Journal of Molecular Graphics and Modelling 54 (2014) 107–113

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Lessons from molecular modeling human α-L-iduronidase

Danieli Forgiarini Figueiredo[a,d,1], Dinler A. Antunes[a,d,1], Maurício M. Rigo[a,d], Marcus F.A. Mendes[a,d], Jader P. Silva[a], Fabiana Q. Mayer[b], Ursula Matte[c,d], Roberto Giugliani[c,d], Gustavo F. Vieira[a,d], Marialva Sinigaglia[a,d,*]

[a] NBLI – Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[b] Molecular Biology Laboratory, Instituto de Pesquisas Veterinárias Desidério Finamor, Fundação Estadual de Pesquisa Agropecuária Porto Alegre, Rio Grande do Sul, Brazil
[c] Gene Therapy Center, Experimental Research Center, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil
[d] Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

ARTICLE INFO

Article history:
Accepted 8 October 2014
Available online 18 October 2014

Keywords:
Homology modeling
Low identity template
α-L-Iduronidase (IDUA)
Model evaluation tools
Molecular dynamics
Secondary structure assessment

ABSTRACT

Human α-L-iduronidase (IDUA) is a member of glycoside hydrolase family and is involved in the catabolism of glycosaminoglycans (GAGs), heparan sulfate (HS) and dermatan sulfate (DS). Mutations in this enzyme are responsible for mucopolysaccharidosis I (MPS I), an inherited lysosomal storage disorder. Despite great interest in determining and studying this enzyme structure, the lack of a high identity to templates and other technical issues have challenged both bioinformaticians and crystallographers, until the recent publication of an IDUA crystal structure (PDB: 4JXP). In the present work, four alternative IDUA models, generated and evaluated prior to crystallographic determination, were compared to the 4JXP structure. A combined analysis using several viability assessment tools and molecular dynamics simulations highlights the strengths and limitations of different comparative modeling protocols, all of which are based on the same low identity template (only 22%). Incorrect alignment between the target and template was confirmed to be a major bottleneck in homology modeling, regardless of the modeling software used. Moreover, secondary structure analysis during a 50 ns simulation seems to be useful for indicating alignment errors and structural instabilities. The best model was achieved through the combined use of Phyre 2 and Modeller, suggesting the use of this protocol for the modeling of other proteins that still lack high identity templates.

1. Introduction

Classical studies with globins provided the first example of a "molecular disease" [1,2], the first clues regarding the impact of amino acid exchanges on protein structure/function, and their consequences to human health [3–5]. The globin family also provided evidence that protein structures evolve more conservatively than protein sequences [6], and the tertiary structure conservancy observed among homologues enables comparative structure predictions. Since then, accurate prediction of the 3D structure of proteins related to human diseases has been a major goal in structural bioinformatics [7].

Great progress has occurred over the past few decades, with improvements in the algorithms and computational resources. The Critical Assessment of protein Structure Prediction (CASP) competitions, which biannually challenge researchers to provide accurate models of an unreleased crystal structure, provide a wealth of information to this field [8,9]. These competitions also highlight the evolution of structure prediction and the persistent limitations and bottlenecks for comparative modeling [10,11].

In the present work, we focus on the homology modeling of the human α-L-iduronidase (IDUA, E.C: 3.2.1.76), a member of the glycoside hydrolase family [12]. Mutations in this enzyme are responsible for an inherited lysosomal storage disorder, called MPS I (Mucopolysaccharidosis I, OMIM #607014, #607015, #607016) [13,14]. The enzyme has 653 amino acids and is synthesized on the rough endoplasmic reticulum (ER), presenting a signal peptide with

* Corresponding author at: Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, Building 43323, Room 225, Brazil. Tel.: +555133089938
E-mail address: msinigaglia@gmail.com (M. Sinigaglia).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jmgm.2014.10.004
1093-3263/

26 residues in its N-terminal portion. There are six glycosylation sites in its sequence, which are important for enzyme trafficking to the lysosome [15]. The enzyme architecture corresponds to a $(\alpha\beta)_8$ domain, which is a conserved structure that is also known as a TIM barrel fold [12].

Determining the 3D structure of this important enzyme would be pivotal for understanding phenotypic variations presented by MPS I [13], for predicting the impact of mutations [16] and for devising new pharmacological treatments [14]. However, technical issues postponed by years the determination of a crystal structure and the absence of a high identity template has been a major challenge for accurate modeling [12,17,18]. The recent publication of a human α-L-iduronidase crystal structure [19] allowed us to directly assess the efficacy and limitations of different comparative modeling protocols, all performed with a low identity template (only 22%). This convenient experiment is able to provide important lessons on comparative modeling and structural analysis, which can be applied for other challenging molecular targets involved in human diseases.

2. Materials and methods

2.1. Molecular modeling of α-L-iduronidase (IDUA)

The human α-L-Iduronidase FASTA sequence was recovered from UNIPROT [20] (code P35475), and the first 26 amino acids, which belong to the signal peptide, were removed. Three different models of IDUA were then generated using alternative approaches. In each case, one model was selected among several, after an evaluation with the proper tools.

