This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952110

MARILIA deliverable: *Report on MD trajectories*
Deliverable number: *D3.1*

MARILIA			
MARA-BASED INDUSTRIAL LOW-COST IDENTIFICATION ASSAYS			
Project nr:	952110	Call reference:	H2020-EIC-FETPROACT-2019
Start date:	September 1st, 2020.	Duration:	30 months

Deliverable identification			
Leading beneficiary:	*UZ*	Planned delivery date:	*M18*
Related WP:	*WP3*	Actual delivery date:	*28.2.2022.*
Dissemination level:	*Public*		

Contributors	
Beneficiary name	Contributor(s)' name(s)
University of Zagreb	Branimir Bertoša
	Antun Barišić
	Sanja Škulj
	Zoe Jelić Matošević

Deliverable Reviewers		
Version	Reviewer	Date
1.0	Ivan Barišić	28.2.2022

Table of content

1. Context and objectives

The first-time synthesis and purification of peptide oligonucleotide conjugates (POCs) and recombinant proteins requires significant resources (material and time) until a laboratory protocol is established. The unexpected folding pathways of biomolecules tertiary structures can lead to non-functional analytic and catalytic components. The fusion of proteins to other peptides, proteins and oligonucleotides, can significantly change their structure. Thus, *in silico* studies of the 3D structures of the artificial peroxidases are highly beneficial to reduce tedious experimental evaluations. Computational simulations allow insight into molecular properties of the studied systems, investigation of their structural and dynamical properties.

Studying systems in this way can help to avoid potential traps or explain observed phenomena, thus reducing the workload concerning empirical approaches. During the first project period, we worked closely with the experimentalists from AIT and RBI by simulating the horseradish peroxidase in different forms and with varying conditions, trying to predict and explain the outcomes of experiments. The details of the computational simulations and the results of their analyses are presented in this report.

2. Description of the performed tasks and obtained results

2.1 Methodology

2.1.1 System preparation

Starting from the available crystal structure of horseradish peroxidase (HRP) C1A (PDB code: 1H5A)[1], five main forms of enzyme were built *in silico*:

1. Native horseradish peroxidase (HRP)

2. Structure with six mutations introduced by Martell and coworkers, but without split indroduced (mHRP) [2]

3. Split structure with six mutations introduced by Martell and coworkers (sHRP).[2]

4. Separated subunit A of sHRP - sHRP-A (residues 1-213)

5. Separated subunit B of sHRP - sHRP-B (residues 214-308)

Main structural motifs, mutation sites of mHRP, sHRP, sHRP-A and sHRP-B as well as amino acids where the split occurs in the sHRP protein are shown in Figure 1 with aligned primary sequences. Starting from five *in silico* prepared structures, systems that were subjected to molecular dynamics (MD) simulations were prepared. The split form (sHRP) was generated from the X-ray structure by introducing the cut-site after G213 and six mutations (T21I, P78S, R93G, N175S, N255D, and L299R) identified by Martell *et al*.[2] Missing residues 307-308 in the X-ray structure were generated by the solution builder module of CHARMM-GUI.[3–5] Hydrogen atoms were added using CHARMM-GUI in a way that the side chains of all arginines and lysines were positively charged, histidines (with hydrogen on epsilon nitrogen – HIE) and cysteines were in neutral form, while side chains of glutamates and aspartates were deprotonated and negatively charged. Four disulfide bonds (Cys11-Cys91, Cys44-Cys49, Cys177-Cys209, and Cys97-Cys301), as well as the bond between His170 and Fe^{2+} from the heme cofactor, were defined in structures in which they were present. The CHARMM36m force field was used for parametrization of protein structure, glycans, heme and ions.[6] Solvatation effects were simulated using a cubic box filled with the TIP3P model of water molecules, with a distance of at least 20 Å between the solute and the edge of the box . Chloride ions were added to neutralize the systems. Molecular dynamics (MD) simulations were performed using periodic boundary condition (PBC). The size of a rectangular box depends on the system, but it was on average around 10 nm x 10 nm x 10 nm.

