A new view of missense mutations in α-mannosidosis using molecular dynamics conformational ensembles

Špela Mandl,[†] Bruno Di Geronimo,[*,†,‡] Santiago Alonso-Gil,[¶,§] Gibu George,[‖]
Ulrika Ferstl,[†] Sereina Annik Herzog,[⊥] Bojan Žagrović,[¶,§] Christoph Nusshold,[†]
and Pedro A. Sánchez-Murcia[*,†,#]

†*Laboratory of Computer-Aided Molecular Design, Division of Medicinal Chemistry,
Otto-Loewi Research Center, Medical University of Graz, Neue Stiftingtalstr. 6/III, A-8010
Graz, Austria*

‡*Present address: School of Chemistry and Biochemistry, Georgia Institute of Technology,
Atlanta, Georgia, USA*

¶*Max Perutz Labs, Vienna Biocenter Campus (VBC), Campus Vienna Biocenter 5, 1030,
Vienna, Austria*

§*Department of Structural and Computational Biology, Vienna BioCenter University of
Vienna Campus-Vienna-Biocenter 5, A-1030 Vienna, Austria*

‖*Institut de Química Computacional i Catàlisi and Departament de Química, Universitat
de Girona, C/ Maria Aurèlia Capmany 69, 17003 Girona, Catalonia, Spain*

⊥*Institute for Medical Informatics, Statistics and Documentation, Medical University of
Graz, Neue Stiftintalstr. 6/III, A-8010 Graz, Austria*

#*BioTechMed-Graz, Mozartgasse 12/II, A-8010 Graz, Austria*

E-mail: bruno.digeronimo@medunigraz.at; pedro.murcia@medunigraz.at

1

Abstract

The mutation of remote positions on enzyme scaffolds and how these residue changes can affect enzyme catalysis is still far from being fully understood. One paradigmatic example is the group of lysosomal storage disorders, where the enzyme activity of a lysosomal enzyme is abolished or severity reduced. In this work, we analyze molecular dynamics simulation conformational ensembles to unveil the molecular features controlling the deleterious effects of the 43 reported missense mutations in the human lysosomal α-mannosidase. Using residue descriptors for protein dynamics, their coupling with the active site, and their impact on protein stability, we have assigned the contribution of each of the missense mutations into these three categories. We demonstrate here that the use of conformational ensembles is a powerful approach not only to better understand missense mutations at the molecular level, but to revisit the missense mutations reported in lysosomal storage disorders in order to aid the treatment of these diseases.

Introduction

Lysosomal storage disorders (LSDs)[1–3] are a paradigm of how single-point mutations in an enzyme can disrupt or abolish its activity and cause a pathology.[1–4] Looking at the location of these mutations in LSDs-related enzymes, the reported clinical mutational landscape includes not only positions close to the active site, but also remote ones more than 10 Å away from it. Computationally, the effect of the orthosteric mutations can in many cases be assessed using the 'chemical sense' or by calculating properties of the active site in the presence of the mutation. In contrast, the understanding of how remote mutations allosterically affect the enzyme activity[5] in LSD is still far to be understood.[6] The remote mutation of the side chain of a protein residue in an LSD-related enzyme can impact its network of amino-acid interactions in different ways.[7,8] On the one hand, some of these mutations can prevent the correct folding of the enzyme in the endoplasmic reticulum and/or its transport

to the lysosome.[9,10] The computation of the change of the free energy of the protein scaffold upon mutation as $\Delta\Delta G_{X\to Y}$ has been shown useful to assess the impact of such pathogenic mutation on the integrity of the enzyme.[11,12] On the other hand, other mutations may not affect the enzyme folding, but reduce significantly its enzyme activity.[13] Whereas the former group of mutations affect the folding process of the enzyme, the latter group may introduce changes on the local interactions and/or the dynamics and flexibility of the protein scaffold, causing 'frustrated networks in proteins'.[14] Recently, the possibility to access a protein structure from its sequence via deep-learning algorithms has allowed the exploration of these amino-acid networks from a structural perspective.[15–17] As an example, there are available computational tools to detect the network of interactions between remote mutations and the active site when the structure is available.[18–20] The protein structure can also be used as input file to study the role of enzyme conformational dynamics, an important factor in catalysis.[21–23] Indeed, via classical molecular dynamics (MD) simulations, conformational ensembles of the enzyme with explicit solvation can be obtained.[24,25] Importantly, by analysis of those conformational ensembles it can be detected the role of remote residue on catalysis.[26–28]

Human lysosomal α-mannosidase (hLAMAN, EC 3.2.1.24) is part of the GH38 family of Zn-dependent aspartic glycosidase hydrolases.[29–32] It catalyzes the cleavage of α-mannosidic linkages $\alpha(1 \to 2)$, $\alpha(1 \to 3)$, and $\alpha(1 \to 6)$ from oligosaccharides. Lack of hLAMAN activity causes the autosomal recessive disease termed α-mannosidosis (MANSA).[33–36] MANSA is a LSD associated with more than 130 pathogenic variants.[37,38] The effect of the 43 reported missense mutations in MANSA has been classified depending on their effect on the protein activity or stability (Table 1): 17 of them reduce or abolish the enzyme activity, 23 yield an incorrect enzyme folding or incomplete trafficking to the lysosome, and 3 of them has not been totally classified yet. As examples, D74E, H200L, or H200N,[39,40] which lead to alterations in the coordination of metal ions at the active site,[41] belong to the first group of mutations, whereas R750W, which occurs in over 25% MANSA patients, disrupts the

3

proper folding of the enzyme.[36] Although the macroscopic effects of these mutations are known, and the malignancy of missense mutations can be predicted by tools like ClinVal,[42] the specific molecular mechanisms behind these defective mutations are not yet well understood. Amongst other causes, the lack of an experimental structural model of hLAMAN. hLAMAN is closely related to bovine lysosomal α-mannosidase (bLAMAN, EC 3.2.1.24) with an 80% of similarity, for which one crystallographic structure is available (PDB id. 1O7D).[30] bLAMAN is a homodimer, where each of the monomer is formed by five peptides (labelled as A-E). Analogously, it is known that hLAMAN is cleaved into three peptides: one of 70 kDa, one of 42 kDa (referred to as D), and one of 13/15 kDa (referred to as E).[43] The 70 kDa peptide is subsequently cleaved into three additional peptides (A, B, and C). These A, B, and C peptides are connected by disulfide bonds and exhibit a high degree of glycosylation. Interestingly, although the closest human Golgi α-mannosidase II (hAMan II, EC: 3.2.1.114) displays similar kinetic properties, a 81% amino-acid similarity or a similar inhibitor sensitivity/substrate specificity,[43–45] this cytoplasmic enzyme appears as a single monomer.[46]

In this work, we have analyzed all-atom conformational ensembles of hLAMAN generated via classical MD simulations in explicit aqueous solvent to unveil the key residues at the hLAMAN dimer interface and the molecular mechanism behind the reported missense mutations across the protein structure (Figure 1). This way, we have been able to compute the degree of coupling of these mutations with the active site of the enzyme over its protein network, the impact of these mutations on the stability of the enzyme, and the role of enzyme dynamics. As outcome, we have quantified each of the former factors for each of the reported mutations in hLAMAN. We show that by studying both the structure and the dynamic of enzymes in LSDs, new insights can be gained into the disease mechanism for future targeted therapies.[47] Although the correlation between the type of missense mutation and the clinical manifestations is not evident in many LSDs, we believe our approach may help to shed light
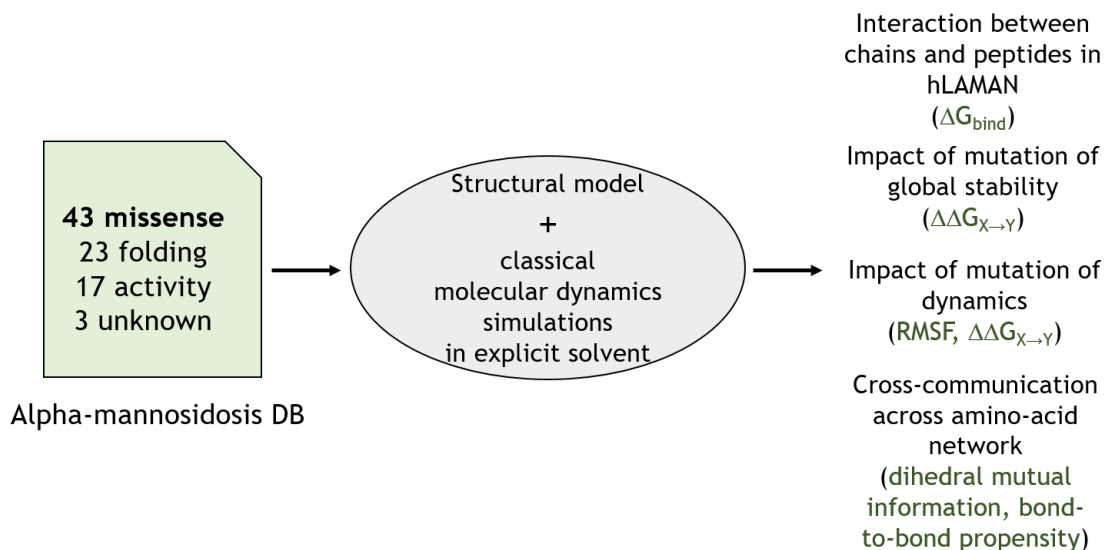
4

in this regard.

Figure 1: Analysis of hLAMAN MD ensemble to unveil the role of malign missense mutations.

