On the Mechanism of the Lysosomal Enzyme Iduronate-2-sulfatase. A Multiscale Approach

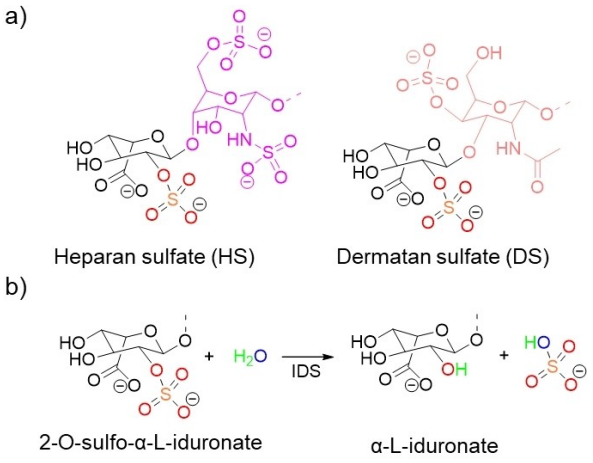
Mario Prejanò,*[a] Isabella Romeo,[a] Cristina Talerico,[a] Angela Parise,[a] Nino Russo,[a] and Tiziana Marino*[a]

Iduronate-2-sulfatase (IDS) is a Ca^{2+}-dependent enzyme belonging to the family of sulfatases that catalyzes the hydrolysis of sulphurylated glycosaminoglycans (GAGs), like dermatan and heparan sulphate. Its deficiency or modification leads to the accumulation of GAGs in the human body and to the occurrence of severe conditions, such as Hunter disease, or Mucopolysaccharidosis type II. Due to its involvement in this syndrome, it is of interest to understand the action mechanism of the enzyme to design new drugs for more efficient medical strategies. In the present work, we carried out a detailed multiscale modelling-based investigation, adopting molecular dynamics simulation (MDs) and QM/MM calculations to study the enzyme-dermatan sulfate interactions and to investigate the reaction mechanism of IDS. The analysis of molecular dynamics trajectories helped to shed light on the contribution of the individual residues of the active site in the recognition of dermatan sulfate. The role of FGly84 and His229, investigated in both neutral and protonated states in the case of the latter, is highlighted for the binding of the substrate. QM/MM calculations demonstrated that the reaction mechanism is a two-steps process occurring via a sulphurylation-desulphurylation mechanism where the FGly84 first attacks the sulphate group of the dermatan sulphate (DS), and later is desulphurylated.

Introduction

Mucopolysaccharidosis type II (MPS II, OMIM 309900) is known also as Hunter syndrome. Its name arises from Charles Hunter, the professor of medicine in Manitoba, Canada, who first described and diagnosed the disease in two brothers in 1917.[1] MPS II belongs to the category of X-linked recessive lysosomal storage diseases and is an extremely rare metabolic disease presenting more than 300 different mutations in the *IDS* gene. Currently, MPS II can be managed and treated with enzyme replacement therapy (ERT) on infusion of the recombinant enzyme, with limited benefits due to the inability to cross the blood-brain barrier (BBB) of the recombinant enzyme, reducing the effect of the therapy for the central nervous system.[2,3]

The modifications of the gene lead to a deficiency of the lysosomal enzyme iduronate-2-sulfatase (IDS, EC 3.1.6.13) with a wide range of symptoms that vary in severity among patients. The defect of IDS implicates the accumulation of two specific complex glycosaminoglycans (GAGs), dermatan sulphate (DS) and heparan sulphate (HS), which are covalently bound to the core proteins of proteoglycans present in the extracellular matrices and at cell surfaces (see Scheme 1a). HS and DS, at the non-reducing end, present 2-O-sulfo- α -L-iduronate residues that can be hydrolyzed at the C2-sulfate ester bond position by the IDS in healthy systems (see Scheme 1b), in the first step of the biodegradation pathway



Scheme 1. Heparan and dermatan sulfate (a) and mechanism of hydrolysis of the C2-bound ester bond of 2-O-sulfo-α-L-iduronate molecule catalyzed by IDS enzyme (b).

of these species.[2,4] In absence of IDS, HS and DS are accumulated and result in damage to the brain, respiratory tract, heart, liver, spleen, face, bones, and joints.[5]

The decreased production and catalytic activity, and/or protein misfolding represent the most reliable biomarkers for the onset of the disease with respect to the total lack of enzyme.[5–7] GAGs are implicated in many biological processes and conditions and their accumulation over time may affect multiple organs. For example, HS binds many proteins and is implicated in numerous biological processes and diseases including cancer[5] and viral infection.[8] The biological roles of DS, comprising GalNAcβ(1→3)IdoAα(1→4)-disaccharide units and that is the object of the current investigation, have been

less widely studied than HS. Endogenous DS plays crucial roles in the development of the central nervous system and has been further associated to cancer development also.[9–12]

In addition to the cases of DS and HS, sulfatases are hydrolases that cleave sulfate esters of other substrates, including sulfolipids and steroid sulfates also.[13,14] In the case of IDS, the X-ray crystallographic structure occurred in 2017 (PDB: 5FQL),[15] breaking new ground in the structural understanding of IDS enzyme and opening to new strategies for the treatment of Hunter disease.[15] The crystal structure, referred to human IDS, has sulfate anion covalently bound in the active site, and supplied useful insights for the catalytic mechanism, evidencing a striking similarity with other human sulfatases also.[15] By pairwise alignment of IDS with other human sulfatases, only ~20% sequence identity emerged with major conservation in N-terminal regions, concentrated in the central β-sheet of subdomain SD1 in particular. This subdomain is further characterized by a central 7-stranded β-sheet that contains the catalytic core. Furthermore, IDS main chain includes subdomain SD2, which consists of fragments due to lysosomal proteolytic processing events. Despite these divergent aspects, the active site residues are highly conserved among the human arylsulfatases A (ARSA), arylsulfatases B (ARSB), steroid sulfatases (STS), galactosamine-6 sulfatase (GALNS) and N-sulfoglucosamine sulfohydrolase (SGSH) supporting a possible common catalytic mechanism.[15]

