J Mol Model (2007) 13:1133–1139
DOI 10.1007/s00894-007-0232-5

ORIGINAL PAPER

The studies on substrate, product and inhibitor binding to a wild-type and neuronopathic form of human acid-β-glucosidase

Igor Z. Zubrzycki · Agnieszka Borcz · Magdalena Wiacek · Wojciech Hagner

Received: 26 March 2007 / Accepted: 12 July 2007 / Published online: 23 August 2007

Abstract Gaucher disease is a lysosomal storage disorder caused by deficiency of human acid β-glucosidase. Recent x-ray structural elucidation of the enzyme alone and in the presence of its inhibitor was done, which provided an excellent template for further studies on the binding of substrate, product and inhibitor. To draw correlations between the clinical manifestation of the disease driven by point mutations, L444P and L444R, and the placement and function of putative S-binding sites, the presented theoretical studies were undertaken, which comprised of molecular dynamics and molecular docking methods. The obtained results indicate the D443 and D445 residues as extremely important for physiological functionality of an enzyme. They also show, although indirectly, that binding of the substrate is influenced by an interplay of E235 and E334 residues, constituting putative substrate binding site, and the region flanked by D435 and D445 residues.

Keywords Human acid-β-glucosidase · Molecular dynamics · Molecular docking · S-binding place · Mutations

I. Z. Zubrzycki (✉) · A. Borcz · M. Wiacek
Department of Biotechnology, University of Rzeszow,
ul Sokolowska 26,
36-100 Kolbuszowa, Poland
e-mail: izubrzyc@univ.rzeszow.pl

W. Hagner
Department of Rehabilitation, Collegium Medicum,
Nicolas Copernicus University,
ul. M. Sk odowskiej-Curie 9,
85-094 Bydgoszcz, Poland

Introduction

The Gaucher disease is genetically inherited and primarily encompasses the patients with defective intracellular hydrolysis of glucosylceramide leading to common lysosomal storage disease. The disease is caused by mutations in the gene encoding the enzyme acid-β-glucosidase (EC 3.2.1.45), Glycoside Hydrolase Family 30 of the Carbohydrate-Active Enzymes database (CAZy) [2]. It is a lysosomal hydrolase cleaving β-glycosyl bonds of glucosylceramide. The normal enzyme is synthesized as a precursor and contains a secretory signal sequence. The proteolytic clipping of the signal sequence occurs during the transport through the endoplasmic reticulum and Golgi apparatus to the lysosome [1, 3].

The deficiencies of the human enzyme result in the disease affecting circa 30,000 individuals. Mutations affect the turnover number, substrate affinity, and/or activator binding [4] leading to reduction of a lysosomal enzyme concentration. This compromises folding in the endoplasmic reticulum and results in proteasomal degradation of the protein [5]. The clinical manifestation of the disease is described on the basis of involvement of central nervous system as 1/ non-neuronopathic, 2/ acute neuronopathic and 3/ subacute neuronopathic [6]. Human acid β-glucosidase is a homomeric protein, most probably the peripheral membrane protein [7,8]. So far, about 20 mutations of acid-β-glucosidase are known [9], however, most of them influence the enzymatic activity leading to disease with the severity depending on the placement of the mutation. In the recent years several attempts to answer the question of the placement of the catalytic center were undertaken. These included site-directed mutagenesis [4], homology modeling [10] and electrospray tandem mass spectrometry [11]. The published results differed in the placement of a substrate-binding site.

Several studies signified the residues D435 and D445 as the sites binding the substrate, whereas others indicated E235 and E340 as a S-binding place. Tandem mass spectrometry studies identified E340 as a nucleophile [11], while site-directed mutagenesis identified the E235 residue as an acid/base catalyst [12]. The point mutation studies revealed the important steric role of D445 residue, but disproved it as a catalytic nucleophile [4]. Recently, the x-ray crystal structure of the enzyme [13] as well as its interactions with an inhibitor [14] have been published making the theoretical analysis approaches more visible and less prone to the methodological errors. In the presented research, the molecular dynamics studies as well as molecular docking were used to address the question of the structural changes induced by the two physiologically important point mutations, L444P and L444R.