Results and discussion

Quaternary structure of hLAMAN

We built-up a model for the quaternary structure of hLAMAN as a polypeptide homodimer $(A-E)_2$ based on experimental evidences (Figure 2A). The expression of the enzyme in *Komagataella phaffii* shows a band at ca. 230 KDa in both SDS-PAGE gels and Western-Blot under native conditions, which corresponds to the hLAMAN homodimer of the polypeptide A-E (Figure 2B). These results are in line with others,[43] and previous size exclusion chromatography data.[48] Noteworthy, we had difficulties to observe this band in the secreted protein fraction (supernatant fraction). This fact may indicate the need of the intracellular processing to complete the maturation of the protein. It is reported that each of the polypeptides of hLAMAN consist of five chains (named A-E). hLAMAN is biosynthesized as a polypeptide of ca. 120 KDa in the endoplasmatic reticulum (ER) and transported then to the lysosome.[49] Along this journey, the maturation of the hLAMAN polypeptide

yields three fragments of ca. 70 (chains A, B, C cross-linked via four disulfide bonds), 40 (chain D), and 15 (chain E) KDa, respectively. Additionally, the recombinant hLAMAN expressed in different eukaryotic heterologous systems (e.g., Chinese hamster ovary (CHO) cells,[48] COS-1 cells,[49] or *K. phaffii*[43]) may be secreted as the whole precursor polypeptide (120 kDa). Interestingly, this recombinant precursor evolves when concentrating the protein crude by cleavage between chains C and D, with the appearance of bands at ca. 70 and 55 KDa in SDS-PAGE gel.[48] Looking at the crystallographic symmetry in bLAMAN, the most plausible interaction between polypeptides in the homodimer of bLAMAN and hLAMAN ((A-E)$_2$) is via chains A and A' (Figure 2A). hLAMAN structure consist of 23 α-helices and 50 β-sheets (Figure S1). The analysis of the molecular interfaces of our hLAMAN structure and the analog structure of bLAMAN (PDB id. 1O7D)[30] with the PISA server[50] points out that the most probable assembly would be the one corresponding to the quaternary structure of a single A-E assembly (chains A-E, Figure 2A). The active site of hLAMAN is located at the interface between chains A, C and E within each A-E unit. It is composed of a Zn^{2+} atom coordinated by three residues of chain A (H72, D74, and D319) and one residue of chain C (H446). D74 and D319 on chain A act in catalysis as nucleophile and general acid, respectively (Figure 2C).[51] Interestingly, all these residues are located on loops between α-helices (Figure S1).
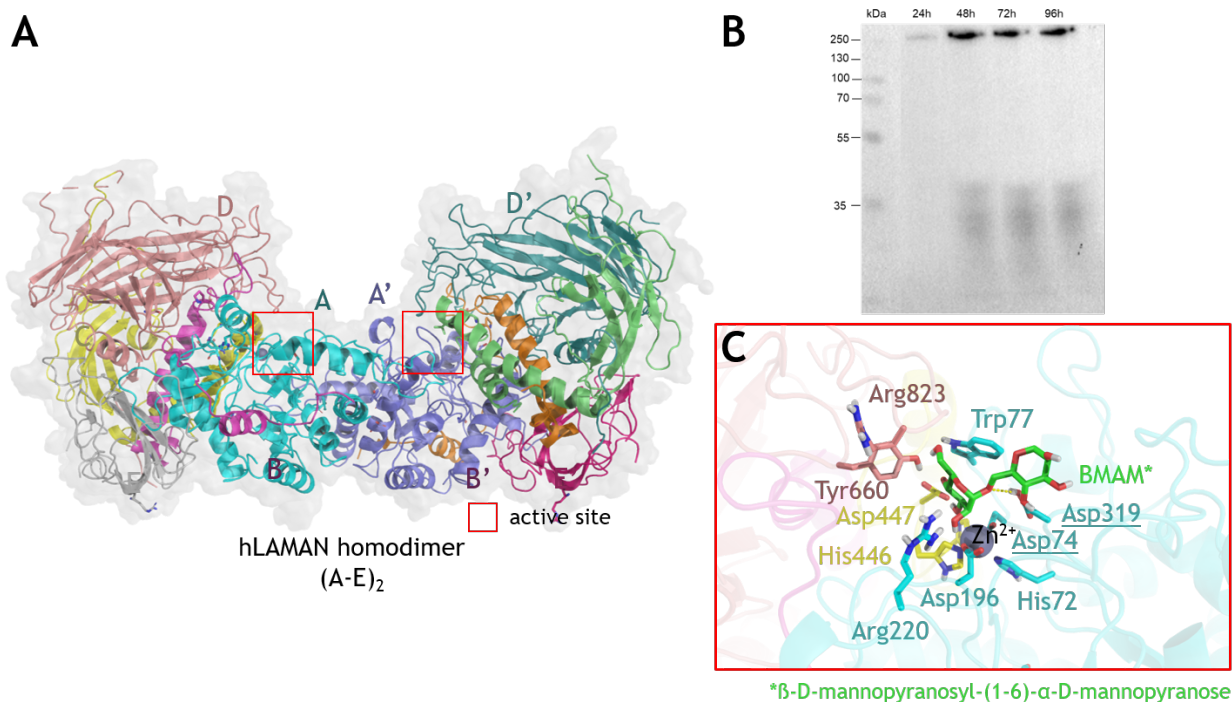
6

Figure 2: (A) hLAMAN quaternary structure. Peptides of the second polypeptide are labeled as A'-E'. (B) Western-Blot of hLAMAN expressed in *K. phaffii* under native conditions at different incubation times. (C) Detail of the active site of hLAMAN in the presence of the substrate BMAM (β-D-mannopyranosyl-(1,6)-α-D-mannopyranose). Relevant residues at the active site are shown in stick. The two catalytic acid residues are underlined.

We studied the interactions between subunits in hLAMAN. To do this, we embedded hLAMAN into a box of water molecules and ran μs-scale molecular dynamics (MD) simulations of the systems. We studied the energetic contributions between interfaces (see Methods for further details). In Figure 3A are shown the binding energy distributions between peptides A-E in the hLAMAN monomer computed along the MD simulation trajectories. The interaction between peptides A and B has the larger energetic contribution to the stability of the protein complex, whereas the other cross-interactions between chains (e.g., B-C, D-E or C-D) contribute to a smaller extend. There are still some differences between them: whereas chain B has a strong interaction with chain C, the interaction between D and E, although favorable, has a much smaller contribution. In all these cases, their mean values for the interactions between A and C, B and E or D and E are below 10 kcal mol^{-1}.
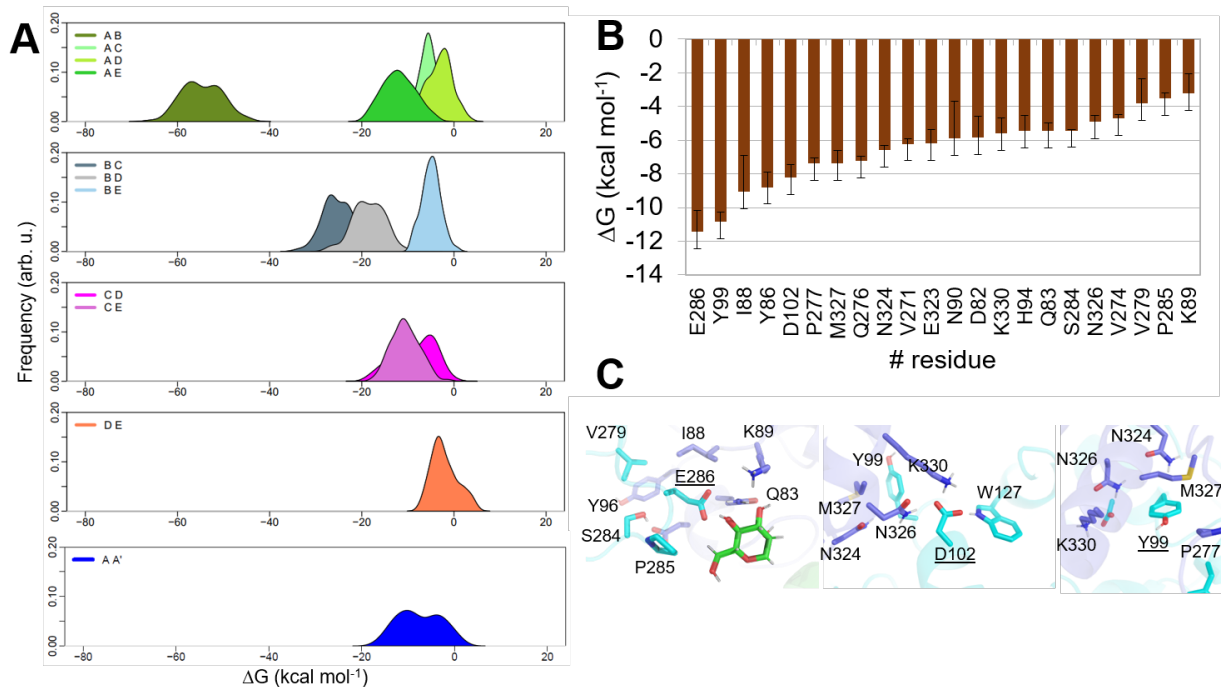
7

Figure 3: (A) Computed binding energy (kcal mol^{-1}, ± SD) between peptide chains in hLAMAN along the MD simulations. (B) Relevant residues at the A-A' interface and energy contribution per residue (kcal mol^{-1}). (C) Detail of the interactions of E286, D102 and Y99 at the homodimer interface of hLAMAN.