The mature IDS polypeptide has two disulphide bonds (Cys171–Cys184 and Cys422–Cys432) that are functionally important since their mutation produces Hunter syndrome. Out of the two free cysteines, Cys84 is post-translationally modified as a key catalytic residue in the active site in all eukaryotic sulfatases; this thiol group is oxidized to Cα-formylglycine (FGly, 2-amino-3-oxopropionic acid) by FGly-generating enzyme (FGE).[15] This unique posttranslational modification, essential for catalytic activity, occurs within the endoplasmic reticulum before the sorting of sulfatase to different cellular compartments.[16] Next to the modified residue, an extra electron density for a metal ion modelled as Ca^{2+} was observed, coordinated by one nitrogen and five oxygen atoms. Calcium ion was also present in metal-binding pocket of other sulfatases including *Pseudomonas aeruginosa* arylsulfatase, (PAS) whose catalytic mechanism was reported in a previous work.[17] The role played by Ca^{2+} in the IDS is typical of metal ions present in other biological systems, and it is related to active-site stabilization and sulfate-ester formation.[18] The study of GAG-enzyme interactions is required both to increase the understanding of processes occurring during development and disease, and as a potential tool for carbohydrate-based drug discovery.

Furthermore, it is important to assess the influence on the mechanism of the numerous pathogenic mutations in the IDS gene identified in patients with Hunter syndrome by exploring the mechanism of the wild type enzyme, in order to better clarify the respective role of the catalytic pocket residues implicated in the reaction mechanism.[15,19] This understanding can be crucial also in the optimizations of synthetic GAGs production at larger scale.[20]

For all these reasons, a detailed multiscale approach-based investigation of IDS enzyme has been carried out. Molecular dynamics simulations (MDs) of the apo-form enzyme and of the complex with DS structure have been performed in order to investigate the conformational behavior, especially of the active site residues, and to obtain dynamics factors helpful for the next mechanistic study at QM/MM level. Such calculations, can offer new valuable mechanistic insights into the working mechanism of IDS, allowing an accurate description of all the involved steps to drive a rational strategy to fight the rare disease.

In addition, due to condition required by the IDS activity at pH 5.0 (sulfate eliminations occurs efficiently at optimum lysosomal pH ~4.8), histidine residues of the active site can exist in the protonated form. At this purpose we analyzed the substrate-binding modes and their contribution on the enzyme and to the rate determining step having His229 in protonated and in neutral form. MDs revealed important differences between the two forms of His229, evidencing as this residue can represent a determinant factor into molecular recognition with a closer focus at the binding step. This takes place in analogy to the active site of other sulfatases where one of the histidine residues of outer coordination shell of metal ion, in particular that in direct contact with sulphate oxygen, results to be protonated.[21,22]

It is worth to note that, at the best of our knowledge, the present work represents a further improvement of the atomistic scenario of IDS, since the current knowledge of Michaelis-Menten complex (ES) oft he enzyme is built on previous attempts based on docking calculations.[15] The information gained from the present study can be therefore considered as the "first real representative image" of the IDS. Due to the increasing efforts and interest in contrasting IDS-related diseases,[23–27] the adopted computational protocols, based on a well consolidated strategy previously applied in other investigations,[28–31] are indispensable to make predictions and provide insights that might not be available from experiments and that can crucially boost the design of new therapies.

Computational Details

MD simulations. The crystallographic structure selected to carry out the study (PDB code: 5FQL)[15] consists of 525 residues with the sulphated modified residue (ALS84, according to the PDB file nomenclature). In order to proceed with the investigation, ALS84 was replaced by the catalytically active formylglycine (FGly84, see Figure 1). The parameters for the non-standard FGly84 residue have been obtained after geometry optimization at B3LYP level of theory and by using the 6-31G** basis set, as implemented in Gaussian 09 D.01.[32] Electrostatic potential-derived charges were computed according to the Merz-Kolmann partitioning scheme and converted into RESP charges with the *resp* module of AmberTools 2016.[33] For the enzyme, amber ff99SB force field (FF) was selected[34] and the protonation states for all amino acids at pH 4.5, *i.e.* at lysosomal pH, were calculated from the H++ web-software.[35] The Ca^{2+} ion present in the active site was described using the Improved Model of Hydrated Calcium Ion parameters,[36] maintaining the coordination sphere to the metal observed in the crystal structure.

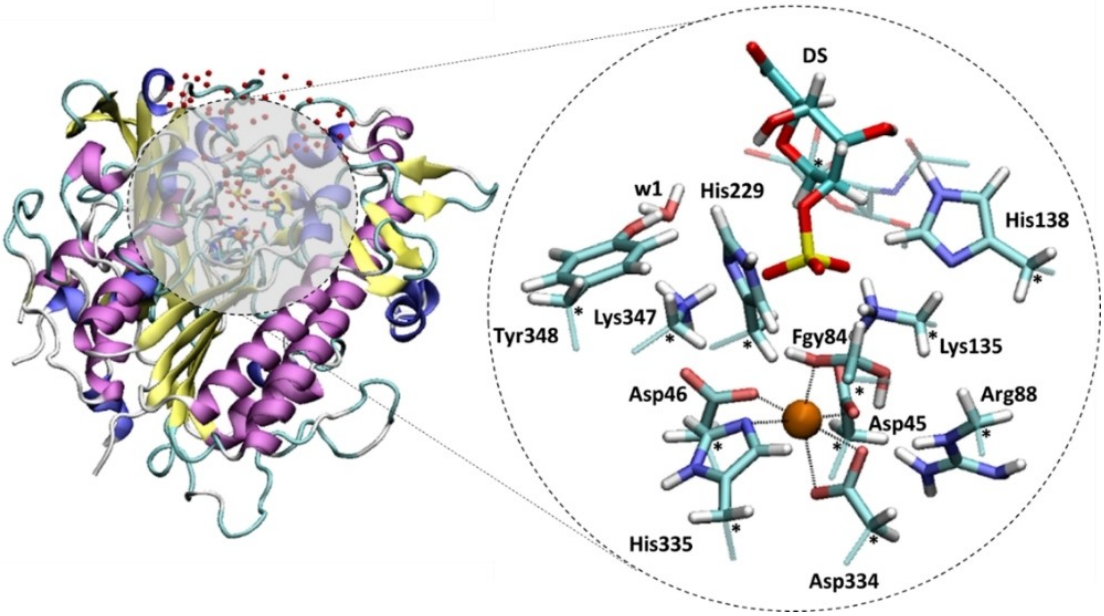


Figure 1. On the left, the adopted model of IDS and the selected QM layer, on the right. Atoms labelled with "*" indicate where the truncation of QM and MM layers occurs.