2.1.1. Model IDUA I-TASSER

The I-TASSER server was used to produce five models, which were ranked by the C-score and TM-score [21]. After the server is provided with the amino acid sequence of the target, it retrieves template proteins of similar folds from the PDB by LOMETS (Local Meta-Threading-Server). Afterwards, contiguous fragments recovered from the PDB are reassembled into full-length models, filling the missing regions using *ab initio* modeling. Additional steps are performed to remove the steric clash and to refine the global topology of the models (more information at http://zhanglab.ccmb.med.umich.edu/I-TASSER/).

For the IDUA sequence, I-TASSER identified 1uhvA as the best template for modeling (id1 = 0.21, id2 = 0.18, cov 0.75 and *z*-score = 2.14). It should be noted that the models produced by this approach lack a C-terminal region that was not modeled, and modeled structures correspond to residues 27–635 of the wild-type IDUA. The best model presented a C-score of −1.58 and TM-score of 0.52 ± 0.15, which are indicative of a reliable model with a correct global topology.

2.1.2. Model IDUA Rempel et al./Modeller (IDUA-RM)

Modeller 9v9 [22] was applied to produce a model using the crystal structure of beta-D-xylosidase from *Thermoanaerobacterium saccharolyticum* (PDB: 1UHV) as a template. The template structure was obtained after a combined search using psiBLAST and BLASTp from NCBI as well as HHPRED from the Max Plank Institute (http://toolkit.tuebingen.mpg.de/hhpred). The sequence alignment used for this approach was the same as previously published by Rempel and colleagues [17] (PDB: 1Y24). One hundred models were generated that covered residues from 36 to 504 of the wild-type IDUA, and the best model was selected using the DOPE score [22] and Procheck [23].

2.1.3. Model IDUA Phyre 2.0/Modeller (IDUA-PM)

In this approach, Modeller 9v9 [22] was applied using the same template (PDB: 1UHV) but with a different alignment. Sequence alignment covered residues from 27 to 542 of the wild-type IDUA (Figure S1), presenting 22% identity with the template sequence, and this was performed by folding recognition with Phyre 2 (http://www.sbg.bio.ic.ac.uk/phyre2) [24]. The alignment was further checked manually and adjusted, considering the location of insertion/deletion in loops and the correct positioning of crucial residues from the catalytic domain. One hundred models were generated, and the best model was selected using the DOPE score [22] and Procheck [23].

Supplementary Figure S1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

2.1.4. Parameters for homology modeling with Modeller

Homology modeling with Modeller 9v9 was performed in a semi-automated fashion through the use of python scripts previously developed by our group. The modeling protocol followed the default optimization and refinement protocol, as described in the Modeller online manual (available at http://salilab.org/modeller/9.13/manual/node19.html). Briefly, each model is optimized with the variable target function method (VTFM) with conjugate gradients (CG), and later refined using molecular dynamics (MD) with simulated annealing (SA).

2.2. Comparison of generated models with IDUA crystal structure

The best model for each approach was evaluated using Procheck [23], Verify 3D (http://nihserver.mbi.ucla.edu/Verify_3D/) [25,26] and ModFOLD (http://www.reading.ac.uk/bioinf/ModFOLD/) [27], and the results were compared with those obtained for a crystal structure of IDUA (PDB: 4JXP). The IDUA model previously published by Rempel and colleagues [17] (PDB: 1Y24) was also included in this analysis.

2.3. Refinement of an IDUA crystal structure for molecular dynamics

The recently published crystal structure of human α-L-iduronidase (PDB: 4JXP) had two missing sites (residues 55–61 and 103–106) [19]. These sites had to be modeled before submitting this structure to a molecular dynamics (MD) simulation. Using the full sequence of human IDUA from UNIPROT [20] (code P35475), a new crystal structure containing missing loops (IDUA-Crystal) was generated with Modeller 9v9 [22].

Of note, the 4JXP crystal structure was later exchanged at PDB by the 4MJ4 structure. There are no relevant structural differences between 4JXP and 4MJ4, and both present the same sequence gaps. However, the 4MJ4 sequence presents three amino acid exchanges with respect to 4JXP and to the human α-L-iduronidase sequence recovered from UNIPROT [20] (code P35475). Because this work was performed before this exchange at PDB and because 4JXP presented the same reference sequence used for all models, we did not exchange our reference structure for 4MJ4.

2.4. Molecular dynamics simulations

IDUA models (IDUA-ITASSER, IDUA-RM, IDUA-PM and 1Y24) and the IDUA crystal structure were subjected to 50 ns of a molecular dynamics simulation using the GROMACS v4.5.1 package [28] on a Linux platform using the GROMOS96 (53a6) force field. An appropriate number of sodium (Na+) and chloride (Cl−) counterions was added to neutralize the system, with a final concentration

of 0.15 mol/L. A cubic box was defined with at least 15 Å of solvation layer around the protein, using a *SPC* water model and periodic boundary conditions. The *v-rescale* (tau_t = 0.1 ps) and *parrinello-rahman* (tau_p = 2 ps) algorithms were used for temperature and pressure coupling, respectively. Cutoff values of 1.2 nm were used both for van der Waals and Coulomb interactions, with *Fast Particle-Mesh Ewald* electrostatics (PME).