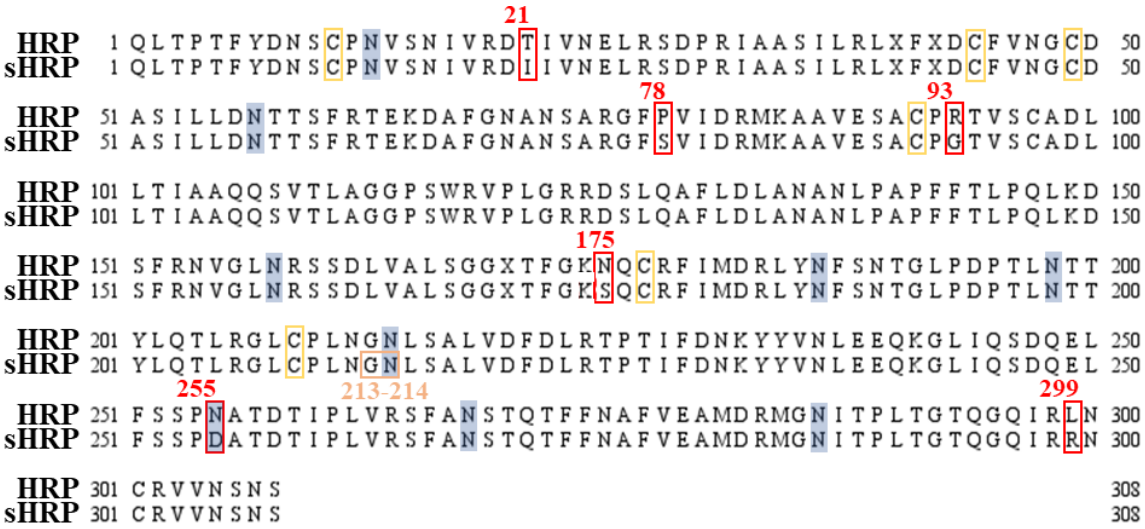


Figure 1. Alignment of the HRP and mutated mHRP, sHRP, sHRP-A and sHRP-B structures. Positions of mutated amino acids are shown in red squares, positions of glycosylated sites – asparagine amino acids (which follow pattern Asn–X–Thr/Ser) are shaded in blue squares, cysteine amino acids which form disulfide bridges are shown in yellow squares and the cut-site of sHRP is shown in an orange square.

2.1.2 Molecular dynamic (MD) simulations

Prior to MD simulations, all systems were energy minimized (geometry optimized) in 1000 cycles and then equilibrated in the equilibration process provided by the CHARMM-GUI solution builder module and different restraints were subsequently applied.[3] After energy minimization, systems were equilibrated for 10 ns. The production phase of molecular dynamics (MD) simulations lasted for 500 ns for each system with a time step of 2 fs and the LINCS algorithm to keep all bonds constrained.[7] MD simulations were performed in the isobaric-isothermal ensemble (NPT) employing periodic boundary conditions (PBC) in all directions at T = 300 K, which was maintained via a Nosé-Hoover thermostat[8] with a coupling constant of 1.0 ps^{-1}. Pressure was set to 1.013 bar and was controlled with a semi-isotropic Parrinello-Rahman barostat[9] with a time constant for pressure coupling of 5 ps^{-1}. Long range electrostatics were calculated using the particle-mesh Ewald (PME) method[10] with real space Coulomb interactions cut off at 1.2 nm using a Fourier spacing of 0.12 nm and the Verlet cut-off scheme. All simulations were run using the GROMACS 2018.6 software package[11]. Analyses of trajectories were performed using Gromacs analysis tools and the VMD program[12]. Electrostatic potential was calculated using the PMEpot plugin for VMD for each frame and then averaged over the entire trajectory. It was calculated using all atoms of the system with a three-dimensional grid (48 x 48 x 48) and Ewald factor of 0.25 at T = 300 K.