The enzyme was then solvated in an orthorhombic box of 95×85×99 $Å^3$ containing 71169 TIP3P explicit water molecules,[37] within 12 Å from the enzyme, and neutralized with Na^+ ions, using the Joung and Cheatman parameters.[38] After the minimization of the entire system, carried out following the procedure reported in previous works,[28–31,39,40] a progressive heating phase of 50 ps, from 0 to 310 K was performed. The final phase consisted in the productive simulation for 150 ns at 310 K, 1 atm and NPT ensemble. A cutoff distance of 12 Å, the SHAKE algorithm and PME method were used during all the MD performed simulations. With the aim to analyze the influence of the protonation state of the active site His229 in neutral and protonated form (named Hip229, see below), two different MDs of 150 ns were performed.

Different representative conformations were obtained from RMSD-based clustering of the trajectories by RMS-fitting of the Cα atoms, adopting the average linkage clustering algorithm, as implemented in cpptraj, and identifying 10 representative conformations for each simulation (see SI for details). The most populated structure was then selected to dock the DS substrate, adopting AutoDock Vina program[41] and selecting the Ca^{2+} as box center for the molecular docking study. This procedure allowed getting an initial geometry for the MDs of enzyme-substrate complex (**ES**), for both IDS-His229 and IDS-Hip229 systems. Thus, MDs were performed on the two different ES complexes, following the same protocol adopted for the apo-form enzyme. In order to proceed with the simulations, the parameters for DS were additionally calculated in agreement to the procedure selected to obtain the parameters of FGly84 (see tables in SI).

QM/MM calculations. A QM/MM model was obtained from minimized geometry in water of docked DS to the crystal structure of IDS.[15] The QM/MM calculations were performed applying the two layers ONIOM formalism,[42] including in the high-layer region the Ca^{2+} and its first coordination shell (Asp45, Asp46, FGly84, Asp334, His335) and second coordination shell (Arg88, Lys135, His138, His229 or Hip299, Lys347, Tyr348). One water molecule and the 2-O-sulfo-a-L-iduronate moiety of the substrate were included in the high-layer region, as reported in Figure 1. The QM region of initial model was of 125 atoms, with a total charge of 0 (+1 in the case of Hip229). The low-layer portion of the model consisted of the remaining protein and water molecules within 5 Å from the active site to a final number of 8203 atoms. ONIOM calculations were performed using Gaussian 09[32] software, following the electrostatic embedding scheme,[43] and adopting the B3LYP functional.[44,45] The link atom approach, which consists of saturating the valences on the carbons where the truncation of bonds across the high and low layers occurs (see atoms labeled with "*" in Figure 1), has been applied. The LANL2DZ pseudopotential for Ca^{2+} and the 6-31G(d) basis set for the other atoms were selected for the geometry optimizations,[46] while the MM region was treated at ff99SB level of theory. Atoms within a radius of 15 Å from the high-layer region were geometry-optimized, while the water molecules and the remaining enzyme atoms were fixed in their initial positions. The reaction transition states and intermediates were found using a relaxed linear transit scan. The resulting geometries were then optimized, and the corresponding Hessian matrices were analyzed to confirm that all the reported structures actually represent transition states or minima on the corresponding potential energy surfaces.

As discussed below and in detail, the adopted strategy was designed to identify the rate determining step of the reaction mechanism of IDS, studying the sulphurylation and desulphurylation steps separately.

Results and Discussion

Molecular dynamics simulations of apo-form IDS. The first part of the investigation has been focused on the conformational behavior of the IDS enzyme. In detail, for both apo-form and substrate-bound-IDS, we have considered either neutral (N_δ/N_εH) and protonated (N_δH/N_εH) states of the histidine at position 229. Then, for each system we performed MDs, for a total number of 4×150 ns of trajectories.

As regards of apo-form IDS, the analysis of Root Mean Square Deviations (RMSD, see Figure 2a) plots for IDS-His229 and IDS-Hip229 and the inspection of clustered geometries reveal that both systems reached the equilibrium (see Figure S1). Furthermore, during MDs, neither the secondary structure composition of the protein nor the metal coordination sphere suffered any important variations, as shown in the Figure S2 and Table S1.

Interestingly, RMSD values for IDS-Hip229 slightly oscillate in proximity of average value of 1.92 Å, while those obtained for the IDS-His229 system are at 1.49 Å, (see Figure 2a). In analogy to the RMSD, the Root Mean Square Fluctuations (RMSF) revealed, in general, higher values in the case of IDS-Hip229 system also (see Figure 2b). The behaviors can be related to the more frequent oscillations of amino acid residues composing the active site of IDS-Hip229, as confirmed by the analysis of RMSD of their sidechain atoms (see Figure S3). Such movement is further caused by the presence of higher number of water molecules in proximity of the Ca^{2+}. This observation is supported by the visual inspection of clustered structures, which highlights the increased presence of water molecules in the case of IDS-Hip229 (see Table S2), and by the radial distribution function (RDF) of O^W-O^W pair in the active site proximity (see Figure S4). In addition, the principal component analysis (PCA) reveals, that the solvent-exposed loop in proximity of the active site, composed by residues from Gly171 to Gly184, shifts with respect to its initial position in IDS-Hip229 MDs (see Figure 2c).

Despite these differences, the overall conformation of the active site is well conserved in both systems due to the strong hydrogen bonding network present during the whole simulation time (see Figure S5 for further details). The guanidinium group of Arg88 interacts with carboxylate of Asp45 and Asp334 (salt bridge) and with the backbone of Lys135 and Val135, in both systems. Analogous results are obtained for the imidazole group of His138, which is involved in interaction with the adjacent carbonyl group of the Pro139 and with the hydroxyl group of FGly84, acting as a hydrogen-bond acceptor and donor respectively. The amino group of Lys135 faces the imidazole of His229 and, less frequently, the carboxylate group of Asp46 and Glu166 (see Figure S5).

In the case of the phenol ring of Tyr348, hydrogen bond interaction is observed with Asp269 and Lys347, with major oscillation for the latter in the case of the IDS-Hip229. Indeed, Lys347 further establishes salt-bridge with the carboxylate of Asp46, as observed in the case of IDS-His229, and cation-π interaction with the aromatic moiety of Trp337.