Both mutations, although of entirely different physico-chemical nature, result in severe neuronopathic disease. It may be caused by conformational changes and alteration in domains folding [15]. The analysis of clinical manifestation of the disease on the ground of previously published results and the recent x-ray studies attempted to draw correlations between biological activity and point mutations and to shed light on the mechanisms that render the activity of the enzyme.

Methods

MD simulations

Mutant construction

The coordinates of the 3D-structure of the human acid-β-glucosidase (PDB-code: 1OGS) were retrieved from the protein data bank. Single amino acid mutations, L444P and L444R, were placed using the program Swiss-View [16]. In this method an amino acid rotamer library is browsed to to change amino acids side chains. The method allows to place the side chains, which do not cause sterical conflicts with their local environment. The three molecules, the wild form, L444P and L444R mutants, of the enzyme were subjected to molecular dynamics procedure described below. The average structure of last 200 ps of the simulation was subjected to steepest descent minimization procedure to obtain a template for molecular docking experiment.

Molecular dynamics

All the studied proteins, wild type as well as the mutants, were embedded in a solvent box consisting of 9939 and 6264 simple point charges (spc) water molecules [17] and octanol molecules, respectively. A twin-range cutoff was used for long-range interactions: 1.5 nm for electrostatic interactions and 1.2 nm for van der Waals interactions [18]. The Shake algorithm was used to constrain hydrogen bond length [19]. The simulation was performed under the constant pressure of 1 bar applied independently in all three directions. A coupling constant t_P of 0.5 ps and compressibility of 4.5×10^{-5}bar^{-1} were applied. The protein and solvent molecules were coupled separately to an external bath at 300 K, using a coupling constant of t_T= 0.1 ps [20]. In all the studies, the simulation time was 1 ns and the trajectory was integrated with 1 fs time step. All simulations and data analysis were performed using the Gromacs v. 3.0 software [21]. The protein, as well as the water parameters, was those of the oplsa (OPS-AA/L all-atom) force field [22]. Structural diagrams, were prepared using the programs VMD [23], and Pymol [24].

Docking experiments

Following the results reported by Hetenyi and van der Spoel blind docking procedure was used for search for putative binding places [25]. Docking experiments were carried out using the program AutoDock, version 3.0.5 [26] supported by the program ADT [27]. The structure of the macromolecule as well as the substrate, product and inhibitor was held fixed. All the docking runs were performed using the Lamarckian genetic algorithm (LGA). The docking procedure began with a population of random ligand conformations in random orientations and at random translations. During the docking experiment, the population size was 50 and 250000 energy evaluations or 27000 generations were used. The final 150 docked conformations were obtained in 150 different runs. The elitism number, the rate of gene mutation and the rate of gene crossover were equal to 1, 0.02 and 0.8, respectively. The probability that docking solution in the population would undergo a local search was set to 0.06 and the constraint was set to a maximum of 300 iterations per search. The maximum number of successes or failures before changing the size of local search space (rho) was both set to 4. A pseudo-Solis and Wets local search was used to minimize energy of the population. Translations were set to have a maximum limit of 2 Å/step and the orientation step size for the angular component and the torsions had a maximum limit at 50 degrees/step. The data analysis was performed using the custom written Perl and tcl scripts.

The distances between the center of masses of ring carbons for substrate product and inhibitor and the center of mass of Cα atoms constituting the pairs of metabolically important residues were the criterium used for elucidation of clusters of best fitting conformers.

Results and discussion

Structural fluctuations

To provide a picture of the global drift of the studied model in the liquid environment the analysis of the time-dependent Cα rms deviations (RMSD) was conducted (Fig. 1). It may be observed that the maximal drift for all the studied structures is on the order of ~0.17 nm and reaches a plateau after ~0.75 ns of simulation. An initial rise of ~0.05 nm was observed for all the simulations. This feature is most probably owed to the relaxation of the X-ray derived structure. The following rise of RMSD is most probably attributable to inaccuracies in the force field as well as the relaxation motion of proteins.