We also explored the interactions of the polypeptide interface (A-A' interface). The hLAMAN homodimer interacts mainly via the two A-A' chains through helices H3 and H11. Indeed, it is observed a favorable interaction between A and A' is observed along the MD trajectories (Figure 3B). Looking in detail at the individual energy contribution of the interfacial residues, it is found that E286, Y99, I88, Y86, D102, P277, M327 and Q276, amongst others, as the relevant ones for the stability of the homodimer (Figure 3B). The side chain of E286 establishes a salt-bridge with K89 and a hydrogen bond with Q83. A hydrogen bond is also observed with an α-D-mannosepyranose from the other subunit A'. Importantly, the carboxylic acid of the side chain of D102 (located on H3) establishes a hydrogen bond with the indole nitrogen of W127 (chain A), the amide nitrogen of the side chain of N326, and ammonium group of K330. The latter two residues are located on the other chain A' (Figure 3C). Two residues important for the energetics of the interface have been reported as

8

missense mutations in patients. The mutation D102N causes a drop of the enzyme activity up to 85 % of the wild-type counterpart activity. The hydrogen-bond-acceptor character of both oxygen atoms of D102 side chain are essential to keep these three interactions between A and A'. As a consequence, the introduction of an amide function would not make such interaction network possible. The change Y99H also reduces significantly the enzyme activity. The phenol ring in Y99 interacts with the side chain of M327. A hydrogen bond between the phenol oxygen of Y99 and the backbone of Y86 orientates the phenol ring. Thus, the introduction of an imidazol in such position may imbalance the polarity of this section of the interface. Altogether, E386, Y99 and D102 are in close proximity on helices H3 and H11. This regions is critical for the integrity of the hLAMAN homodimer.

Reported missense mutations and hLAMAN structure

43 deleterious missense mutations in hLAMAN has been reported in MANSA so far (Table 1).[38] A first group of 23 mutations has been shown to prevent the protein folding of the enzyme in the ER; 17 of them to reduce the activity of hLAMAN, but do not to affect its folding, and 3 of them has not been classified yet. Looking at the location of the reported mutations affecting the activity on the structure of hLAMAN, it can be seen that most of them are at a distance above 10 Å from the active site. Very few of the mutations (e.g., H72L, D74E or P197R) are located in the proximity of the active site. This highlights the challenging nature of explaining the loss or reduction of enzyme activity by simple visual inspection of the protein structure.

Table 1: Reported deleterious missense mutations in hLAMAN.[38]

Mutant	Chain	Rel. activity (%)*	Distance (Å)**	Sec. structure	Contact***
(a) Protein folding					
C55F	A	9	27.3	SS-bond	B
A95P	A	10	13.0	loop	-
P197R	A	12	7.0	loop	C
R202P	A	24	20.0	α-helix 8	C,D
L352P	A	11	26.0	β-sheet 9	-
T355P	B	9	19.4	$alpha$-helix 13	-
P356R	B	6	17.9	α-helix 13	-
G390C	B	7	17.1	β-turn	D
H445Y	C	6	8.2	loop	A
S453Y	C	19	10.2	loop	A
S453F	C	19	10.2	loop	A
L565P	C	22	45.1	β-sheet 20	D
W714R	D	5	32.1	β-sheet 31	-
R750W	D	8	24.7	β-sheet 34	A,E
G800W	D	5	20.8	β-sheet 38	B
G800R	D	8	20.8	β-sheet 38	B
L809P	D	4	31.7	β-sheet 39	-
G891R	E	5	40.7	loop	B,C
L892P	E	11	41.8	loop	B
R916C	E	16	29.8	β-sheet 39	A
R916H	E	13	29.8	β-sheet 39	A
L956R	E	8	24.8	β-turn	A
F1000S	E	11	32.4	β-sheet 47	-
(b) Activity defective					
H72L	A	17	6.5	active site	-
D74E	A	11	5.0	active site	-
Y99H	A	9	18.5	α-helix 3	-
D102N	A	15	21.0	α-helix 3	-
G153V	A	19	14.9	loop	-
D159N	A	10	9.7	loop	B,C
H200L	A	25	13.9	loop	-
H200N	A	53	13.9	loop	-
R229W	A	30	23.7	α-helix 8	C,D
P263L	A	9	15.4	loop	-
S318L	A	32	9.3	loop	-
P379L	B	19	18.1	loop	D
G420V	B	15	37.6	loop	E
G451C	C	30	12.9	loop	B,D
V457E	C	33	14.0	α-helix 17	A
T745R	D	18	25.5	β-sheet 33	-
R950P	E	16	41.2	β-sheet 47	-
(c) Not classified					
L518P	C	N.D.	34.5	β-sheet 15	D
G801D	D	N.D.	24.3	β-sheet 38	B
R916S	E	N.D.	29.8	β-sheet 45	A

*Relative activity (%) vs wild-type counterpart.[38] N.D.: no data; **Distance of the mutated residue to the catalytic Zn^{2+} atom; *** (-): no contact with another chain.

10

Impact of the missense mutations in hLAMAN stability

Almost half of the reported missense mutations affecting the folding of the protein (13 out of 23) are located in ordered regions, while most of the mutations causing decrease of the enzyme activity (10 out of 17) are located in loops or disordered regions. This observation can be linked to the fact that changes in α-helices or β-sheets (ordered regions) can deeply alter the folding outcome in hLAMAN. In order to support this hypothesis, we measured the impact of each of the 23 missense mutations affecting the folding in hLAMAN on the global stability of the enzyme (Figure 4 and Table S1). We calculated the change of the global free energy of the enzyme variant by replacing residue X by the deleterious mutation Y (as $\Delta\Delta G_{X\to Y}$, see Materials and Methods). In most of the folding mutations, the system increases its free energy significantly (median of 5.0 kcal mol^{-1}, with 48 % of them over $\Delta\Delta G_{X\to Y} > 10.0$ kcal mol^{-1}). This highlights the impact of these mutations on the secondary structure of hLAMAN. As an example, the changes on Pro and Gly residues has the larger impact on $\Delta\Delta G_{X\to Y}$ (e.g., G800 in chain D). In contrast to the former folding mutations, the values for the ones affecting the enzyme activity, but not the folding exhibit smaller $\Delta\Delta G_{X\to Y}$'s (median of 6.00 kcal mol^{-1}) and only a 39 % of them are above 10 kcal mol^{-1}. Only G153V T745R or G451C within this group do decrease the stability of the protein scaffold. This may indicate that in the former group of mutations, the lack of the activity is not due to a loss of the structure of the protein, but rather to a change of the dynamics and the network of interactions within hLAMAN. Those mutations located at the active site (H72L or D74E) have residual impact on the global stability of the protein scaffolds, with $\Delta\Delta G_{X\to Y}$ close to zero or even negative. Regarding the mutations that have not been classified, only G801D shows a value close to 10 kcal mol^{-1}.
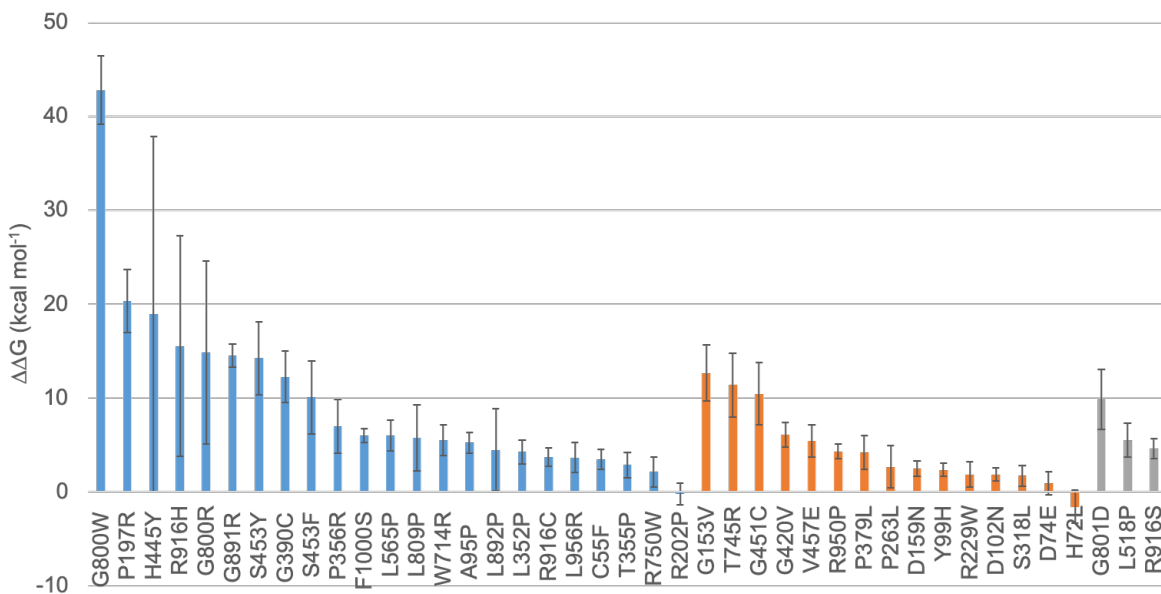
Figure 4: $\Delta\Delta G_{X\to Y}$ (kcal mol^{-1}, $\pm$ SD) for the missense mutations affecting the enzyme folding (blue bars), the enzyme activity (orange bars) or of unknown effect (grey bars) computed using MD ensembles and FoldX5.[52] Mutations affecting residue H200 present values above 200 kcal mol^{-1} and have been excluded.