Such behavior can be related to the rotation of imidazole group of Hip229 (see Figure 2d) in IDS-Hip229 system that, to interact with the backbone carbonyl of Ile230, causes the shift of Lys347. The neutral imidazole of His229, instead, establishes hydrogen bond with Asp46, analogously to Hip229, and solvent molecules. The most populated geometry obtained from hierarch-

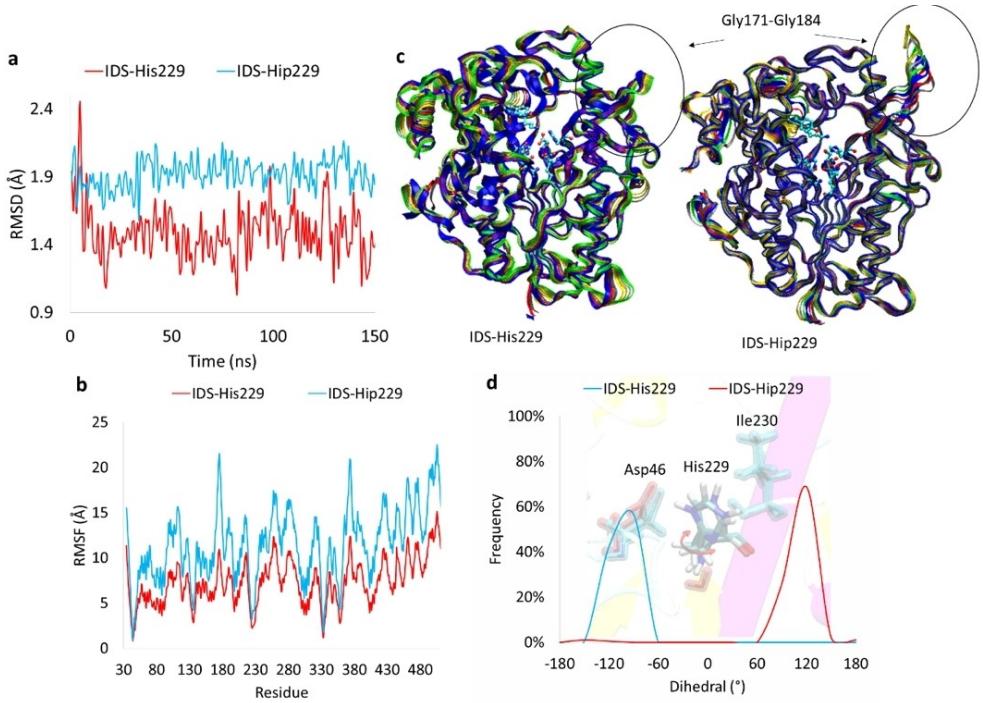


Figure 2. RMSD (a) and RMSF (b) plots of IDS-His229 and IDS-Hip229 systems. Principal component analysis (c) of apo-form IDS enzymes, with His229 (on the left) and Hip229 (on the right). Frequency plot of dihedral angle (d) of His229/Hip229 residue in the apo-form MDs.

ical clustering analysis on both the trajectories, has been selected for the docking of the DS substrate and subsequent MDs on IDS-His229:DS and IDS-Hip229:DS complexes.

Molecular dynamics of substrate-bound IDS. The IDS-His229:DS and IDS-Hip229:DS complexes showed a more stable behavior of protein architecture, as confirmed by lower RMSD (with values oscillating at 1.39 Å and 1.56 Å respectively), and RMSF than the respective apo-forms (Figure S6). The substrate, which shows a stable trend of the RMSD in the active site (see Figure S7), interacts with a different number of amino acids during the simulation time of IDS-His229:DS and IDS-Hip229:DS.

In particular, the carboxylate group of DS is involved in interactions with His229 and Asn177, in IDS-His229:DS, and with Lys469 in IDS-Hip229:DS. In the case of substrate hydroxyl groups, the main hydrogen bonds are established with Tyr315 and Asp269, in IDS-His229:DS, and with FGly84, in IDS-Hip229:DS. The OSO_3^- moiety further interacts with the FGly84 and Lys347, in both systems, and Lys469 and Hip229, in the IDS-His229:DS and IDS-Hip229:DS, respectively (see Figure 3a).

Indeed, the positively charged Hip229 interacted with the negatively charged OSO_3^- group of DS more frequently than the neutral His229, as further confirmed by the RDF calculation considering the S_{DS} and its imidazole group (see Figure S8). Overall, the results of MDs highlighted that second shell residues like His229, either in neutral or protonated state, Lys347, Lys469 and Asn177 can come into play in the recognition and binding phases of the substrate, being crucial for the adequate orientation of DS in the active site. Similar conclusion can be carried out for FGly84, which resulted important for the prior to the catalysis stages, in accordance with previous works.[15,47]

Due to its catalytic role, attention was further focused on the analysis of conformational behavior of FGly84 (see Figure 3b). The O1 atom of the residue averagely lies at 3.88 Å (IDS-His229:DS) and 2.64 Å (IDS-Hip229:DS) from the substrate and at 3.68 Å (IDS-His229:DS) and 3.49 Å (IDS-Hip229:DS) from the Ca^{2+}. Interestingly, during the simulations, the sulfate group of DS faces the $O2_{FGly84}$ more frequently than the $O1_{FGly84}$, as confirmed by the analysis of RDFs of the $O1_{FGly84}$-S_{DS} and $O2_{FGly84}$-S_{DS} pairs. For both protonation states of His229, RDF of $O2_{FGly84}$-S_{DS} presents a peak close to 4.75 Å, two times more intense than the $O1_{FGly84}$-S_{DS} in proximity of 5.50 Å (see Figure 3b). This led to consider both hydroxyl groups of FGly84 as nucleophile agents in the next step of the study on the catalytic mechanism. As will be later highlighted in the QM/MM calculation section, activation of the OH_{FGly84} occurring via deprotonation driven by a general base is required for the catalysis. During the simulations, it was observed that the hydroxyl groups interact via hydrogen bonds with carboxylate group of Asp residues engaged in coordination of the Ca^{2+} (see RDFs in Figure S9). For this reason, the possibility of nucleophile activation practiced by Asp45, Asp46 and Asp334 was further inspected, and it will be discussed in the next section.