To examine the fluctuations of the structure on a residue-by-residue basis, the time averaged rms fluctuations (RMSF) of Cα atoms during the last 200 ps of simulation

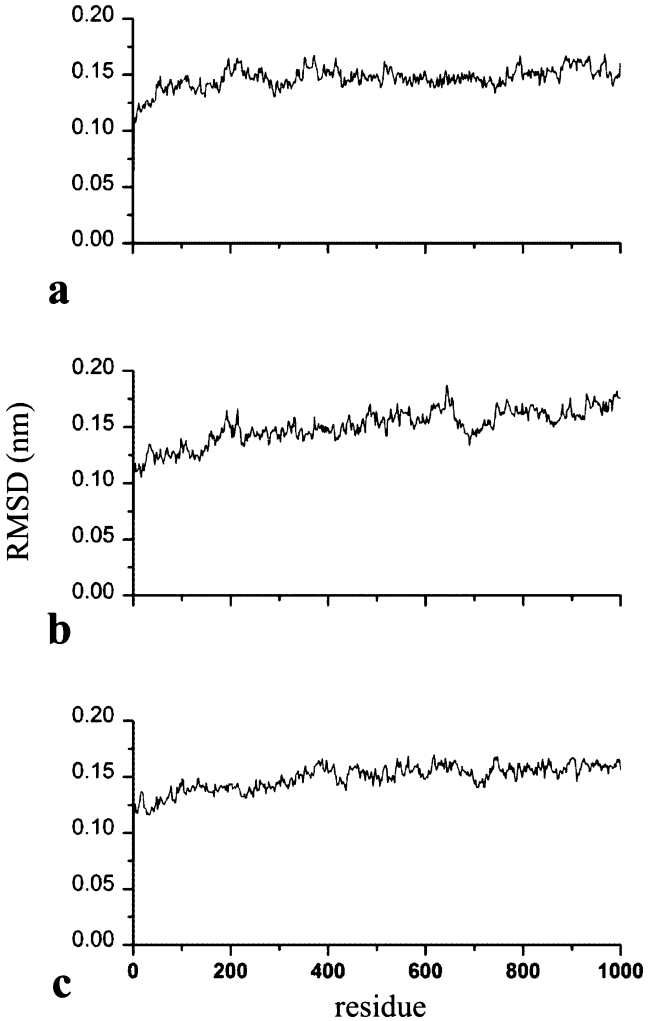

Fig. 1 Drift of protein structure from the initial model. The RMSD of all Cα atoms from the starting structure is shown as a function of time. **a**) wild-type structure; **b**) L444P mutation; **c**) L444R mutation

