswold, K. E. & Bailey-Kellogg, C. Structure-guided deimmunization of therapeutic proteins. *J. Comput. Biol.* **20**, 152–165 (2013).
46. King, C. *et al.* Removing T-cell epitopes with computational protein design. *Proc. Natl Acad. Sci. USA* **111**, 8577–8582 (2014).
47. Voznyi, Y. V., Keulemans, J. L. & van Diggelen, O. P. A fluorimetric enzyme assay for the diagnosis of MPS II (Hunter disease). *J. Inherit. Metab. Dis.* **24**, 675–680 (2001).
48. Tolun, A. A. *et al.* A novel fluorometric enzyme analysis method for Hunter syndrome using dried blood spots. *Mol. Genet. Metab.* **105**, 519–521 (2012).
49. Winter, G. *xia2*: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2009).
50. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 48–57 (2006).
51. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
52. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
53. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
54. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601 (2009).
55. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
56. Read, R. J. & McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 338–344 (2011).
57. Terwilliger, T. C. Rapid automatic NCS identification using heavy-atom substructures. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2213–2215 (2002).
58. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
60. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
61. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
62. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
63. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
64. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
65. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704–721 (2005).
66. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
67. Laskowski, R. A. & Swindells, M. B. LigPlot +: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).
68. Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
69. van Zundert, G. C. *et al.* The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
70. Wassenaar, T. A. *et al.* WeNMR: structural biology on the grid. *J. Grid Comput.* **10**, 743–767 (2012).
71. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
72. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
73. Bond, C. S. & Schuttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 510–512 (2009).

## Acknowledgements

We acknowledge Diamond Light Source for time on beamline I03 under proposal MX6641. We thank Shire Pharmaceuticals for providing Elaprase (idursulfase) and Tom Terwilliger for helpful advice on multi-crystal averaging. We also thank Alexandre Bonvin for modifying HADDOCK to support the non-standard amino acid FGH (PDB code DDZ). R.J.R. is supported by a Principal Research Fellowship funded by the Wellcome Trust (Grant 082961/Z/07/Z), which also supported C.H.H. and M.D. A.Z. was supported by a Senior Research Fellowship from the British Heart Foundation (PG/09/072/27945). J.E.D. is supported by a Royal Society University Research Fellowship (UF100371). Support received from the US National Institutes of Health (grant P01GM063210 R.J.R.) is gratefully acknowledged. The research was facilitated by a Wellcome Trust Strategic Award (100140) to the Cambridge Institute for Medical Research.

## Author contributions

M.D. prepared protein samples, performed biochemical experiments, collected crystallographic data, carried out the initial refinement and co-wrote the manuscript. A.Z., P.E.S. and D.M. assisted with protein chemistry and crystallization. R.J.R. solved the structure. G.B. assisted with refinement. C.H.H. refined and analysed crystallographic data, carried out docking calculations, prepared figures and co-wrote the manuscript. P.E.S., J.E.D. and R.J.R. co-wrote and edited the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Demydchuk, M. *et al.* Insights into Hunter syndrome from the structure of iduronate-2-sulfatase. *Nat. Commun.* **8**, 15786 doi: 10.1038/ncomms15786 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>