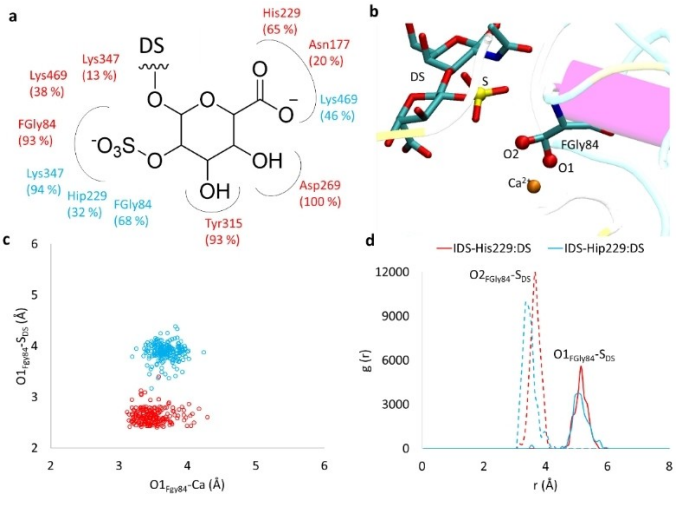


Figure 3. Hydrogen bond network (a) of DS substrate in the MDs IDS-His229:DS (red) and IDS-Hip229:DS (cyan) systems. Atoms involved in the early stage of catalytic mechanism of IDS enzyme (b) and relative distances plot (c). RDFs (d) of $O1_{FGly84}$-S_{DS} and $O3_{Flgy84}$-S_{DS} pairs calculated during the MDs of substrate-bound IDS.

Reaction mechanism. The investigated catalytic mechanism proceeds through a S_N2-E2 elimination mechanism (see Figure 4), in agreement with previous proposals.[15,18,48]

Thus, the mechanistic proposal can be summarized in two different stages, as pointed out in the Figure 4, concerning the sulphurylation and the desulphurylation acted by FGly84 that have been individually investigated. In the first stage, the O1H of FGly84 of **ES** is initially deprotonated by Asp46 (**TS1**), which acts as a proton acceptor, to generate the effective nucleophilic agent (**INT1**). In **TS2**, later, the sulfate group suffers the attack by O1 nucleophile, to form the sulfate ester bounded to the Ca^{2+} and to release the α-L-iduronate (ROH in Figure 4), as described in **INT2**. After a Ca^{2+}-$O1_{FGly84}$ to Ca^{2+}-$O2_{FGly84}$ switch, the intramolecular elimination of hydrogen sulfate (HSO_3^-), occurring in **TS3**, led to the obtainment of FGly84 aldehyde (**EP**), or $FGly84_{ox}$, and concludes the second part of the reaction.

The discussion of the results will mainly focus on the His229-containing system (see energy surface in Figure 4), since the reaction channel generated with the Hip229 resulted in higher energetics, as reported in supporting information.

Sulphurylation of FGly84 and formation of α-L-Iduronate. In the optimized **ES**, the oxygens of the substrate interact with two positively charged residues, Lys135 (1.98 Å) and Lys347 (1.96 Å, see Figure 5). In addition, His229 establishes H-bond with the –OSO_3 group (N–H–O distance of 2.02 Å). Such interactions generate a fruitful orientation of sulfate group in facing the O1 (3.31 Å) and O2 (4.42 Å) atoms of FGly84 (see Figure 5). In **TS1**, the deprotonation on O1 takes place by the base Asp46, as confirmed by the distances of 1.23 Å and 1.19 Å, for O1-H and H-O_{Asp46} respectively, with a relative barrier of 3.3 kcal/mol, as reported in Figure 4. In the **INT1** (+0.9 kcal/mol above **ES**), the activated O1 coordinates the Ca^{2+} (2.35 Å) and faces the S of the substrate (4.38 Å, see Figure S10). In **TS2**, at 21.6 kcal/mol with respect to the **ES**, the nucleophilic substitution on the S atom occurs through a S_N2 mechanism. Indeed, the formation of S–O1 bond (1.88 Å) and the breaking of O_{DS}-S_{DS} (2.00 Å) take place almost concertedly, as further confirmed by the frequency of 198*i* cm^{-1}.

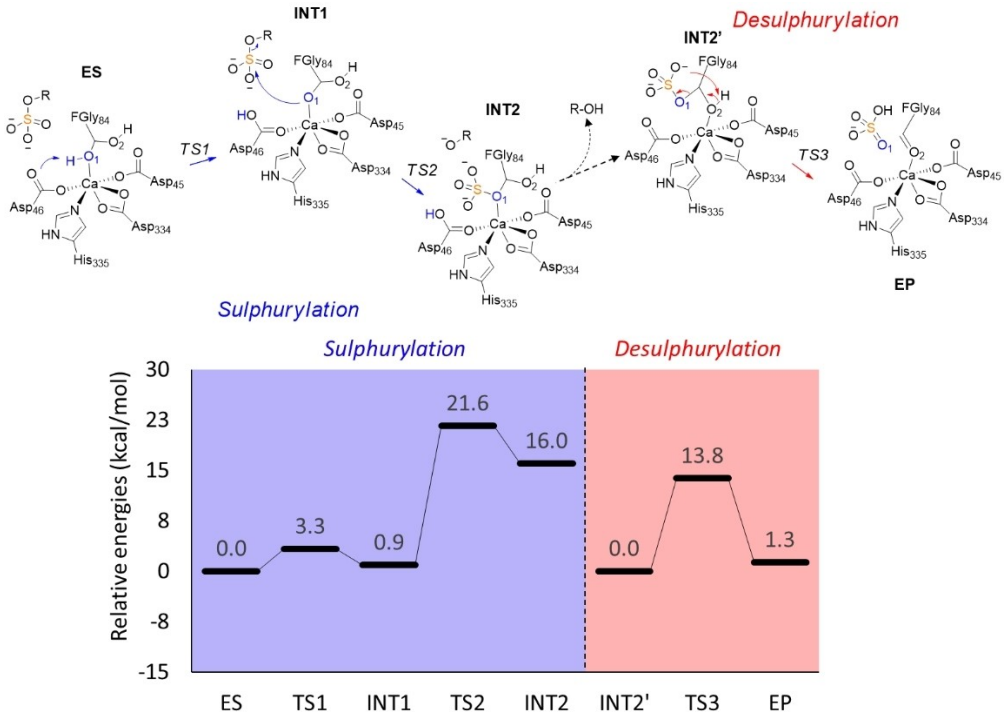


Figure 4. The two stages of the catalytic mechanism of IDS enzyme (top) and the related energy profiles calculated for the sulphurylation and desulphurylation stages of reaction catalyzed by IDS (bottom).