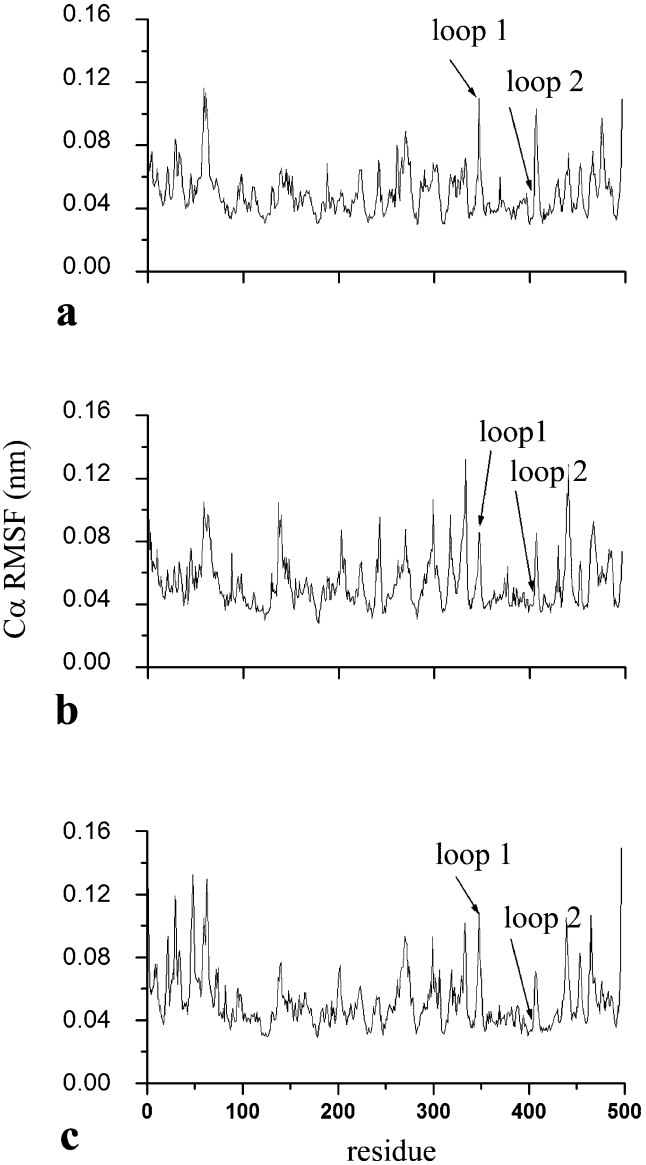

Fig. 2 Fluctuation of protein coordinates. The RMSFs of the Cα coordinates from their time-averaged values (last 200-ps simulation) are shown as a function of residue number for the wild-type (**a**), L444P (**b**) and L444R (**c**) structures. The arrows indicate the positions of loop1 and loop2 comprising the residues 345–349 and 394–399

were analyzed (Fig. 2). The average time period (0.8-1 ns) was determined on the observation of Cα RMSD drift. The analysis of Fig. 2 reveals that RMSFs adopt large values in the vicinity of N- and C-terminus. This behavior is attributed to the presence of loops preceding the terminal ends. Following the studies of Premkumar et al. [14], a special attention was given to the regions Ser^{345}-Glu^{349} (loop1) and Val^{394}-Asp^{399} (loop2). Thus, after rejection of 10 terminal residues from both ends, the standard deviations of Cα fluctuation are equal to ±0.0169 nm, ±0.0160 nm and ±0.0140 nm for wild-type, L444P and L444R structure, respectively. The analysis of standard

deviations of loop 1 and loop 2 resulted in the following values: wild type structure: ±0.020 nm, ±0.0044 nm; L444P: ±0.0142 nm, ±0.0045 nm and L444R: ±0.022 nm, ±0.0065 nm. Taking the standard deviation of RMFS as a flexibility measure, the obtained results point to the loop 1 as more flexible than the overall structure for both wild type and L444R structures. One may also observe that the flexibility of loop 2 is within the L444P and L444R structures significantly smaller than loop 1.

Following the previously presented logic [14], a conclusion may be drawn that indeed only loop 1 may act as a doorkeeper of an active site comprising E235 and E340 residues. Additional information on the structural flexibility is offered by the analysis of time-dependent secondary structure fluctuations (Fig. 3) [28]. Analysis of Fig. 3 reveals high stability of α-helices in the region encompassed by the first 250 and 360–370 residues. The β-sheets are very stable in the region comprising the residues 460–497. The lowest time-dependent stability and lowest structure conservation, induced by the point mutations, is observed for the regions comprising the residues 65–80 and