Our MD simulations were divided into four main stages: energy minimization (EM), solvation, thermalization and production. The EM stage was subdivided into three steps. First, the *steepest-descent* algorithm with position restraints for protein heavy atoms (5000 kJ^{-1} mol^{-1} nm^{-1}) was applied, allowing only the solvent to relax. Afterwards, an EM with the same algorithm and no restraints was performed, allowing relaxation of the entire system. Finally, an EM using a *conjugate gradient* (CG) algorithm with no restraints was performed. The solvation stage was divided into several steps. First, an MD simulation with an *md* integrator algorithm and position restraints for all protein heavy atoms (5000 kJ^{-1} mol^{-1} nm^{-1}) was performed at a temperature of 300 K for a period of 500 ps to allow for solvation layers formation. Then, temperature was reduced to 20 K (2 steps, total of 20 ps), after which position restraints were gradually reduced to 0.2 kJ^{-1} mol^{-1} nm^{-1} (11 steps, total of 130 ps). During thermalization, the system was gradually heated from 20 to 300 K (with no restraints), increasing by approximately 50 K at each 320 ps step. Together, these equilibrium stages complete 2500 ps of simulation. This is the initial time for the production stage, in which the system was held at a constant temperature (300 K) and had no restraints up to 50 ns.

Simulation plots were generated with the respective programs from GROMACS v4.5.1 package [29] and visualized with xmgrace, the full-featured GUI-based version of Grace (http://plasma-gate.weizmann.ac.il/Grace/). Visual inspection of the MD trajectories was performed with VMD 1.9.1 [30], PyMOL 1.0 [31] and UCSF Chimera [32].

2.5. RMSD weighted by reference structure dynamics

Protein frames of each simulation were recovered (each 5 ns) and used to calculate the RMSD (root-mean-square deviation) against the reference structure (IDUA-Crystal). This was obtained following the rationale: "$(Smod_{t(x)} - Scryst_{t(0)}) - (Scryst_{t(x)} - Scryst_{t(0)})$, where "$Smod_{t(x)} - Scryst_{t(0)}$" is the RMSD of the modeled structure at a time x, in relation to the crystal structure (time = 0). Only sections of β-sheets and the α-helix of the TIM barrel (excluding loops) were considered because the great variability of the loops was expected to be even for the reference structure. The limits of each secondary structure were defined by considering the crystal structure sequence (PDB: 4JXP). An alternative weighted RMSD was calculated by discounting the variability presented by the IDUA-Crystal at each time point of the simulation. In both cases, the RMSD was calculated for all atoms (backbone and side chains).

3. Results and discussion

3.1. Previous models of human α-L-iduronidase (IDUA)

The human α-L-iduronidase is an enzyme from the glycoside hydrolases group and is involved in the catabolism of glycosaminoglycans (GAGs), heparan sulfate (HS) and dermatan sulfate (DS) [12,13]. Accurate modeling of α-L-iduronidase was a major goal for researchers because several mutations are related to a broad variety of clinical manifestations of mucopolysaccharidosis type I (MPS I) [16,33]. However, lack of high identity templates and difficulties in producing a crystal structure of the enzyme have challenged researchers. Rempel and colleagues published the first model for the IDUA enzyme in 2005 [17], which was deposited at PDB under the code 1Y24. This model was obtained with the SWISS-MODEL online server [34], which performs homology modeling in an automated fashion, and used the beta-D-xylosidase of *T. saccharolyticum* (PDB: 1UHV, 1PX8) as a template. This model was downloaded and included in our analysis to evaluate the impact of alternative alignments and molecular modeling tools.

Recently, a second model of IDUA was published by Chandar & Mahalingman using the automated tool Schrödinger PRIME [18]. The structure, however, was not made available.

The IDUA crystal structure was finally published in November 2013 by Bie and colleagues [19] and provided insightful information on the complete structure of this important enzyme. This structure was initially made available under the PDB code 4JXP, which was afterwards replaced by the 4MJ4 structure (see Section 2). It is important to consider that all models from the present study were generated before the publication of the IDUA crystal structure and were therefore predicted and evaluated without any influence from these experimental data.

3.2. Evaluation of generated models and comparison with a reference crystal structure

A Ramachandran plot is a well-known evaluation tool to assess the stereochemical quality of a given model through the analysis of *phi* and *psi* angles for all protein residues [23]. This analysis was performed with Procheck software [23] for the best models produced by three alternative approaches (Table 1) as well as for the previously published model (PDB: 1Y24) and for the reference crystal structure (PDB: 4JXP). Considering this analysis, the best models were IDUA-RM and IDUA-PM, which had 87.3% and 85% of residues in the most favored regions, respectively. Both models were produced with Modeller [22], which indicates the ability of this software to assign stereochemical properties. The percentage of residues in disallowed regions was also lower for these two models when compared to 1Y24 and IDUA-ITASSER, which was the worst model in this analysis (Table 1).