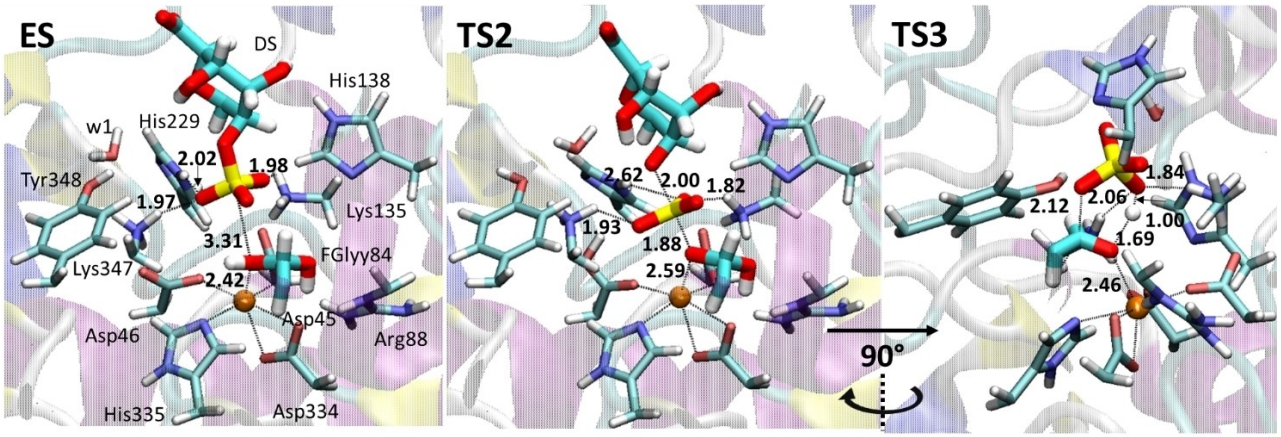


Figure 5. Optimized geometries of **ES**, **TS2** and **TS3**. Distances are reported in Å. For clarity, most of the atoms have been omitted.

The possibility of concerted deprotonation of $O1H_{FGly84}$ and nucleophilic attack to the S_{DS} was further tested, but during the optimization such transition state returned the **ES** structure. The first stage of the reaction mechanism is completed by the formation of α-L-iduronate in **INT2**, which lies at + 16.0 kcal/mol (see Figure 4). In the optimized **INT2**, the alkoxide moiety of α-L-iduronate lacks stabilization from amino acids in the active site, thus resulting in an endothermic reaction. It cannot be excluded however that, after **INT2**, the forming α-L-iduronate can be neutralized by neighbor residues during the release-step of the catalytic cycle with a more favorable thermodynamic. It is worth nothing that the study of this stage was not object of the current investigation. By the way, calculations on a model of the uncatalyzed reaction confirm the thermodynamic of the considered chemical event, highlighting furthermore the role played by the IDS in lowering the activation barrier by *ca* 40 kcal/mol (see Figure S11).

Since the MD revealed the presence of -O2H moiety of FGly84 in proximity of the $-OSO_3$ group of the substrate, its involvement in the catalytic reaction has been also tested. In Figure S12 are shown the optimized structures of the two corresponding "key" transition states, **TS1-O2** and **TS2-O2**. In the case of **TS1-O2**, the Asp334 was considered for the deprotonation of hydroxyl group. The energy barriers of **TS1-O2** and **TS2-O2** involving $O2_{FGly84}$ resulted 11.4 and 34.6 kcal/mol, 8.1 and 13.0 kcal/mol higher than the relative transition states with $O1_{FGly84}$ respectively (see Fig-

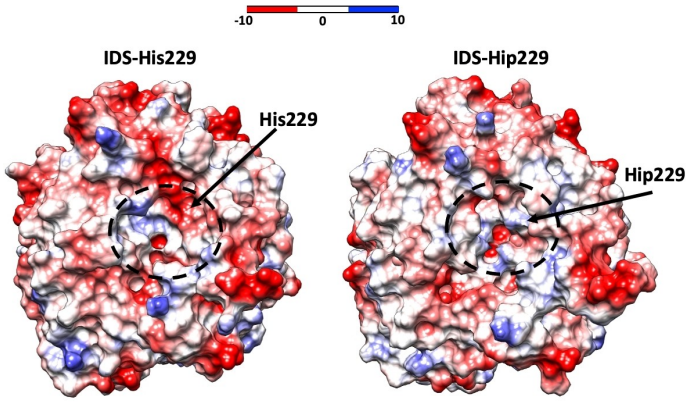


Figure 6. Electrostatic potential surfaces of the IDS-His229 (left) and IDS-Hip229 (right) structures. The circled region indicates the catalytic pocket.

ure S13). This difference can be attributed to the presence of Ca^{2+}, which in **TS1** and **TS2** is coordinated by $O1_{FGly84}$. Due to the Lewis acidity of Ca^{2+}, $O1_{FGly84}$ can be more easily deprotonated, (in **TS1**) and can be considered a better nucleophile than $O2_{FGly84}$ (in **TS2**). This behavior of Ca^{2+} highlights its catalytic role in the process. Furthermore, with the involvement of Asp334 for the deprotonation of the $O2_{FGly84}$, the hexacoordination of Ca^{2+}, observed in the **ES**, is not retained, thus leading to higher energies of the transition states.

The contribution to the catalysis carried out by the positively charged Hip229 was further tested, starting from the optimized **ES** and **TS2**, adding one proton to the His229. For the **TS2-Hip229**, a relative energy of 26.3 kcal/mol has been calculated with respect to **ES-Hip229**, which is 4.6 kcal/mol higher than the respective His229 (see Figure S13).