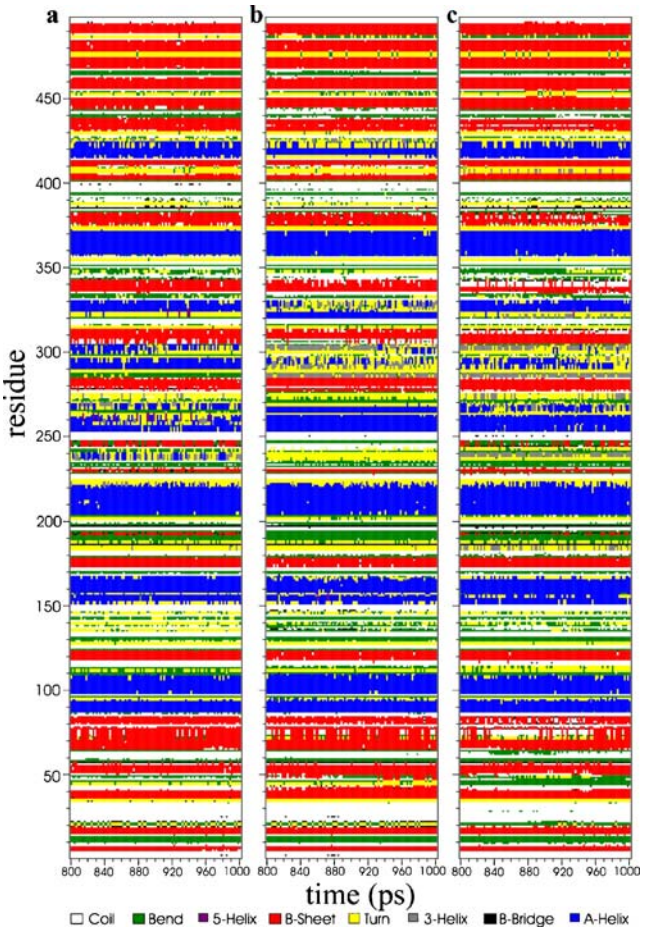

Fig. 3 Secondary structure fluctuations (analyzed using the DSSP program [28] as a function of time for the simulation of **a**) wild-type structure; **b**) L444P mutation; **c**) L444R mutation of human acid-β-glucosidase

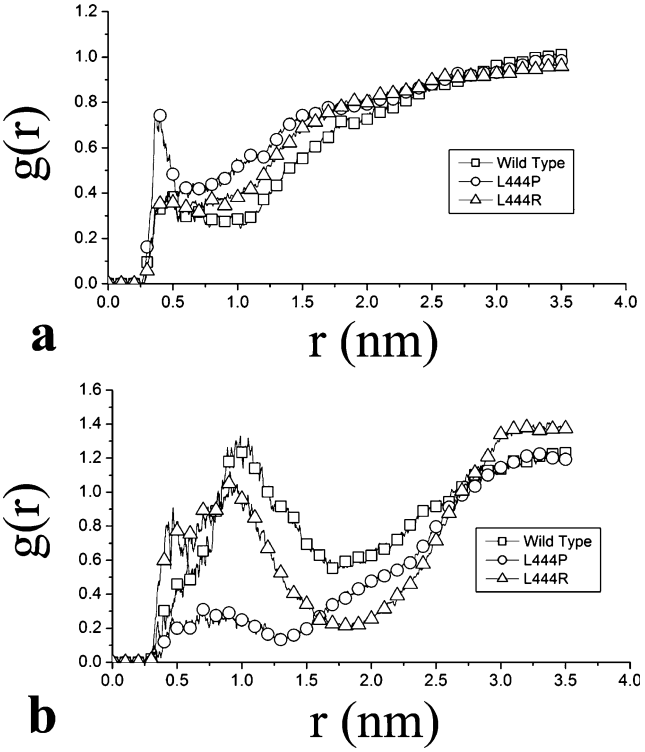

Fig. 4 The radial distribution function (last 200-ps simulation). **a**) water molecules; **b**) octanol molecules (-□- wild type protein, -○- L444P mutation, ¯∈- L444R mutation)

260–340. The region 340–350, i.e. the region comprising the loop 1, is also highly unstable upon introduction of both point mutations. Furthermore, the obtained results indicate that L444R mutation does not influence the flexibility of the loop 1. However, it influences its structural conservation. This observation is in agreement with the biological function of the wild-type and mutated enzyme showing the point mutation having an influence on the activity of an enzyme [29].

Solvent distribution near mutations points

To study the influence of point mutations on saturation of metabolically important residues with solvent molecules, the analysis of the radial distribution function g(r) around the center of masses of Cα carbons of residues 443–445 was performed. According to the analysis, the g(r) function follows the changes in the average density of a medium as a function of a radius from the reference point. Therefore, this function is appropriate for the examination of the degree of solvent saturation of the studied residues. The results of g(r) analysis are shown in Fig. 4. One may observe that the wild-type and L444R structures are defined by similar water distribution (Fig. 4a). The L444P mutation results with a clear maximum of solvent saturation at a distance of approximately 0.4 nm away from the center of masses of Cα atoms of 443–445 residues. However, the