The overall quality of all models was also evaluated using other two well-known pieces of software, ModFOLD [27] and Verify3D [26] (Table 2). Models 1Y24 and IDUA-RM had a global quality score lower than the others at ModFOLD and have presented a low percentage of its residues (<60%) with an averaged 3D-1D score >0.2 at Verify 3D, which configures that this model failed in this evaluation. IDUA-PM and IDUA-ITASSER presented high confidence at ModFold (probability of an incorrect model at lower than 1/100) and were approved at Verify 3D. Moreover, IDUA-PM presented the highest value of global quality (0.3734). Of note, values above 0.4 indicate high similarity with the native structure. Considering these three independent methods for viability assessment, IDUA-PM was considered to be the best IDUA model generated in this work.

As a crystal structure of IDUA was recently made available (PDB: 4JXP), it is possible to directly assess which model was closest to the native structure (Fig. 1). Visual inspection indicates a similar arrangement of the catalytic domain (TIM barrel) for all models. However, important differences were observed among the models when considering the specific amino acids sequence that composes the structure of each TIM barrel. These differences are highlighted by performing the root mean square deviation (RMSD) for the crystal structure backbone with respect to the same residues in each proposed model (Table 3). Again, the closest model to the native structure was IDUA-PM, which presented an RMSD of 5.78 Å for the catalytic domain and 6.17 Å for the whole modeled structure. Although this value was high, it is 2-fold smaller than that of the previous model (1Y24), which had an RMSD of 12.24 Å for

Table 1
Percentage (%) of protein residues in each region of the Ramachandran Plot.

Regions	4JXP (crystal)	1Y24 (model)	IDUA-RM (model)	IDUA-PM (model)	IDUA-ITASSER (model)
Most favored	86.5	72.5	87.3	85.0	76.7
Additional allowed	12.9	23.4	11.0	12.3	16.5
Generously allowed	0.6	2.4	1.0	1.6	4.1
Disallowed	0.0	1.7	0.7	1.1	2.7

All values are given as %.

Table 2
Additional evaluation tools for model viability assessment.

Models	ModFOLD			Verify 3D	
	Global model quality score	*P*-value	Confidence	Residues with an averaged 3D-1D > 0.2[a]	Results
1Y24	0.2081	2.663E−2	Intermediate	56.29	Reproved
IDUA-RM	0.2160	2.288E−2	Intermediate	56.08	Reproved
IDUA-PM	0.3734	1.898E−3	High	86.05	Approved
IDUA-ITASSER	0.2961	5.747E−3	High	80.79	Approved
4JXP	0.8555	1.239E−4	Cert	97.07	Approved

[a] Values given in %.

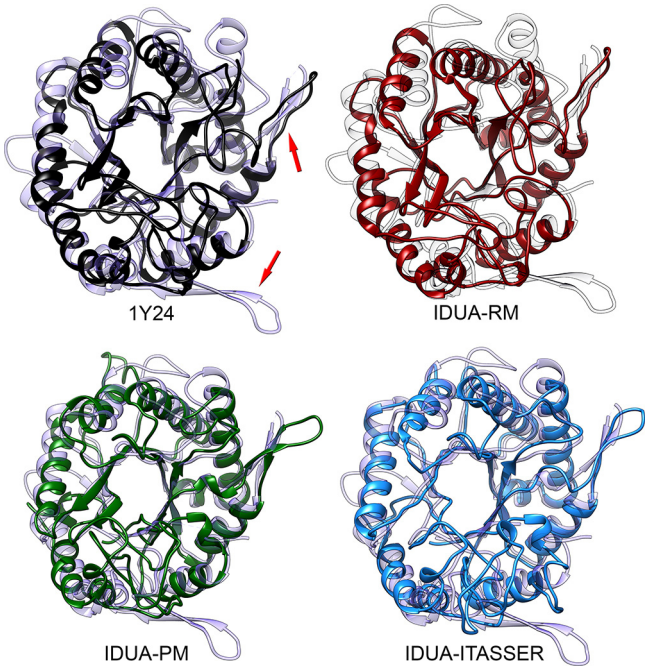

Fig. 1. Catalytic domain of α-L-iduronidase (TIM Barrel). Superposition of different models over IDUA crystal structure (4JXP), which is depicted in light gray. Red arrows indicate two β-hairpins which protract from TIM barrel domain. Observe that one of these β-hairpins was not properly folded in any of the models, and the other was completely folded only in IDUA-PM.

Table 3
RMSD against crystal structure.

Model	RMSD[a] (TIM barrel)	RMSD[a] (protein)
1Y24	12.24 Å	16.65 Å
IDUA-RM	11.84 Å	16.7 Å
IDUA-PM	5.78 Å	6.17 Å
IDUA-ITASSER	6.35 Å	9.12 Å

[a] Backbone Root Mean Square Deviation (RMSD). Values correspond to the RMSD of specific amino acids which compose each model TIM barrel, against the same residues in IDUA crystal structure (PDB: 4JXP).

the TIM barrel and 16.65 Å for the entire structure. The second model, which had better results in this analysis, is IDUA-ITASSER, which presented a slightly larger RMSD for the catalytic domain (6.35 Å).