The presence of Hip229 generated a more positive environment of active site, as can be observed from the analysis of APBS (Adaptive Poisson-Boltzmann Solver) potential exhibited in Figure 6, and a different hydrogen bonds network in proximity of FGly84 and the SO_3^- group, as similarly observed in previous work.[28] Indeed, the analysis of the geometries highlighted that the **ES-Hip229** presents the hydrogen bonds between the $-OSO_3^-$ group of DS and $O1H_{FGly84}$ and $-NH_3^+$ moiety of Lys347, which are not observed in the **TS2-Hip229** (see Figure S14). These interactions can be the source of stronger stabilization of the enzyme-substrate complex with a consequent increasing of the transition state energy barrier. At this purpose it is interesting to observe an opposite behavior in presence of His229, with the above discussed hydrogen bond occurring in the transition state but not in the corresponding enzyme-substrate complex giving rise to a more stabilized transition state in agreement with the theory of the enzymatic catalysis.[49]

As final comment, at this stage the enzyme has completed its function of catalyst, and the main product has been released. As previously mentioned, we have also considered formation of the HSO_4^-, as discussed in the next section.

Desulphurylation of FGly84/formation of HSO_4^-. In this step of the reaction mechanism, the optimized **INT2'** is characterized by the presence of Ca^{2+} coordinate to the O2 of FGly84. The relative intermediate with the Ca^{2+}–O1 bond is *ca* 10 kcal/mol higher in energy and attempts in the finding of TS, concerning the O1–Ca^{2+} to O2–Ca^{2+} switch were not fruitful (see Figure S14). For this reason, to study the second stage of the reaction, we considered **INT2'** as starting point. The intramolecular elimination of HSO_4^- takes place through deprotonation of O2 hydroxyl group and concerted breaking of C_{FGly84}–$O1_{FGly84}$ bond (2.12 Å vs 1.46 Å **INT2'**, see Figure 5). The related **TS3** has an energy of +13.8 kcal/mol, with respect to the **INT2'**. The subsequent **EP** (1.3 kcal/mol, respective to **INT2'**) is characterized by the oxidized form of FGly84 (FGly84ox), having the remaining O2 linked to the Ca^{2+}.

The overall analysis of the reaction energy profiles reveals that the **TS2** describing the S_N2 attack at the sulphur atom of the DS substrate, *i.e* the sulphurylation of FGly84, is the rate determining step. The calculated barrier of 21.6 kcal/mol is in agreement with the measured k_{cat} of 2 s^{-1} in vitro, corresponding to ~18 kcal/mol.[50,51] Notably, the obtained value fits well with previous studies concerning the sulphurylation of many compounds, as similarly calculated for human ARSA,[17] arylsulfatase from *Pseudomonas aeruginosa*[52] and computationally and experimentally observed for different diester substrates.[53–55]

Conclusions

In the present work, Ca^{2+}-dependent IDS, an enzyme involved into the lysosomal degradation of dermatan sulfate and heparan sulfate which are two important glycosaminoglycans with many biological roles, was studied *via* a multiscale modelling approach (MDs, QM/MM). IDS deficiency and modifications are crucially linked to the Hunter syndrome, a rare childhood disease, and therefore represents a target for new possible therapies.

Starting from the recently published crystal structure, a model of IDS was built-up and used for molecular dynamics simulations aiming to study the conformational behavior of the enzyme. The simulations were carried out for apo and DS-bounded forms of IDS. Two different protonation states for His229, a residue belonging to the second coordination sphere of the Ca^{2+}, were considered due to the crucial role in the

binding of the substrate and in the catalysis. In the case of apo-form IDS, the active site architecture is affected by two protonation states of the His229, since the presence of different hydrogen-bond interactions involving the amino acid residues of the catalytic pocket was observed.

In the case of IDS:DS complexes, the post-MD analysis highlighted the role of His229/Hip229 and Lys347, Asn177 and Lys469 residues in the recognition and binding of the substrate. In addition, MD analysis permitted to identify that the FGly84 can further play an important role in the stages prior to the catalysis.

Moreover, QM/MM calculations were performed to investigate the chemical events catalyzed by IDS. The studied reaction mechanism is a two-steps process taking place in the sulphurylation and desulphurylation stages, and the rate determining step concerns the attack of $O1_{FGly84}$ to the sulphate group of IDS, *i.e.* in the former step. The protonated histidine (Hip229) has turned higher energy in the rate determining step. The desulphurylation takes place through an intramolecular E2 mechanism leading to the formation of $FGly84_{ox}$.

The insights provided by the current calculations can be helpful for a deeper understanding at molecular level of IDS, and authors hope it can support the design of new applications targeting enzyme in new therapies.