case is otherwise for octanol distribution. On the other hand, the wild-type protein and the L444P mutant are described by similar octanol distribution up to a distance of 1.0 nm away from the reference point. The obtained results are very interesting when confronted with hydrophobic and steric properties of the mutated amino acids. Taking into account their physical features, one could expect that larger amount of octanol molecules should be attracted by proline than the lysine or arginine. However, the obtained data was contrary to that expected. Moreover, L444P mutation leads to a greater saturation with water molecules, whereas L444R mutation to greater saturation with octanol molecules. Such a behavior can be explained considering the steric properties of amino acids. It is possible that the resultant of L444R mutation is due to the exposure of the hydrophobic core of domain II, thus remotely attracting the octanol molecules, whereas L444P mutation is a consequence of the hydrophobic core, which is concealed in the structure of the protein.

Automated docking of a substrate, product and inhibitor

To confirm the placement of the metabolically important residues, a set of blind docking experiments were performed. Thus, the substrate (glucosylceramide - GlcCer), the product (glucose - Gluc), and the inhibitor (conduritol-β-epoxide) (CBE) were docked onto the native x-ray derived structure of GlcCerase (Fig. 5). Such an approach allowed for systematic sampling of the protein surface and for testing the regions of the protein for the substrate, product and inhibitors binding capabilities. The analysis of the results indicates the residues E235, E340 and D443, D445 to be contributing in the construction of the metabolically important sites of the enzyme. On the other hand, analysis of distances between the center of masses of sugar carbons for glucose and GlcCer, the epoxide ring of CBE, and the center of mass of Cα atoms constituting the pairs of metabolically important residues allowed for elucidation of clusters of best fitting conformers. Conse-