Impact of the missense mutations on the dynamics of hLAMAN

We further analyzed the dynamics of hLAMAN in MD simulations. The core globular domain of hLAMAN shows low dynamical fluctuations along the trajectories, as can be seen in the root-mean square fluctuation (RMSF) values per residue (mean RMSF of 0.83 Å, Figure 5A). In contrast, the β/γ-loops exposed to the solvent (e.g., residues 539-543, 551-562, 583-593, 759-768, and 983-985) and the spliced regions between peptide chains (A-B: K345-G346; C-D, S600-W601; D-E, G882-A883) have a noticeable mobility. The computed RMSF values on MD trajectories show that positions R229 and P263 moves significantly along the MD trajectory (Figure 5B and Table S2). Whereas P263 is located in a loop, R229 is placed in the α-helix H9. However, since we analyzed the wild-type hLAMAN scaffold, the impact of the reported mutations on the dynamics of the protein cannot be anticipated. In addition to that, most of the *loci* for the described mutations have an RMSF value below the mean value (Figure 5B). Therefore, we decided to look for alternative descriptors that

12

allowed us to evaluate the impact of the dynamical effect of the mutations. Since we are using conformational ensembles of hLAMAN, it is possible to quantify the impact of the residue substitution coupled with the dynamics of the protein, independently of the mean value. The differential width of the former $\Delta\Delta G_{X \to Y}$ energy distribution for each of the explored mutations is related to the dynamics of the local environment (Table S1). As an example, the change G891R on β42-β43 loop of subunit E shows a broad $\Delta\Delta G_{X \to Y}$ (standard deviation of 4.38 kcal mol^{-1}), whereas the change of L352P in the β-sheet 9 of subunit B has a much smaller bandwidth (standard deviation of 1.24 kcal mol^{-1}, Figure 5B). Indeed, for some residues (e.g., C55F removing the disulfide on chain A), more than one maximum in $\Delta\Delta G_{X \to Y}$ is found, which illustrates different conformational states of hLAMAN in MD simulations. Compared to the RMSF values, the standard deviation in $\Delta\Delta G_{X \to Y}$ can be used as a measure of the impact of the missense mutation in the dynamics of the local environment using MD ensembles of the wild-type enzyme.
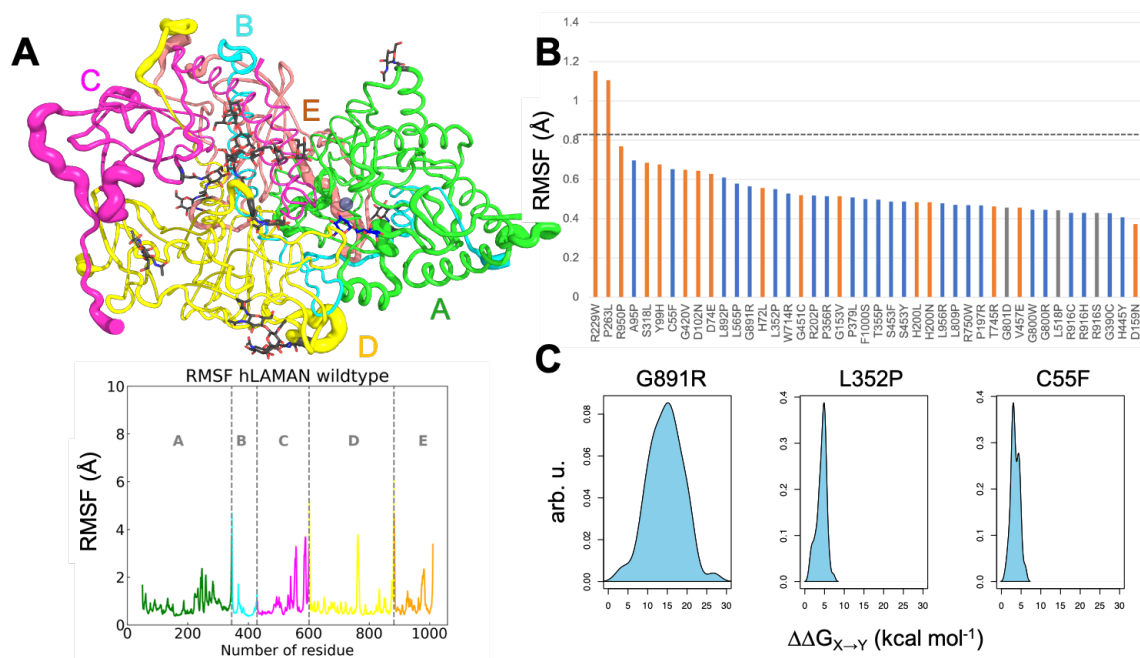

Figure 5: (A) Root-mean-square fluctuation (RMSF, Å) in MD simulations and dynamical regions on wild-type hLAMAN. Peptides A-E are shown. (B) RMSF values for the reported missense mutation positions. The grey dotted-line highlights the RMSF mean value in hLAMAN. (C) $\Delta\Delta G_{X \to Y}$ (kcal mol^{-1}) distribution of G891R, L352P, and C55F in MD simulation.

13

Coupling of the remote missense mutations with the active site

As we commented before, most of the reported mutations are not in the proximity of the active site of the enzyme. But how do they affect the enzyme activity if they are located remotely? In order to answer this questions, we interrogated our MD trajectory data to quantify the degree of 'cross-talk' between residues in the protein. We computed the mutual information (MI) between each of reported positions containing a mutation and ten residues at the active site[28] (Figure 6 and Table S2). Those mutations that are located at positions at the active site (H72 and D74) are not included in the analysis. MI refers to the dependency between two variables or elements, like residues in this case. We use as metrics the total pairwise MI between the ϕ and ψ dihedral angles of ten residues from the active site and these two dihedral angles from the rest of the residues of the protein scaffold. The larger the MI, the stronger the communication between residues. In general, we observe a strong coupling of the reported positions with three residues at the active site: D74, R220, and to a lesser extend, W77. In contrast, there is a weak coupling with Y660. Within the group of mutations affecting the activity of the enzyme, 54 % of them have a total dihedral MI value above the mean value (1.09). D159 is the one with the highest total dihedral MI value with the residues of the active site (mainly via D74 and R220). V457 and R229 are the following residues with the larger dihedral MI values within this group of mutations. In the case of mutations affecting the folding of the enzyme, only 13 % of the mutations show a total dihedral MI value above 1.0 (mean MI value per residue 0.66), with positions S453 and H445 as the ones with larger coupling to the active site. Very weak coupling is observed for residues P379, P263, R950, and G420 in the group of enzyme defective enzymes and P356, L352, T355, G800, C55, and F1000 in the missense mutations affecting the folding.
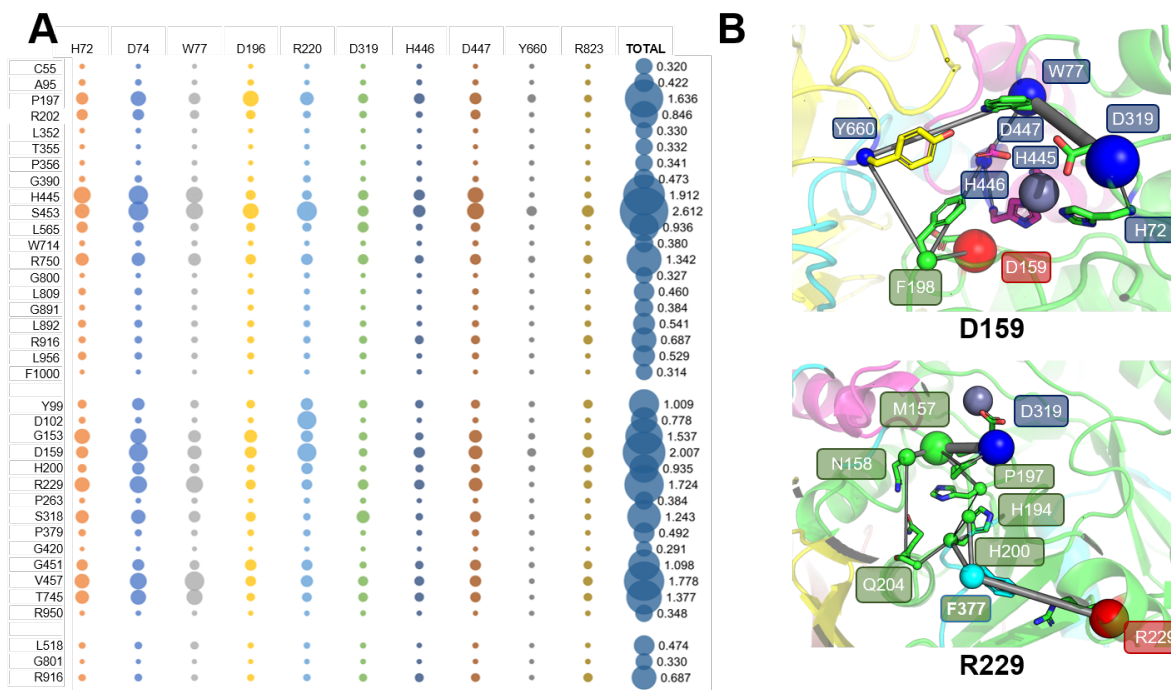
14

Figure 6: (A) Coupling of missense mutations and ten residues at the active site. No data are shown for H72 and D74 since they are included into the ten residues of the active site. (B) WISP trajectories obtained from the source (red spheres) D159 and R229 to the active site (blue spheres, sink). The connecting nodes are also as spheres of different size.

Further, we studied the pathways between the missense mutations and the residues at the active site using the weighted implementation of suboptimal path (WISP) analysis[27] (see Material and Methods). Since the folding mutations may prevent the quaternary structure of hLAMAN, we focused our analysis on the positions of the reported mutations affecting the enzyme activity. In Figure 6B are represented the WIPS trajectories for two of the strong coupled residues with the active site: D159 and R229. The side of the spheres denote the frequency of a node (residue) and the width of the cylinders the frequency of an edge (residue–residue interaction) in the generated path. Larger sizes indicate higher frequencies, what means an increased importance in the allosteric communication. The strongly coupled position D159, only needs four residues to reach the active site (77, 159, 198, and 445). In contrast, the residue R229 needs eight residues to reach it. On the way to the active site, other positions listed in Table 1 (e.g., H200) are found. The pathways for the rest of

15

positions of the missense mutations affecting the activity are plotted in Figures S3 and S4, summarize in Table S4, and the contributions of the node in Figure S5.