The TIM barrel domain was already present at the structure of beta-D-xylosidase of *T. saccharolyticum* (PDB: 1UHV), which was used as the template for three of the models evaluated in our study (1Y24, IDUA-RM and IDUA-PM). Comparison between this template and the reference crystal structure for human α-L-iduronidase (PDB: 4JXP) reveals some differences that contributed to the high RMSD values observed for the models (Figure S2). In spite of the great structural similarity between the two catalytic domains, it is important to note the presence of two β-hairpins that protract from the IDUA barrel. One of these structures was not present in the template and, consequently, was not properly folded in any of the models (Fig. 1). Moreover, it is also possible to observe the absence of the fibronectin domain in the beta-D-xylosidase structure (a domain that was not included in our modes) as well as small differences at the β-sandwich domain.

Supplementary Figure S2 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

The choice of the best model and the sequence alignment are key points for molecular homology modeling [35,36]. When sequence identity between the target and template is higher than 90%, highly accurate models are obtained. On the other hand, sequence alignment represents the major bottleneck for homology modeling in cases where the sequence identity is lower than 25%, which may produce crucial errors in the generated models [37,38]. In our work, IDUA has sequence identity of only 22% with the chosen template, and sequence alignment was decisive for obtaining a more reliable model. Our IDUA-PM model, which was generated with Modeller [22] using an alignment from Phyre 2 [24], was the closest model in relation to the crystal structure. The most dramatic errors were observed for the previously published model, 1Y24, and for IDUA-RM, which used the same alignment, regardless of the use of different software for modeling (Swiss PDBViewer [34] and Modeller [22], respectively).

Differences among models become clearer when comparing the maps of secondary structures (Figure S3). A shorter sequence was used to model the catalytic domain of the 1Y24 and IDUA-RM models (the region between the red boxes). Moreover, great divergence of secondary structures is observed when comparing these models with the crystal structure, especially in the region between the α7 and α8 helices. These models also lack a β1 sheet and two β-hairpins (indicated with "C" and "D"). Although present in these models, the β6 and β8 sheets were formed in an incorrect region of the sequence. Finally, there are also problems with the size and location of the α3, α6 and α8 helices. On the other hand, IDUA-PM

and IDUA-ITASSER models presented the TIM barrel in the same region as the crystal structure, with extensive agreement in the alignment of their secondary structures. The IDUA-PM model lacks β1 and β8 sheets, but all helices are in place. Moreover, one of the β-hairpins was correctly folded (Figure S3).

Supplementary Figure S3 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

3.3. Molecular dynamics of the modeled structures

Molecular dynamics simulations were also used to evaluate the models and to compare the dynamics to the reference structure (IDUA-Crystal). This powerful approach allows us to assess structure dynamics in solution, which provides insightful information on protein stability, and has been successfully applied to evaluate structural models [39,40]. In some cases, it has also been used to refine models [41,42]. However, a recent study presenting simulation results for 24 proteins selected from CASP showed that, in most cases, proteins drifted away from crystal structure [43]. According to the authors, force field accuracy seems to be a limiting factor for this purpose.

In the present work, great stability was observed for the IDUA-Crystal regarding the maintenance of secondary structures during a 50 ns simulation (Figure S3). Greater variation was observed for the proposed models, especially considering the catalytic domain of the 1Y24 and IDUA-RM models. The alignment error becomes evident when considering the secondary structures of these models. The end of the TIM barrel domain is defined by α8 and β8 structures, which in these models were located before the correct position. This error was highlighted by molecular dynamics, which indicated great instability at this region.

The IDUA-PM and IDUA-ITASSER models had an overall stability similar to the reference structure. The IDUA-PM model presented partial alterations in some helices and the unfolding of the β7 sheet, which was also observed for the other models. Of note, the correctly predicted β-hairpin was stable throughout the simulation. The IDUA-ITASSER model had similar results. In this case, it is interesting to observe the formation of the β8 sheet after 10 ns of simulation. This structure was not predicted in the input model.

Considering the RMSD of the catalytic domain, the IDUA-Crystal presented stability after 16 ns of simulation, with lower values when compared to the models (Fig. 2). Models 1Y24, IDUA-RM and IDUA-ITASSER presented a similar RMSD, achieving stability after only 30 ns. Greater instability was observed in this analysis for the IDUA-PM model, which achieved stability after only 35 ns, with higher RMSD values. This analysis refers to the deviation of catalytic domain (including loops) for each of the models in relation to its own conformation at the beginning of the simulation. The main idea was to observe the overall variability of initial structures throughout the simulation.