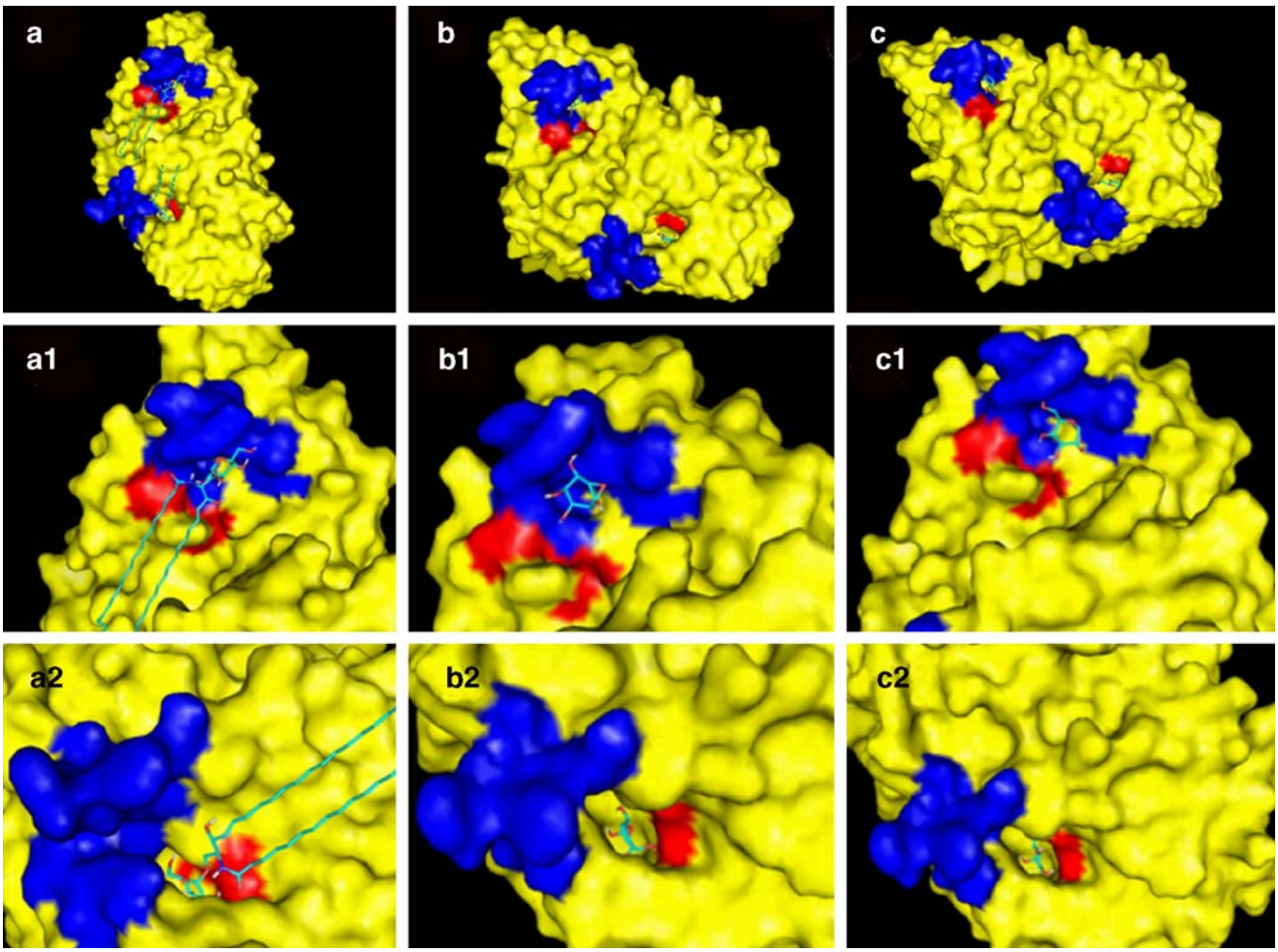

Fig. 5 The binding of an arbitrarily chosen structure of glucosylceramide (**a**), conduritol-β-epoxide (**b**), glucose (**c**) to the active site D443/D445 (A1, B1, C1) and E320/E340 (A2, B2, C2) of the wild-type structure of human acid-β-glucosidase. A1, B1, C1 blue mask represents the residues D443–D445; red mask represents the residue D444; A2, B2, C2 blue mask represents loop1 (Ser^{345}-Glu^{349}) and loop2 (Val^{394}-Asp^{399}), whereas red mask the residues E235 and 340

quently, the substrate docking experiment (Fig. 5a–a2) resulted in only one structure fitting well to E235 and E340 residues and eight structures to the D443–D445 region. The results of the inhibitor (Fig. 5b–b2) and the product (Fig. 5c–c2) docking gave an evidence for good fitting (over 10 conformers fit well to the metabolically important residues) of these molecules to the both metabolically active places.

Conclusion

In 2003, the X-ray structure of GlcCerase at 2.0 Å resolution was reported [13]. This report was followed by the publication of an x-ray structure of an enzyme coupled to its inhibitor [14]. The refined X-ray structure comprises two molecules per asymmetric unit with the overall fold consisting of the three domains. Domain 1 comprises the residues 1–27 and 383–414; domain 2 consists of residues 30–75 and 431–497; and domain 3 is based on residues 76–381 and 416–430. The recent x-ray studies clearly indicate the Glu340 and Glu235 residues as an active center of the enzyme. It has been shown that the activity of an enzyme may be steered by the two flexible loops: loop1 and loop2 (residues 345–349 and 394–399, respectively), which guard the entrance to the proposed active site of an enzyme. The molecular dynamics studies show that only loop 1 is flexible with flexibility higher than that observed for loop 2. The obtained results clearly depict that L444P mutation lowers the flexibility of the loop 1, along with the fact that both the mutations affect the structure stability of the region comprising the loop 1.The solvent distribution studies clearly show the differences in octanol and water saturation suggesting that L444P mutation leads to concealing of the hydrophobic core while the L444R mutation to the exposure of the hydrophobic core of the molecule. Furthermore, the analysis of charge distribution reveals a similar picture for the wild-type and L444P structures. Apparently, L444R mutation yields significantly different results. It is striking that both mutations although of completely different physical properties, result in the same neuropathicity and render the enzyme inactive. This observation leads to the conclusion that the region comprising the residues 443–445 is somehow involved in or steers the binding of substrate, product and/or inhibitor. It is possible that it takes part in hydrophobic/hydrophilic interactions between the enzyme and its environment (lipid membrane). Although both mutations have similar clinical manifestation, the rendering of the enzyme activity is driven by different physical backgrounds. Subsequently, the obtained results indicate that L444P mutation influences directly the substrate binding by decreasing the flexibility of the loop 1 and destabilization of the region comprising this loop. L444R mutation, on the other hand, results in exposure of the hydrophobic core and rendering the enzyme activity. Possibly, it influences the interactions between the enzyme and lipid bilayer. Summarizing, the obtained results show that mutations results in macroscopic changes influencing the overall physical properties of an enzyme, thereby influencing its physiological properties.