We further analyzed the coupling of the reported mutations with the same 10 residues from the active site but via bond-to-bond propensity descriptor (Figure 7A and Table S2).[18,53] Bond-to-bond propensity describes the effects of fluctuations of given bonds on any other bond in the protein. As before, the bond-to-bond scoring is defined between 0.0 and 1.0, but in this case 1.0 is the maximum value for each position. In contrast to the dihedral-dihedral MI analysis, with the bond-to-bond propensity, we can see the contribution due to the network of covalent bonds. Positions related to the misfolding of hLAMAN have in general high bond-to-bond propensity scoring values. The ones related to the loss of the activity of the enzyme tend to have smaller values. More of the half of the positions of the misfolding group shows a mean scoring value above 0.60. In contrast, only 4 positions out of 12 have a mean value ≥ 0.60 in the group of mutations affecting the enzyme activity. For the non-characterized missense mutations, the positions R916 and L518 show large scoring values (≥ 0.60). G801 presents a small value. We represented the total dihedral MI vs the bond-to-bond scoring mean values (Figure 7B). We normalized all the data by dividing by the maximum value within the series. This way, all data are distributed within {0,1}. Few positions (D159, R229, R750, and P197) present both high MI and bond-to-bond scoring values (right top quadrant). In the left top quadrant (low total dihedral MI value and large bond-to-bond scoring) are located most of the positions related to folding missense mutations, with only two mutations affecting the activity (P379 and G451), which are glycine/proline residue. Regarding the position of the missense mutations that are not classified yet, two of them (R916 and L518) are present in this quadrant, while G801 presents both low MI and bond-to-bond scoring. Altogether, with this projection MI vs bond-to-bond scoring, a tendency is observed for the missense mutations affecting the folding of the enzyme: low MI values and large bond-to-bond propensities.
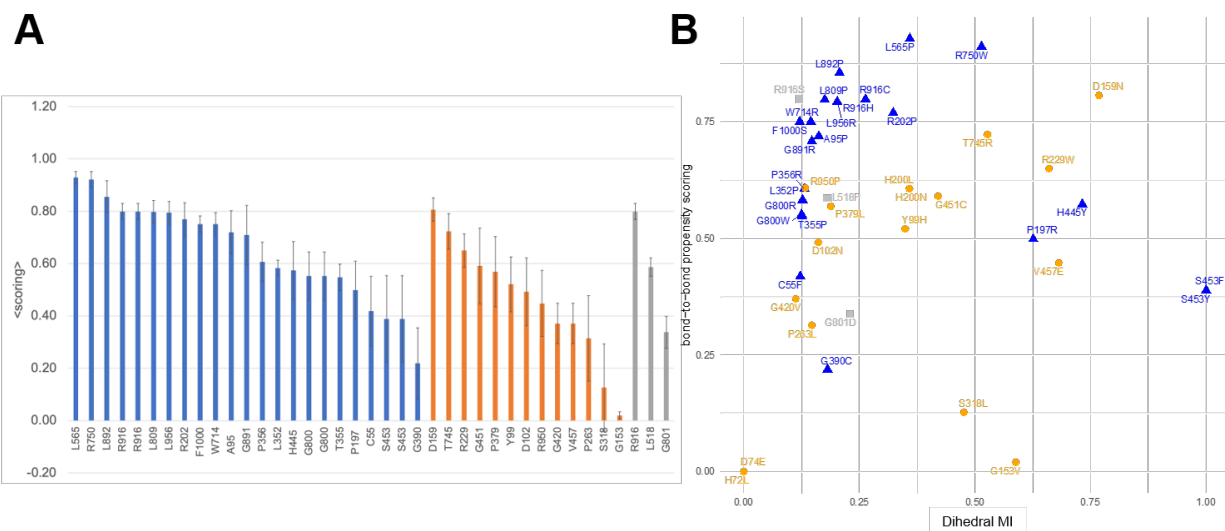
16

Figure 7: (A) Scoring for bond-to-bond propensities for the positions. (B) Total dihedral MI vs bond-to-bond propensity scoring for the reported missense mutations.

A new view of hLAMAN missense mutations

The analysis of MD-derived conformational ensembles can provide useful information about the molecular basis for each of the reported mutations. As discussed above, we have analyzed the impact of the reported mutation by computing from conformational ensembles the global stability of the enzyme ($\Delta\Delta G_{X \to Y}$ values), the cross-talking of each of the positions with the active site (total dihedral MI and bond-to-bond propensity scoring), as well as the dynamical component for each of the positions (width of the $\Delta\Delta G_{X \to Y}$). The different scoring of each of the residues in each of these four descriptors can unveil the molecular basis for causing malignancy. We merged the data of the four descriptors for all missense mutations and we plotted the data as radar chat (Table 2 and Figure S6). All values are defined for a range between 0 and 1 after normalizing by the maximum value of the two series. Looking at Table 2, the relative impact of one mutation in each of the descriptors can be identified. In the graphical representation, the radar chart of the mutations can provide a quick overview of the molecular reasons of malignancy for a given variant. As an example, in Figure 8 are shown the three unknown variants, which have been not yet classified: L518P, G801D and R916D.

17

The median values for the folding and defective activity variants are also shown. Based on their projections, the G801D enzyme is more similar to those mutation causing unfolding, whereas L518P and R916S mimics the radar chart of the group of mutations causing the loss of enzyme activity, with a strong coupling with the active site via the backbone bones. Of the three mutations, G801D is the one with the larger impact on the stability of hLAMAN and L518D the one with a larger dihedral coupling with the active site (total dihedral MI values).
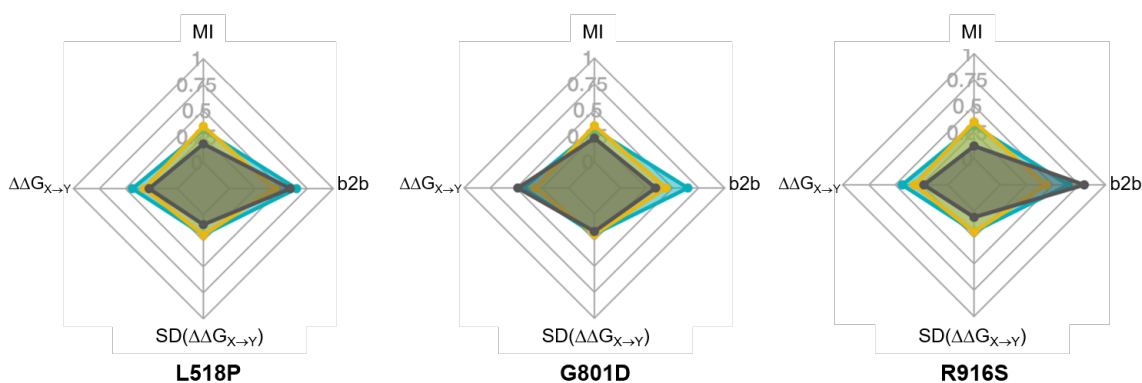

Figure 8: Radar chart of variants L518P, G801D and R916D (grey surface) for the descriptors total dihedral MI, bond-to-bond scoring (b2b), $\Delta\Delta G_{X \to Y}$, and the standard distribution of the latter. All values are normalized and defined between 0 and 1. The median values of the defective activity (orange surface) and folding (blue surface) missense groups are also shown.

Table 2: A new analysis of the reported missense mutations in hLAMAN analyzing MD ensembles.

Mutant	Chain	Total dihedral MI	$\Delta\Delta G_{X \to Y}$	SD($\Delta\Delta G_{X \to Y}$)	bond-to-bond propensity
(a) Protein folding					
C55F	A	0.122	0.172	0.056	0.419
A95P	A	0.162	0.257	0.058	0.720
P197R	A	0.626	1.000*	0.178	0.499
R202P	A	0.323	0.011	0.059	0.770
L352P	A	0.127	0.209	0.066	0.583
T355P	B	0.126	0.141	0.071	0.548
P356R	B	0.131	0.343	0.152	0.607
G390C	B	0.181	0.603	0.145	0.219
H445Y	C	0.732	0.929	1.000	0.574
S453Y	C	1.000	0.495	0.205	0.389
S453F	C	1.000	0.495	0.205	0.389
L565P	C	0.359	0.294	0.087	0.929
W714R	D	0.145	0.270	0.087	0.751
R750W	D	0.514	0.106	0.085	0.912
G800W	D	0.125	1.000*	0.192	0.552
G800R	D	0.125	1.000*	0.515	0.552
L809P	D	0.175	0.285	0.186	0.798
G891R	E	0.147	0.714	0.066	0.709
L892P	E	0.207	0.220	0.232	0.856
R916C	E	0.263	0.182	0.052	0.799
R916H	E	0.263	0.764	0.620	0.799
L956R	E	0.202	0.182	0.085	0.794
F1000S	E	0.121	0.295	0.040	0.751
(b) Activity defective					
H72L	A	0.000**	0.077	0.094	0.000**
D74E	A	0.000**	0.048	0.065	0.000**
Y99H	A	0.349	0.116	0.039	0.521
D102N	A	0.161	0.091	0.038	0.492
G153V	A	0.588	0.623	0.157	0.020
D159N	A	0.768	0.122	0.043	0.807
H200L	A	0.358	1.000*	1.000*	0.607
H200N	A	0.358	1.000*	1.000*	0.607
R229W	A	0.660	0.092	0.072	0.650
P263L	A	0.147	0.132	0.120	0.314
S318L	A	0.476	0.085	0.057	0.127
P379L	B	0.188	0.208	0.096	0.569
G420V	B	0.112	0.301	0.069	0.371
G451C	C	0.420	0.514	0.175	0.591
V457E	C	0.681	0.266	0.091	0.448
T745R	D	0.527	0.559	0.179	0.723
R950P	E	0.133	0.213	0.043	0.609
(c) Not classified					
L518P	C	0.180	0.272	0.096	0.587
G801D	D	0.230	0.485	0.168	0.338
R916S	E	0.120	0.227	0.058	0.799

*Data above 20.35 kcal mol^{-1}, the normalizing value.**Residue defined as part of the active site.