Using as the input the structural data from molecular dynamics, an alternative validation method was performed. First, the RMSD against the reference structure (IDUA-Crystal) was calculated for ten frames (each 5 ns) that were recovered from each simulated model (Figure S4). Afterwards, aiming to highlight the dynamic similarities in respect to the reference structure, a "weighted RMSD" was calculated by discounting the IDUA-Crystal variability at each time of simulation (Fig. 3). This analysis indicates a tendency of all models to converge with the reference structure throughout the simulation (reduction of RMSD values). Moreover, considering the maintenance of the 3D scaffold of the catalytic domain, IDUA-ITASSER and IDUA-PM were the two models that most closely matched the dynamic behavior of the reference structure. In this analysis, the average RMSD for IDUA-ITASSER and

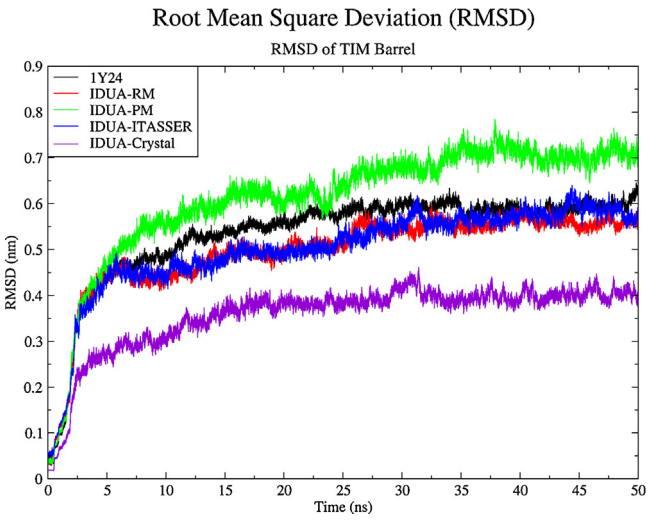

Fig. 2. TIM Barrel Root Mean Square Deviation (RMSD) of each simulated model and crystal structure (Model-4JXP), throughout a 50 ns molecular dynamics simulation. Each line indicates the divergence of a given structure in relation to its own initial conformation.

IDUA-PM was 3.66 and 4.28 Å, respectively (for all atoms). Thus, despite the fact that the RMSD against initial structures (Fig. 2) indicates greater stability for 1Y24 and IDUA-RM, it actually reflects an alternative (incorrect) folding of the catalytic domain in these models and a higher value of RMSD against the reference structure (Fig. 3).

Supplementary Figure S4 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

The higher RMSD values observed for IDUA-PM in Fig. 2 reflect a small increase in diameter of the catalytic domain during the simulated time (Figure S5), without domain unfolding, which might be influenced by the absence of the other domains of the enzyme. In spite of that, IDUA-PM presented a more precise modeling of the TIM barrel, which is sustained throughout the simulation (Fig. 3 and Figure S3). Regarding the radius of gyration (RoG) analysis, a better

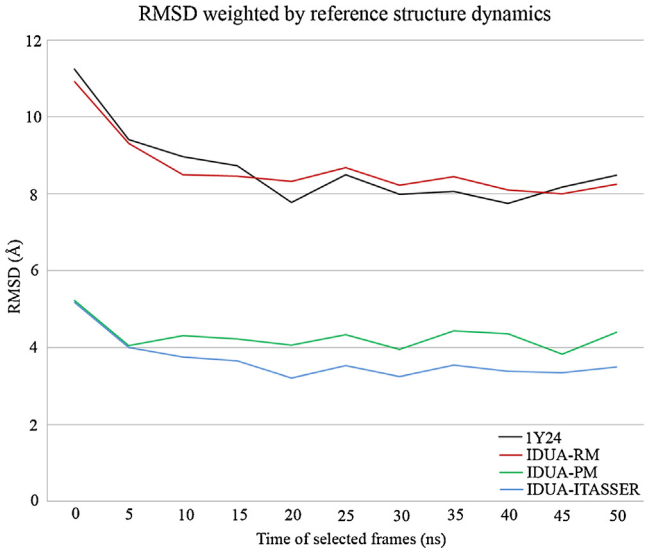

Fig. 3. TIM Barrel Root Mean Square Deviation (RMSD) weighted by reference structure dynamics. Each line indicates the divergence of a given structure in relation to the reference structure (Model-4JXP), discounting the divergence observed during Model-4JXP simulation. Direct RMSD against Model-4JXP (without discounts) can be seen in Figure S3.

fit can also be observed of IDUA-PM and IDUA-ITASSER values with those from the reference structure at the first half of simulation (Figure S5). The IDUA-RM and 1Y24 models presented lower values of RoG, which was also a consequence of using a shorter sequence for catalytic domain folding (wrong alignment). Additionally, the IDUA structure is highly glycosylated *in vivo*, and the absence of glycans in our models can also contribute to the observed instability. Recent structural analyses revealed the importance of N372 glycosylation for enzyme activity, indicating a direct interaction between an N-glycan mannose residue and the substrate [19,44].

Supplementary Figure S5 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

Finally, a root mean square fluctuation (RMSF) analysis was also performed for the TIM barrel residues of each model during the simulation. Little overall fluctuation was observed for catalytic domain residues of the IDUA-Crystal, presenting only six peaks of higher flexibility (Figure S6). These peaks are related to coil regions (near residues 60, 187 and 350) and two β-hairpins (near residues 105 and 373). One exception is made for the third peak (near residue 165), which corresponds to the final portion of α3, a region that was unfolded during simulation (Figure S3). Regarding the models, IDUA-ITASSER had the most similar behavior in comparison to the IDUA-Crystal, and 1Y24 presented the greatest divergence.