[1] C. Hunter, *Proc. R. Soc. Med.* **1917**, *10(Sect Study Dis Child)*, 104–116.
[2] S. Mohamed, Q. Q. He, A. Singh, V. Ferro, *Mucopolysaccharidosis type II (Hunter syndrome): Clinical and biochemical aspects of the disease and approaches to its diagnosis and treatment.* In: Baker, D. C editors. Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, London, **2020**, pp. 71–117.
[3] J. E. Wraith, M. Scarpa, M. Beck, O. A. Bodamer, L. D. De Meirleir, N. Guffon, A. Meldgaard Lund, G. Malm, A. T. Van der Ploeg, J. Zeman, *Eur. J. Pediatr.* **2008**, *167*, 267–277.
[4] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, *Essentials of Glycobiology, 2nd edition*, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, **2009**.
[5] E. F. Neufeld, J. Muenzer, *The mucopolysaccharidoses*, In: C. R. Scriver, A. L. Beaudet, W. S. Sly, et al., editors. The Metabolic and Molecular Bases of Inherited Disease, 8th ed., New York: McGraw-Hill, **2001**, pp. 3421–3452.
[6] V. A. McKusick, E. F. Neufeld, *The mucopolysaccharidoses storage diseases.* In: J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, et al., editors. The Metabolic and Molecular Bases of Inherited Disease, 5th ed., New York: McGraw-Hill, **1983**, pp. 751.
[7] M. Schwarz, P. Skrinjar, M. J. Fink, S. Kronister, T. Mechtler, P. I. Koukos, A. M. J. J. Bonvin, D. C. Kasperb, H. Mikula, *Chem. Sci.* **2020**, *11*, 12671–12676.
[8] M. P. Cosma, S. Pepe, I. Annunziata, R. F. Newbold, M. Grompe, G. Parenti, A. Ballabio, *Cell* **2003**, *113*, 445–456.
[9] K. Sugahara, T. Mikami, *Curr. Opin. Struct. Biol.* **2007**, *17*, 536–545.
[10] J. M. Trowbridge, R. L. Gallo, *Glycobiology* **2002**, *12*, 117R–125R.
[11] A. Malmström, B. Bartolini, M. A. Thelin, B. Pacheco, M. Maccarana, *J. Histochem. Cytochem.* **2012**, *60*, 916–925.
[12] M. A. Thelin, B. Bartolini, J. Axelsson, R. Gustafsson, E. Tykesson, E. Pera, Å. Oldberg, M. Maccarana, A. Malmstrom, *FEBS J.* **2013**, *280*, 2431–2446.
[13] G. Parenti, G. Meroni, A. Ballabio, *Curr. Opin. Genet. Dev.* **1997**, *7*, 386–391.
[14] J. J. Hopwood, A. Ballabio, *Multiple sulfatase deficiency and the nature of the sulfatase family.* In: C. R Scriver, A. L Beaudet, D Valle, W. S Sly (Eds.), The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill, New York, **2001**, pp. 3725–3732.
[15] M. Demydchuk, C. Hill, A Zhou, G. Bunkóczi, P. E. Stein, D. Marchesan, J. E. Deane, R. J. Read, *Nat. Commun.* **2017**, *8*, 15786.
[16] R. Froissart, G. Millat, M. Mathieu, D. Bozon, I. Maire, *Biochem. J.* **1995**, *309*, 425–430.
[17] T. Marino, M. Toscano, N. Russo, *Chem. Eur. J.* **2013**, *19*, 2185–2192.
[18] D. Ghosh, *Methods Enzymol.* **2005**, *400*, 273–293.
[19] S. R. Hanson, M. D. Best, C. H. Wong, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 5736–5763.
[20] X. Zhang, L. Lin, H. Huang, R. J. Linhardt, *Acc. Chem. Res.* **2020**, *53*, 335–346.
[21] G. Lukatela, N. Krauss, K. Theis, T. Selmer, V. Gieselmann, K. von Figura, W. Saenger, *Biochemistry* **1998**, *37*, 3654–3664.
[22] C. S. Bond, P. R. Clements, S. J. Ashby, C. A. Collyer, S. J. Harrop, J. J. Hopwood, J. M. Guss, *Structure* **1997**, *5*, 277–289.
[23] Z. Zhang, M. Ma, W. Zhang, Y. Zhou, F. Yao, L. Zhu, M. Wei, Z. Qiu, *Front. Genet.* **2023**, *14*, 1103620.
[24] R. Mashima, M. Ohira, T. Okuyama, M. Onodera, S. Takada, *Sci. Rep.* **2023**; *13*: 7865.
[25] P. M. Franchi, N. Kulagina, A. Ilinskaya, B. Hoffpauir, M. G. Qian, H. Sugimoto, *AAPS J.* **2023**, *25*, 61.
[26] C. Schöneich, *Biomol. Eng.* **2023**, *13*, 1142.
[27] J. E. Wraith, M. Beck, R. Giugliani, J. Clarke, R. Martin, J. Muenzer, *Genet. Med.* **2008**, *10*, 508–516.
[28] M. Prejanò, I. Romeo, L. Sgrizzi, N. Russo, T. Marino, *Phys. Chem. Chem. Phys.* **2019**, *21*, 23338–23345.
[29] M. Prejanò, I. Romeo, M. A. La Serra, N. Russo, T. Marino, *J. Chem. Inf. Model.* **2021**, *61*, 5883–5892.
[30] M. Prejanò, T. Marino, N. Russo, *Front. Chem.* **2018**, *6*, 606.
[31] M. Prejanò, I. Romeo, N. Russo, T. Marino, *Int. J. Mol. Sci.* **2020**, *21*, 4551.
[32] Gaussian 09, Revision D. 01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta,

F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

[33] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman (2016), AMBER 2016, University of California, San Francisco.

[34] L. Wickstrom, A. Okur, C. Simmerling, *Biophys. J.* **2009**, *97*, 853–856.

[35] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537–541.

[36] J. Yoo, J. Wilson, A. Aksimentiev, *Biopolymers* **2016**, *105*, 752–763.

[37] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[38] I. S. Juong, T. E. Cheatham III, *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

[39] F. E. Medina, M. Prejanò, *ACS Catal.* **2021**, *11*, 4136–4145.

[40] M. Prejanò, F. E. Medina, M. J. Ramos, N. Russo, P. A. Fernandes, T. Marino, *ACS Catal.* **2020**, *10*, 2872–2881.

[41] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, *30*, 2785–2791.

[42] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* **1996**, *100*, 19357–19363.

[43] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, Jr., K. Morokuma, M. J. Frisch, *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

[44] A. D. Becke, *J. Phys. Chem.* **1993**, *98*, 5648–5652.

[45] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[46] W. R. Wadt, P. J. Hay, *J. Chem. Phys.* **1985**, *82*, 284–298.

[47] B. Schmidt, T. Selmer, A. Ingendoh, K. von Figura, *Cell* **1995**, *82*, 271–278.

[48] M. J. Appel, C. R. Bertozzi, *ACS Chem. Biol.* **2015**, *10*, 72–84.

[49] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, M. H. Olsson, *Chem. Rev.* **2006**, *106*, 3210–3235.

[50] C. Kim, J. Seo, Y. Chung, H.-J. Ji, J. Lee, J. Sohn, B. Lee, E. Jo, *J. Hum. Genet.* **2017**, *62*, 167–174.

[51] Y. K. Chung, Y. B. Sohn, J. M. Sohn, J. Lee, M. S. Chang, Y. Kwun, C. H. Kim, J. Y. Lee, Y. J. Yook, A.-R. Ko, D.-K. Jin, *Glycoconjugate J.* **2014**, *31*, 309–315.

[52] B. van Loo, R. Berry, M. F. Mohamed, M. Golicnik, A. C. Hengge, F. Hollfelder, *Biochemistry* **2019**, *58*, 1363–1378.

[53] R. Wolfenden, Y. Yuan, *PNAS* **2007**, *104*, 83–86.

[54] K. Szeler, N. H. Williams, A. C. Hengge, S. C. L. Kamerlin, *J. Org. Chem.* **2020**, *85*, 6489–6497.

[55] J. M. Younker, A. C. Hengge, *J. Org. Chem.* **2004**, *69*, 9043–9048.