Acknowledgments I would like to thank Dr. Arndt Rolfs, Department of Neurology, University of Rostock for the excellent discussion in the preliminary molecular dynamics studies. The molecular dynamics analyses were supported by grant G28-2 of Interdisciplinary Center for Mathematical and Computational Modeling, Warsaw, Poland.

References

1. Erickson AH, Ginns EI, Barranger JA (1985) J Biol Chem 260:14319–14324
2. Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. The Royal Society of Chemistry, Cambridge, pp 3–12
3. Bergmann JE, Grabowski GA (1989) Am J Hum Genet 44:741–750
4. Grace ME, Newman KM, Scheinker V, Berg-Fussman A, Grabowski GA (1994) J Biol Chem 269:2283–2291
5. Sawkar AR, Cheng WC, Beutler E, Wong CH, Balch WE, Kelly JW (2002) Proc Natl Acad Sci (USA) 99:15428–15433
6. Barranger JA, Ginns EI (1989) Glucosylceramide Lipidoses: Gaucher Disease, 6th ed. McGraw Hill Inc., New York
7. Takasaki S, Murray GJ, Furbisch FS, Brady RO, Baranger JA, Kobata A (1984) J Biol Chem 259:10112–10117
8. Berg-Fussman A, Grace ME, Grabowski GA (1993) J Biol Chem 268:14861–14866
9. Horowitz M, Zimran A (1994) Hum Mutat 3:1–11
10. Fabrega S, Durand P, Mornon J-P, Lehn P (2002) J Soc Biol 196:151–160
11. Miao S, McCarter JD, Grace ME, Grabowski GA, Aebersold R, Withers SG (1994) J Biol Chem 269:10975–10978
12. Fabrega S, Durand P, Codogno P, Bauvy C, Delomenie C, Henrissat B, Martin BM, McKinney C, Ginns EI, Mornon JP, Lehn P (2000) Glycobiology 10:1217–1224
13. Dvir H, Harel M, McCarthy AA, Toker L, Silman I, Futerman AH, Sussman JL (2003) EMBO Rep 4:704–709
14. Premkumar L, Sawkar AR, Boldin-Adamsky S, Toker L, Silman I, Kelly JW, Futerman AH, Sussman JL (2005) J Biol Chem 280:23815–23819
15. Morel N, Bon S, Greenblatt H, Wodak S, Sussman JL, Massoulie J, Silman I (1999) Mol Parmacol 55:982–992
16. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
17. Berendsen HJ, Postma JP, van Gunsteren WF, Hermans J (1981) Intramolecular Forces, D. Reidel Publishing Company, Dordrecht
18. Zubrzycki IZ (2002) Biophys J 82:2906–2915
19. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Journal of Computational Physics 23:327
20. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81:3684
21. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) J Comput Chem 26:1701–1718

22. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
23. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33–38, 27–38
24. DeLano W, The PyMOL Molecular Graphics System. 2002, DeLano Scientific, San Carlos
25. Hetenyi C, van der Spoel D (2006) FEBS Lett 580:1447–1450
26. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) J Comput Chem 19:1639
27. Sanner MF (1999) J Mol Graph Model 17:57–61
28. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
29. Uchiyama A, Tomatsu S, Kondo N, Suzuki Y, Shimozawa N, Fukuda S, Sukegawa K, Taki N, Inamori H, Orii T (1994) Hum Mol Genet 3:1183–1184