Supplementary Figure S6 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.004.

Therefore, molecular dynamics provided a series of additional data, which allowed us to compare our models with the crystal structure, highlighting the limitations of each model. Alternatively, a crystal of other proteins from the same family could be used as a reference structure, providing information on the dynamic behavior of the domains of interest (*e.g.*, TIM Barrel). Even in the absence of a reference crystal structure, these analyses could be used as additional tools for model viability assessment. For instance, secondary structure analysis throughout the simulation was shown to be an efficient way to detect alignment problems and segments that were wrongly folded (Figure S3).

3.4. Applicability to other targets

Differences between the results of alternative methods and alignments indicate hidden obstacles to automated production of reliable models, highlighting the importance of careful modeling and viability assessment. The combined use of Phyre 2 [24] and Modeller [22] for accurate alignment and homology modeling provided an improved model of IDUA 3D structure. This goal was achieved using a template with only 22% sequence identity, and the model quality was confirmed by different methods: Ramachandran, ModFOLD, Verify3D, molecular dynamics and direct comparison with crystal structure. Our results suggest the applicability of this combined approach to other important targets that still lack high similarity templates. For instance, the GAGs degradation pathway is composed of 11 enzymes [45,46], five of which have not yet had their structure determined by experimental methods: iduronate 2-sulfatase, heparan N-sulfatase, α-N-acetylglucosaminidase, acetyl CoA:alpha-glucosaminide N-acetyltransferase and N-acetylglucosamine-6-sulfatase, where deficiencies cause MPS II, MPS IIIA, MPS IIIB, MPS IIIC and MPS IIID, respectively. These are highly important targets for structural analysis because mutations of these enzymes are related to different types of MPS [45,46]. Furthermore, mutations of enzymes responsible for the GAGs synthesis are also implicated in human genetic disorders and could also be a target of molecular modeling for structural analysis [47].

4. Conclusions

Molecular homology modeling of proteins with low identity templates remains a highly difficult challenge, which must be addressed by the combined use of alternative approaches. The correct alignment between target and template is a critical step and has a direct impact on model quality. In addition, the careful use of reliable evaluation tools over the best ranked models is a critical step, which also benefits from the combined use of tools with different characteristics. Finally, the conservancy of known secondary structure patterns through molecular dynamics provides a powerful tool for model quality assessment.

In the present work, four alternative models of the human α-L-iduronidase were evaluated and compared with the recently published crystal structure (PDB: 4JXP). Superposition of catalytic domains from the best models (IDUA-PM and IDUA-ITASSER) with the reference crystal structure presented an RMSD of 5.78 Å and 6.35 Å (TIM barrel backbone), respectively. These values are twice smaller than the previously published model (PDB: 1Y24), which presented an RMSD of 12.54 Å (TIM barrel backbone). In conjunction with other conducted evaluations, this result highlights the performance gain of our approach and justifies the use of a similar protocol for modeling of proteins that lack high identity templates, such as other enzymes of the GAGs degradation pathway.

Acknowledgments

This work was supported through scholarships from CNPq, CAPES and FAPERGS. We would also to thank the Centro Nacional de Supercomputação (CESUP-RS) for allowing access to its computational resources.

References

[1] L. Pauling, H.A. Itano, et al., Sickle cell anemia: a molecular disease, Science 110 (1949) 543–548.
[2] V.M. Ingram, A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin, Nature 178 (1956) 792–794.
[3] J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, D.C. Phillips, A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, Nature 181 (1958) 662–666.
[4] M.F. Perutz, M.G. Rossmann, A.F. Cullis, H. Muirhead, G. Will, A.C. North, Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis, Nature 185 (1960) 416–422.
[5] J. Baldwin, C. Chothia, Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism, J. Mol. Biol. 129 (1979) 175–220.
[6] J.T. Lecomte, D.A. Vuletich, A.M. Lesk, Structural divergence and distant relationships in proteins: evolution of the globins, Curr. Opin. Struct. Biol. 15 (2005) 290–301.
[7] J.G. Mullins, Structural modelling pipelines in next generation sequencing projects, Adv. Protein Chem. Struct. Biol. 89 (2012) 117–167.
[8] J. Moult, A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction, Curr. Opin. Struct. Biol. 15 (2005) 285–289.
[9] A. Kryshtafovych, K. Fidelis, J. Moult, CASP9 results compared to those of previous CASP experiments, Proteins 79 (Suppl. 10) (2011) 196–207.
[10] A. Runthala, Protein structure prediction: challenging targets for CASP10, J. Biomol. Struct. Dyn. 30 (2012) 607–615.
[11] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, A. Tramontano, Critical assessment of methods of protein structure prediction (CASP) – round X, Proteins 82 (Suppl2) (2013) 1–6.
[12] A.J. Harvey, M. Hrmova, R. De Gori, J.N. Varghese, G.B. Fincher, Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases, Proteins 41 (2000) 257–269.
[13] E.F. Neufeld, J. Muenzer, The mucopolysaccharidoses, in: C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler, et al. (Eds.), The Metabolic and Molecular Bases of Inherited Disease, vol. III, McGraw-Hill, Medical Publishing Division, 2001, p. 3421.
[14] G. Parenti, Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics, EMBO Mol. Med. 1 (2009) 268–279.
[15] E.H. Schuchman, R.J. Desnick, Mucopolysaccharidosis type I subtypes presence of immunologically cross-reactive material and in vitro enhancement of the residual alpha-L-iduronidase activities, J. Clin. Invest. 81 (1988) 98–105.
[16] G.M. Pastores, P. Arn, M. Beck, J.T.R. Clarke, N. Guffon, P. Kaplan, et al., The MPS I registry: design, methodology, and early findings of a global disease registry