Conclusions

We have analyzed MD-derived conformational ensembles to understand how remote missense mutations affect the loss of enzyme activity and the folding of hLAMAN. The relevant biological quaternary structure of hLAMAN is a homodimer of the polypeptide A-E. By analyzing the key residues for the integrity of the dimer interface between units A and A' in hLAMAN, we could assign an important role for the reported missense mutations like Y99H and D102N. They are located at the hLAMAN homodimer interface and they participate in a well-defined amino acid network. Thus, they mutation may significant disturb this network of interactions between both units. Additionally, the interaction between units A and B in the polypeptide A-E of hLAMAN is the stronger one amongst all subunits. The computation of the impact of the missense mutation on the global stability of the enzyme scaffold as $\Delta\Delta G_{X \to Y}$ shows that the mutations affecting the folding of hLAMAN present larger values than the group affecting the activity. In addition to that, the fluctuations of $\Delta\Delta G_{X \to Y}$ along the MD ensembles indicates the relevance of the mutation on the interactions within the local environment and its dynamical effect. The use of RMSF per residue along the MD simulation does not provide such information. We have also analyzed the coupling of the reported mutations with ten residues located at the active site via two metrics: the total dihedral MI and the bond-to-bond propensity scoring. Using them, we see that the two groups of missense mutations show different trends for each of the descriptors, what could be use as a classification metrics. Altogether, we have reported here the contribution of each of the pathogenic missense mutations into the stability of hLAMAN, its dynamics, and the network of interactions with the active site. Such analysis and understanding at the atomic level can help on deciphering the role of future missense mutations in LSDs as well as to better adapt and designing novel molecules (i.e., pharmacological chaperones or ERT), and potentially, to contribute to the foundation for precision medicine in LSDs.

Materials and Methods

hLAMAN expression and purification

The gene encoding for hLAMAN (MAN2B1) was acquired inserted into the pET28b+ vector (GenScript). After amplification using the primers 5'-TGCGCGCACGTGGATGGGCAGCAGC CATCATCATCATCATCACAG-3' and 3'-TGCGCGGCGGCCGCTTAGCCATCCACTTCC-5' (GenScript) the gen of interest was cloned into the *K. phaffii* expression vector pPICZalpha (Thermo Fischer Scientific) using the restriction enzymes EcoRI and NotI. Following transformation (100 ng of DNA plasmid) and amplification in *E. coli* NEB10-beta cells, the final pPICZalpha was isolated and 13 µg of the DNA plasmid were linearized using SacI as restriction enzyme. *K. phaffii* X33 was transformed with the linearized plasmid. Positive colonies were selected on YPDS agar plates containing 100 μg mL^{-1} of the antibiotic Zeocin^{TM}. To test the efficacy of the hLAMAN expression, a high-throughput expression screening was performed by cultivation of different colonies in 96-deep well plates followed by dot blotting. The 96-deep well plates contained a minimal glucose medium in 0.4 M potassium phosphate buffer pH 6.0 (1.34 % yeast nitrogen base, 1 % D-glucose, 500x biotin) were inoculated with 80 different colonies and were grown in the presence of different concentrations of Zeocin^{TM} (100, 250, 500, and 1000 µg mL^{-1}). hLAMAN expression was induced by addition of methanol (5 % methanol in glucose-deprived minimal medium) every 24 h. After 96 h cells were harvested by centrifugation and the protein fraction from the supernatant were subjected to dot blotting on a nitrocellulose membrane. After blocking the membranes in 5 % (w/v) non-fat dry milk in 1x Tris-buffered saline, 0.1 % Tween$^{(R)}$ 20 Detergent (TBST) media at pH 7.4, the primary antibody anti-HIS IgG (diluted in 5 % (w/v) non-fat dry milk in TBST) was used. After 2 h of incubation, the membranes were washed four times with 1x TBST and then, incubated with HRP-conjugated goat-anti-rabbit IgG (diluted in TBST containing non-fat dry milk (5 %, (w/v)) for 2 h. After another four washing steps with 1x TBST, the immunoreactive bands were visualized by chemiluminescence detection using the Super Signal

West Femto Maximum Sensitivity substrate (Thermo Fischer Scientific) and a Chemidoc detection system (Biorad). The colony showing the highest expression level was selected for a scale-up expression. 25 mL of BMGY media (1 % yeast extract, 2 % peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34 % YNB, 500x biotin, 1 % glycerol) and incubated overnight at 28 øC with a mixing of 180 rpm for growth. Upon an optical density at 600 nm of 6.0, the cells were harvested and resuspended in 200 mL of BMMY media (1 % yeast extract, 2 % peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34 % YNB, 500x biotin, 0.5 % methanol) to induce expression. To maintain expression, 100 % methanol was added every 24 hours to a final concentration of 0.5% at 28 °C. Aliquots were collected at various time points, centrifuged, snap-frozen, and stored for expression analysis of the pellet and the supernatant using SDS-PAGE and Western blot. The anti-His antibodies (Abcam, ab9108) were used for blotting.

Modeling of the quaternary structure of hLAMAN

The initial Cartesian coordinates of the polypeptide of hLAMAN (UniProtKB - O00754) were taken from the AlphaFold Protein Structure Database (AF-O00754-F1).[15] The signal peptide (1-49 amino acids) was removed, and the loops were manually cleaved between amino acids 345-346 (A-B), 429-430 (B-C), 601-602 (C-D), and 882-883 (D-E) (Figure S1). The reported N-glycosylation points at positions Asn137, Asn367, Asn497, Asn645, Asn692, Asn766, and Asn930 as well as the catalytic zinc atom, were placed into hLAMAN by structural superimposition with bLAMAN (PDB id. 1O7D,[30] UniProt code Q29451) or by manual docking. The experimentally reported disulfide bonds for the pairs 55-358, 268-273, 412-472, and 493-5019, were already present in the AlphaFold model and were included when building up the topology of the system. The protonation states of the titratable residues on the protein were calculated via the H++ web server[54] assuming a pH value of 4.8 (close to the lysosomal one). We were careful with the catalytic residue Asp319 (general acid for the first step of the reaction), which was manually protonated to accomplish the proposed mech-

22

anism of action.[55] As hLAMAN substrate, we included the β-D-mannopyranose-(1-6)-α-D-mannopyranose (BMAM), derived from the crystal structure of a GH125 1,6-α-mannosidase variant (PDB id. 5M7I[56]). BMAM was manually docked into the active site of hLAMAN guided by structural superimposition with other glycosydades.

Classical molecular dynamics (MD) simulations

MD simulations were carried out using the suite of programs Amber20.[57] Protein residues and solvated ions were treated with the AMBER ff19SB force field and GLYCAM06 for the N-glycosylated residues and the substrate BMAM.[58,59] Parameters for all atoms involved in the coordination sphere of the catalytic Zn^{2+} were derived by means of the Metal Center Parameter Builder (MCPB) approach.[60] The necessary quantum geometry optimization of the Zn^{2+}-coordination sphere in the gas phase was carried out with Gaussian16[61] at the B3LYP/6-31G* level of theory.[62–65] The hLAMAN monomer (A-E) as well as the hLAMAN dimer (A-E)$_2$ were simulated. All systems were first energetically minimized to avoid steric clashes, and then placed in the center of a cubic box filled with 28,000-29,000 TIP3P water molecules.[66] Then, the solvated systems were minimized in three consecutive steps (first all protons, then solvent, and finally, all system) and heated up in 50 ps from 100 to 300 K using an NVT ensemble with the Langevin thermostat (gamma friction coefficient of 1.0 ps $^{-1}$). Care was taken to constrain the solute during the heating step by imposition of a harmonic force on each atom of the solute of 40 kcal mol^{-1} Å^{-2}. Afterward, these harmonic constraints were gradually reduced up to a value of 10 kcal mol^{-1} Å^{-2} in 4 simulation stages (NVT, 300 K). The systems were switched to constant pressure (NPT scheme, Berendsen barostat 300 K), and the imposed constraints from the heating were totally removed. Each of the systems was further simulated for a total time of 1.5 μs in three independent MD simulations of 0.5 μs. Atom-pair distance cutoffs were applied at 10.0 Å to compute the van der Waals interactions and long-range electrostatics by means of the Particle-Mesh Ewald (PME) method. SHAKE algorithm was applied to restrain the hydrogen atoms on water molecules.[67] MD trajectory

23

analysis was carried out using the *cpptraj*[68] module from Amber20[69] for monitoring the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF), amongst other parameters.

Mutual information (MI) calculation

We considered the communication of all residues in hLAMAN with the active site via the torsional angles. We computed the mutual information (MI) pairwise values using the Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems (PARENT).[28,70,71] As input data set, we analyzed a merged MD ensemble from three independent MD simulations of hLAMAN monomer in explicit aqueous solution (3 x 0.5 μs, 12.000 snapshots; the first 10 ns of the MD trajectories were not taken into consideration). All glycosylated residues where changed to Asn for the analysis afterwards. We analyzed the dihedral-dihedral MI values, which can present values from 0 (no coupling) to 1 (maximum coupling). These dihedral angles span the ones of the backbone (ϕ, ψ). In particular, we analyzed the torsional MI values between each residue i in the protein with 10 selected residues placed at the active site, as done in the bond-to-bond propensity analysis (H72, D74, W77, D196, R220, D319, H446, D447, Y660, and R823). A 962 x 962-contact matrix was generated with the MI values (see Figure S1). The mean value across the matrix is 0.05, with auto-coupling between amino acids excluded from the analysis.