for monitoring patients with Mucopolysaccharidosis Type I, Mol. Genet. Metab. 91 (2007) 37–47.
[17] B.P. Rempel, L.A. Clarke, S.G. Withers, A homology model for human alpha-L-iduronidase: insights into human disease, Mol. Genet. Metab. 85 (2005) 28–37.
[18] S.S. Chandar, K. Mahalingam, Mucopolysaccharidosis type I: homology modeling and docking analysis of the lysosomal enzyme, human α-L-iduronidase, Afr. J. Pharm. Pharmacol. 6 (2012) 2027.
[19] H. Bie, J. Yin, X. He, A.R. Kermode, E.D. Goddard-Borger, S.G. Withers, et al., Insights into mucopolysaccharidosis I from the structure and action of alpha-L-iduronidase, Nat. Chem. Biol. 9 (2013) 739–745.
[20] U. Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[21] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics 9 (2008) 40.
[22] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.
[23] R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, J.M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR 8 (1996) 477–486.
[24] L.A. Kelley, M.J. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server, Nat. Protoc. 4 (2009) 363–371.
[25] J.U. Bowie, R. Luthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164–170.
[26] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85.
[27] L.J. McGuffin, The ModFOLD server for the quality assessment of protein structural models, Bioinformatics (Oxford, England) 24 (2008) 586–587.
[28] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854.
[29] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[30] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, 27-8.
[31] W.L. DeLano, S. Bromberg, PyMOL User's Guide, DeLano Scientific LLC, San Francisco, 2004.
[32] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., UCSF Chimera – a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[33] N.J. Terlato, G.F. Cox, Can mucopolysaccharidosis type I disease severity be predicted based on a patient's genotype? A comprehensive review of the literature, Genet. Med. 5 (2003) 286–294.
[34] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[35] C.N. Cavasotto, S.S. Phatak, Homology modeling in drug discovery: current trends and applications, Drug Discovery Today 14 (2009) 676–683.
[36] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[37] C. Chothia, A.M. Lesk, The relation between the divergence of sequence and structure in proteins, EMBO J. 5 (1986) 823–826.
[38] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins 17 (1993) 355–362.
[39] S. Della-Longa, A. Arcovito, Structural and functional insights on folate receptor alpha (FRalpha) by homology modeling, ligand docking and molecular dynamics, J. Mol. Graphics Modell. 44 (2013) 197–207.
[40] K. Kulleperuma, S.M. Smith, D. Morgan, B. Musset, J. Holyoake, N. Chakrabarti, et al., Construction and validation of a homology model of the human voltage-gated proton channel hHV1, J. Gen. Physiol. 141 (2013) 445–465.
[41] T. Kwon, A.L. Harris, A. Rossi, T.A. Bargiello, Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics, J. Gen. Physiol. 138 (2011) 475–493.
[42] C.B. Platania, S. Salomone, G.M. Leggio, F. Drago, C. Bucolo, Homology modeling of dopamine D2 and D3 receptors: molecular dynamics refinement and docking evaluation, PLoS ONE 7 (2012) e44316.
[43] A. Raval, S. Piana, M.P. Eastwood, R.O. Dror, D.E. Shaw, Refinement of protein structure homology models via long, all-atom molecular dynamics simulations, Proteins 80 (2012) 2071–2079.
[44] N. Maita, T. Tsukimura, T. Taniguchi, S. Saito, K. Ohno, H. Taniguchi, et al., Human alpha-L-iduronidase uses its own N-glycan as a substrate-binding and catalytic module, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 14628–14633.
[45] Z. Banecka-Majkutewicz, J. Jakobkiewicz-Banecka, M. Gabig-Ciminska, A. Wegrzyn, G. Wegrzyn, Putative biological mechanisms of efficiency of substrate reduction therapies for mucopolysaccharidoses, Arch. Immunol. Ther. Exp. (Warsz) 60 (2012) 461–468.
[46] J. Jakobkiewicz-Banecka, E. Piotrowska, M. Gabig-Ciminska, E. Borysiewicz, M. Slominska-Wojewodzka, M. Narajczyk, et al., Substrate reduction therapies for mucopolysaccharidoses, Curr. Pharm. Biotechnol. 12 (2011) 1860–1865.
[47] S. Mizumoto, S. Ikegawa, K. Sugahara, Human genetic disorders caused by mutations in genes encoding biosynthetic enzymes for sulfated glycosaminoglycans, J. Biol. Chem. 288 (2013) 10953–10961.