Bond-to-bond propensity analysis

The bond-to-bond propensities for wild-type hLAMAN were computed using the server ProteinLens.[19,72] The connectivity of the covalent bonds between the source (in our case the active site) and every other bond in the protein structure were measured with the bond-to-bond propensity analysis. This is done by assessing the propagation of a perturbation originating at the source site through an edge formulation of random walks. On they way, the

24

propagation of these 'fluctuations' within the model depends on the strength of interactions between atoms (e.g., hydrogen bonds, hydrophobic interactions). We utilized 9 randomly-selected snapshots from the MD simulation of the wild-type hLAMAN, excluding the first 100 ns of each of the single trajectories. As the primary source site for fluctuation propagation, we defined 10 residues located at or in close proximity to the active site: H72, D74, W77, D196, R220, D319, H446, D447, Y660, and R823. As outcome, the statistical quantile score is obtained, which ranges between 0 (no coupling) to 1 (full coupling).

Weighted implementation of suboptimal path (WISP) analysis

WISP transforms the protein into a network and identifies how distal regions communicate allosterically. By designating a source and sink residue, the analysis reveals the shortest pathways for communication between them, pinpointing key residues and residue interactions. A correlation cutoff of 0.25 (absolute value) is used to determine the 500 shortest paths between source and sink, enabling the analysis of node and edge degenerancies alongside path lengths. We used the Key Interaction Finder (KIF) software.[73] KIF methodology is based on the calculation of a correlation matrix between each of the structural features in the protein, which is transformed into a per-residue matrix by identifying the largest correlation between any pair of residues. This step increases the dimensions from the number of features squared to the number of residues squared, with the diagonal set to 1. Next, the contact maps/matrices are created using the MDTraj library[74] to measure the heavy-atom distances between residue pairs in a given structure. A 6 Å cutoff defines whether two residues are in contact, resulting in a binary matrix containing 1 (for contacts) and 0 (no contacts within the cutoff). These correlation and contact matrices are inputted into the R package Bio3D,[75] facilitating the WISP analysis.

25

Binding energy and $\Delta\Delta G_{X \to Y}$ calculations

A total of 60 frames from the merged MD ensemble were printed and analyzed with the force field FoldX5 v2024[52] to compute the binding energy between hLAMAN subunits (*AnalyseComplex* module) and to calculate the impact on the energy of the system after amino-acid mutation as $\Delta\Delta G_{X \to Y}$ (*PositionScan* module). To calculate the individual energetic contribution of each of the resides at the A-A' interface, we made use of the software MM-ISMSA.[76]

Conflict of interests statement

There are not conflicts of interest.

Data availability

Additional data, like the 3D structural model of hLAMAN, the topology file and the initial coordinates of hLAMAN embedded in a box of water molecules are are publicly available (DOI: 10.5281/zenodo.11638802).

Supporting information

Supplementary data. **Figures**: S1. Scheme for the topology of hLAMAN. Image modified from PDBsum. S2. PARENT analysis contact map in hLAMAN. S3. KIF-WISP calculations from the first-sphere influence on hLAMAN. S4. KIF-WISP calculations from the second-sphere influence on hLAMAN. S5. KIF-WISP node degeneracy for each path derived from the hLAMAN trajectory. S6. Radar charts for data shown in Table 3. **Tables**: S1. Mean $\Delta\Delta G_{X \to Y}$ value (kcal mol^{-1}) for the reported missense mutations computed on the MD ensemble of wild-type hLAMAN. S2. RMSF (Å) for the reported missense mutations computed along the MD simulation of the hLAMAN monomer. S3. Computed bond-to-bond propensity scoring and mutual information (MI) for the reported missense mutations

26

in hLAMAN. S4. Resides identified on the path from the reported position to the active site using the KIF-WISP analysis.

Author contributions

Špela Mandl, Gibu George, Sereina Annik Herzog: investigation; writing – original draft; writing – review and editing. **Bruno Di Geronimo**: validation; investigation; writing – original draft; writing – review and editing; visualization. **Santiago Alonso-Gil**: Conceptualization; methodology; investigation; writing – review and editing. **Ulrika Ferstl, Christoph Nusshold**: investigation; writing – review and editing. **Bojan Žagrović**: writing – review and editing. **Pedro A. Sánchez-Murcia**: Conceptualization; methodology; validation; investigation; writing – original draft; writing – review and editing; visualization.

Acknowledgements

References

(1) Platt, F. M.; d'Azzo, A.; Davidson, B. L.; Neufeld, E. F.; Tifft, C. J. Lysosomal storage diseases. *Nature Reviews Disease Primers* **2018**, *4*, 27.

(2) Filocamo, M.; Morrone, A. Lysosomal storage disorders: Molecular basis and laboratory testing. *Human Genomics* **2011**, *5*, 156.

(3) Moro, E. Lysosomal Storage Disorders: Molecular Basis and Therapeutic Approaches. *Biomolecules* **2021**, *11*, 964.

(4) Futerman, A. H.; van Meer, G. The cell biology of lysosomal storage disorders. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 554–565.

(5) Özlem Tastan Bishop; Musyoka, T. M.; Barozi, V. Allostery and Missense Mutations as Intermittently Linked Promising Aspects of Modern Computational Drug Discovery. *Journal of Molecular Biology* **2022**, *434*, 167610, Allostery: From Mechanisms to Therapies.

(6) Olson, L. J.; Misra, S. K.; Ishihara, M.; Battaile, K. P.; Grant, O. C.; Sood, A.; Woods, R. J.; Kim, J.-J. P.; Tiemeyer, M.; Ren, G.; Sharp, J. S.; Dahms, N. M. Allosteric regulation of lysosomal enzyme recognition by the cation-independent mannose 6-phosphate receptor. *Communications Biology* **2020**, *3*, 498.

(7) Lisi, G. P.; Loria, J. P. Allostery in enzyme catalysis. *Current Opinion in Structural Biology* **2017**, *47*, 123–130.

(8) Stefl, S.; Nishi, H.; Petukh, M.; Panchenko, A. R.; Alexov, E. Molecular Mechanisms of Disease-Causing Missense Mutations. *Journal of Molecular Biology* **2013**, *425*, 3919–3936, Understanding Molecular Effects of Naturally Occurring Genetic Differences.

(9) Wang, S.; Kaufman, R. J. The impact of the unfolded protein response on human disease. *Journal of Cell Biology* **2012**, *197*, 857–867.

(10) Ron, I.; Horowitz, M. ER retention and degradation as the molecular basis underlying Gaucher disease heterogeneity. *Human Molecular Genetics* **2005**, *14*, 2387–2398.

(11) Gerasimavicius, L.; Liu, X.; Marsh, J. A. Identification of pathogenic missense mutations using protein stability predictors. *Scientific Reports* **2020**, *10*, 15387.

(12) Gerasimavicius, L.; Livesey, B. J.; Marsh, J. A. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nature Communications* **2022**, *13*, 3895.

(13) Lim, J. H.; Kang, H. M.; Jung, H.-R.; Kim, D.-S.; Noh, K. H.; Chang, T. K.; Kim, B. J.; Sung, D. H.; Cho, H.-S.; Chung, K.-S.; Kim, N.-S.; Jung, C.-R. Missense mutation of SPAST protein (I344K) results in loss of ATPase activity and prolonged the half-life, implicated in autosomal dominant hereditary spastic paraplegia. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2018**, *1864*, 3221–3233.

(14) D'Amico, R. N.; Murray, A. M.; Boehr, D. D. Driving Protein Conformational Cycles in Physiology and Disease: Frustrated Amino Acid Interaction Networks Define Dynamic Energy Landscapes. *BioEssays* **2020**, *42*, 2000092.

(15) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; others Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(16) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130.

(17) Alamdari, S.; Thakkar, N.; van den Berg, R.; Lu, A. X.; Fusi, N.; Amini, A. P.; Yang, K. K. Protein generation with evolutionary diffusion: sequence is all you need. *bioRxiv* **2023**,

(18) Amor, B.; Yaliraki, S. N.; Woscholski, R.; Barahona, M. Uncovering allosteric pathways

in caspase-1 using Markov transient analysis and multiscale community detection. *Mol. BioSyst.* **2014**, *10*, 2247–2258.

(19) Amor, B. R. C.; Schaub, M. T.; Yaliraki, S. N.; Barahona, M. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature Communications* **2016**, *7*, 12477.

(20) Mersmann, S.; Strömich, L.; Song, F. J.; Wu, N.; Vianello, F.; Barahona, M.; Yaliraki, S. ProteinLens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules. *Nucleic Acids Research* **2021**, *49*, W551–W558.

(21) Yabukarski, F.; Doukov, T.; Pinney, M. M.; Biel, J. T.; Fraser, J. S.; Herschlag, D. Ensemble-function relationships to dissect mechanisms of enzyme catalysis. *Science Advances* **2022**, *8*, eabn7738.

(22) Hammes, G. G. Multiple Conformational Changes in Enzyme Catalysis. *Biochemistry* **2002**, *41*, 8221–8228.

(23) Yon, J.; Perahia, D.; Ghélis, C. Conformational dynamics and enzyme activity. *Biochimie* **1998**, *80*, 33–42.

(24) Childers, M. C.; Daggett, V. Validating Molecular Dynamics Simulations against Experimental Observables in Light of Underlying Conformational Ensembles. *The Journal of Physical Chemistry B* **2018**, *122*, 6673–6689.

(25) Platero-Rochart, D.; Sánchez-Murcia, P. A. In *Biocatalysis in Asymmetric Synthesis*; Gonzalo, G. D., Alcántara, A. R., Eds.; Foundations and Frontiers in Enzymology; Academic Press, 2024; pp 65–99.

(26) Romero-Rivera, A.; Garcia-Borrás, M.; Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catalysis* **2017**, *7*, 8524–8532.

(27) Crean, R. M.; Slusky, J. S. G.; Kasson, P. M.; Kamerlin, S. C. L. KIF—Key Interactions Finder: A program to identify the key molecular interactions that regulate protein conformational changes. *The Journal of Chemical Physics* **2023**, *158*, 144114.

(28) Fleck, M.; Polyansky, A. A.; Zagrovic, B. PARENT: A Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems. *Journal of Chemical Theory and Computation* **2016**, *12*, 2055–2065.

(29) Park, C.; Meng, L.; Stanton, L. H.; Collins, R. E.; Mast, S. W.; Yi, X.; Strachan, H.; Moremen, K. W. Characterization of a Human Core-specific Lysosomal 1,6-Mannosidase Involved in N-Glycan Catabolism. *Journal of Biological Chemistry* **2005**, *280*, 37204–37216.

(30) Heikinheimo, P.; Helland, R.; Leiros, H.-K. S.; Leiros, I.; Karlsen, S.; Evjen, G.; Ravelli, R.; Schoehn, G.; Ruigrok, R.; Tollersrud, O.-K.; McSweeney, S.; Hough, E. The Structure of Bovine Lysosomal -Mannosidase Suggests a Novel Mechanism for Low-pH Activation. *Journal of Molecular Biology* **2003**, *327*, 631–644.

(31) Henrissat, B.; Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **1997**, *7*, 637–644.

(32) Rovira, C.; Males, A.; Davies, G. J.; Williams, S. J. Mannosidase mechanism: at the intersection of conformation and catalysis. *Current Opinion in Structural Biology* **2020**, *62*, 79–92.

(33) Zielonka, M.; Garbade, S. F.; Kölker, S.; Hoffmann, G. F.; Ries, M. Ultra-orphan lysosomal storage diseases: A cross-sectional quantitative analysis of the natural history of alpha-mannosidosis. *Journal of Inherited Metabolic Disease* **2019**, *42*, 975–983.

(34) Kuokkanen, E.; Stensland, H. M. F. R.; Smith, W.; Buvang, E. K.; Nguyen, L. V.; Øivind Nilssen; Heikinheimo, P. Molecular and cellular characterization of novel -mannosidosis mutations. *Human Molecular Genetics* **2011**, *20*, 2651–2661.

(35) Malm, D.; Øivind Nilssen Alpha-mannosidosis. *Orphanet Journal of Rare Diseases* **2008**, *3*, 21.

(36) Gotoda, Y.; Wakamatsu, N.; Kawai, H.; Nishida, Y.; Matsumoto, T. Missense and Nonsense Mutations in the Lysosomal -Mannosidase Gene (MANB) in Severe and Mild Forms of -Mannosidosis. *The American Journal of Human Genetics* **1998**, *63*, 1015–1024.

(37) McCorvie, T. J.; Yue, W. W. *Structure-guided discovery of pharmacological chaperones targeting protein conformational and misfolding diseases*; Elsevier, 2020; pp 281–308.

(38) Stensland, H. M. F. R.; Frantzen, G.; Kuokkanen, E.; Buvang, E. K.; Klenow, H. B.; Heikinheimo, P.; Malm, D.; Øivind Nilssen amamutdb.no: A Relational Database for MAN2B1 Allelic Variants that Compiles Genotypes, Clinical Phenotypes, and Biochemical and Structural Data of Mutant MAN2B1 in -Mannosidosis. *Human Mutation* **2015**, *36*, 581–586.

(39) Sbaragli, M.; Bibi, L.; Pittis, M. G.; Balducci, C.; Heikinheimo, P.; Ricci, R.; Antuzzi, D.; Parini, R.; Spaccini, L.; Bembi, B.; Beccari, T. Identification and characterization of five novel MAN2B1 mutations in Italian patients with alpha-mannosidosis. *Human Mutation* **2005**, *25*, 320–320.

(40) Riise Stensland, H. M. F.; Klenow, H. B.; Nguyen, L. V.; Hansen, G. M.; Malm, D.; Nilssen, Identification of 83 novel alpha-mannosidosis-associated sequence variants: Functional analysis of MAN2B1 missense mutations. *Human Mutation* **2012**, *33*, 511–520.

(41) Berg, T.; Riise, H. M. F.; Hansen, G. M.; Malm, D.; Tranebjærg, L.; Tollersrud, O. K.; Øivind Nilssen Spectrum of Mutations in -Mannosidosis. *The American Journal of Human Genetics* **1999**, *64*, 77–88.

(42) Landrum, M. J.; Lee, J. M.; Riley, G. R.; Jang, W.; Rubinstein, W. S.; Church, D. M.; Maglott, D. R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* **2013**, *42*, D980–D985.

(43) Liao, Y.-F.; Lal, A.; Moremen, K. W. Cloning, Expression, Purification, and Characterization of the Human Broad Specificity Lysosomal Acid -Mannosidase. *Journal of Biological Chemistry* **1996**, *271*, 28348–28358.

(44) Foster, J. M.; Yudkin, B.; Lockyer, A. E.; Roberts, D. B. Cloning and sequence analysis of GmII, a Drosophila melanogaster homologue of the cDNA encoding murine Golgi -mannosidase II. *Gene* **1995**, *154*, 183–186.

(45) van den Elsen, J. M. Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *The EMBO Journal* **2001**, *20*, 3008–3017.

(46) Vallée, F.; Karaveg, K.; Herscovics, A.; Moremen, K. W.; Howell, P. L. Structural Basis for Catalysis and Inhibition of N-Glycan Processing Class I 1,2-Mannosidases. *Journal of Biological Chemistry* **2000**, *275*, 41287–41298.

(47) Gil-Martínez, J.; Bernardo-Seisdedos, G.; Mato, J. M.; Millet, O. The use of pharmacological chaperones in rare diseases caused by reduced protein stability. *Proteomics* **2022**, *22*, e2200222.

(48) Berg, T.; King, B.; Meikle, P. J.; Øivind Nilssen; Tollersrud, O. K.; Hopwood, J. J. Purification and Characterization of Recombinant Human Lysosomal alpha-Mannosidase. *Molecular Genetics and Metabolism* **2001**, *73*, 18–29.

(49) Hansen, G.; Berg, T.; Riise-Stensland, M. F., Hilde; Heikinheimo, P.; Klenow, H.; Evjen, G.; Nilsen, ; Tollersrud, O. K. Intracellular transport of human lysosomal -mannosidase and -mannosidosis-related mutants. *Biochemical Journal* **2004**, *381*, 537–546.

(50) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **2007**, *372*, 774–797.

(51) Suits, M. D. L.; Zhu, Y.; Taylor, E. J.; Walton, J.; Zechel, D. L.; Gilbert, H. J.; Davies, G. J. Structure and Kinetic Investigation of Streptococcus pyogenes Family GH38 -Mannosidase. *PLOS ONE* **2010**, *5*, 1–11.

(52) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic acids research* **2005**, *33*, W382–W388.

(53) Hodges, M.; Barahona, M.; Yaliraki, S. N. Allostery and cooperativity in multimeric proteins: bond-to-bond propensities in ATCase. *Scientific Reports* **2018**, *8*, 11079.

(54) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic acids research* **2005**, *33*, W368–W371.

(55) Petersen, L.; Ardévol, A.; Rovira, C.; Reilly, P. J. Molecular mechanism of the glycosylation step catalyzed by Golgi -mannosidase II: A QM/MM metadynamics investigation. *Journal of the American Chemical Society* **2010**, *132*, 8291–8300.

(56) Alonso-Gil, S.; Males, A.; Fernandes, P. Z.; Williams, S. J.; Davies, G. J.; Rovira, C. Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 -Mannosidases. *Journal of the American Chemical Society* **2017**, *139*, 1085–1088.

(57) Case, D. et al. AMBER 2022: University of California. *San Francisco* **2022**,

(58) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16*, 528–552.

(59) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **2008**, *29*, 622–655.

(60) Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling* **2016**, *56*, 599–604.

(61) Frisch, M. J. et al. Gaussian 16. 2016.

(62) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

(63) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785–789.

(64) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **1980**, *58*, 1200–1211.

(65) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.

(66) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105*, 9954–9960.

(67) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952–962.

(68) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.

(69) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; others *Amber 2021*; University of California, San Francisco, 2021.

(70) Fleck, M.; Polyansky, A. A.; Zagrovic, B. Self-Consistent Framework Connecting Experimental Proxies of Protein Dynamics with Configurational Entropy. *Journal of Chemical Theory and Computation* **2018**, *14*, 3796–3810.

(71) Fleck, M.; Zagrovic, B. Configurational Entropy Components and Their Contribution to Biomolecular Complex Formation. *Journal of Chemical Theory and Computation* **2019**, *15*, 3844–3853.

(72) Mersmann, S. F.; Strömich, L.; Song, F. J.; Wu, N.; Vianello, F.; Barahona, M.; Yaliraki, S. N. ProteinLens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules. *Nucleic Acids Research* **2021**, *49*, 551–558.

(73) Crean, R. M.; Slusky, J. S. G.; Kasson, P. M.; Kamerlin, S. C. L. KIF—Key Interactions Finder: A program to identify the key molecular interactions that regulate protein conformational changes. *The Journal of Chemical Physics* **2023**, *158*, 144114.

(74) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(75) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3D: An R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695–2696.

(76) Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, ; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: An Ultrafast and Accu-

rate Scoring Function for Protein-Protein Docking. *Journal of Chemical Theory and Computation* **2012**, *8*, 3395–3408.

TOC Graphic