The main goal of the conducted simulations was to study the following:

1. Effect of glycosylation on HRP and sHRP structures
2. Effect of cysteine bridge on sHRP structure
3. Influence of short and long His-tag presence on: mHRP, s-HRP, sHRP-A and sHRP-B
4. Effect of the mutations present in mHRP and sHRP structures
5. Availability of Lys for click reactions
6. Simulations without the heme cofactor

2.2 Results

2.2.1 Effect of glycosylation onHRP and sHRP structures

Three different glycolysation patterns of HRP and sHRP were studied. According to the number of mannoses, we studied systems with three glycosylation branching types: $Man_8GlcNAc_2$, $Man_{16}GlcNAc_2$, $Man_{20}GlcNAc_2$ (Figure 2) and compared them to non-glycosylated systems. Since the exact glycosylation pattern of HRP in the expression systems used by AIT was not availble, every one of nine (HRP) /eight (sHRP) corresponding asparagine amino acids were glycosylated with the same branching type.

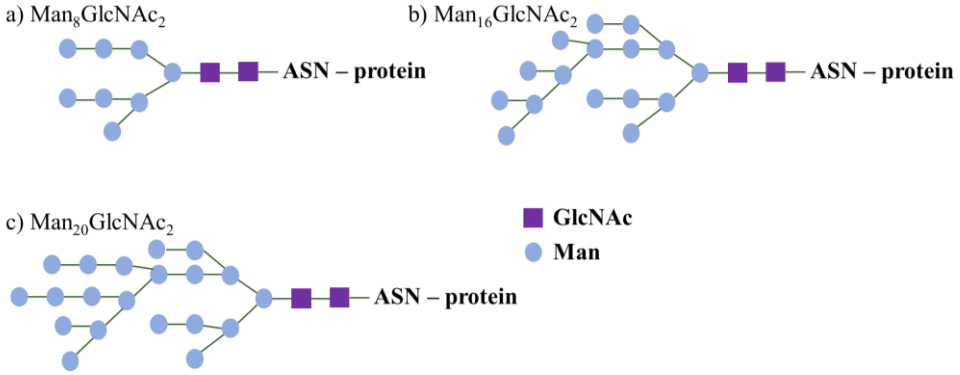


Figure 2. Glycosylation branching types: a) $Man_8GlcNAc_2$, b) $Man_{16}GlcNAc_2$ and c) $Man_{20}GlcNAc_2$. Man stands for mannose, and GlcNAc for *N*-acetylglucosamine.

In order to examine the effects of *N*-glycosylation, both HRP and sHRP structures were prepared without *N*-glycosylation and with each of the aforementioned glycosylation shown in Figure 10.[13] All asparagine amino acids which follow pattern Asn–X–Thr/Ser (X is any amino acid residue other than proline or aspartic acid) were *N*-glycolisated – 9 ASN amino acids in HRP (number of Asn residue: 13, 57, 158, 186, 198, 214, 255, 268, 286) and 8 Asn amino acids in mHRP and msHRP (number of Asn residue: 13, 57, 158, 186, 198, 214, 268, 286). In total, eight systems were prepared (four with HRP and four with sHRP) and MD simulations were run for 500 ns (Table 1), for a total of 4 μs of simulation.

Table 1. Systems prepared for MD simulation.

System	Glycan branching type			
HRP	NO glycan	$Man_8GlcNAc_2$	$Man_{16}GlcNAc_2$	$Man_{20}GlcNAc_2$
sHRP	NO glycan	$Man_8GlcNAc_2$	$Man_{16}GlcNAc_2$	$Man_{20}GlcNAc_2$

2.2.1.1 Effect of *N*-glycosylation on HRP protein structural properties

Protein/glycoprotein size was monitored during the simulations via radius of gyration (R_g) (Table 2 and Figure 3). The most branched glycoprotein $Man_{20}GlcNAc_2$ is almost one half larger, with an R_g approximately 2.9 nm, than the non-glycosylated protein with and R_g approximately 2.0 nm (Figure 4 and Table 2).

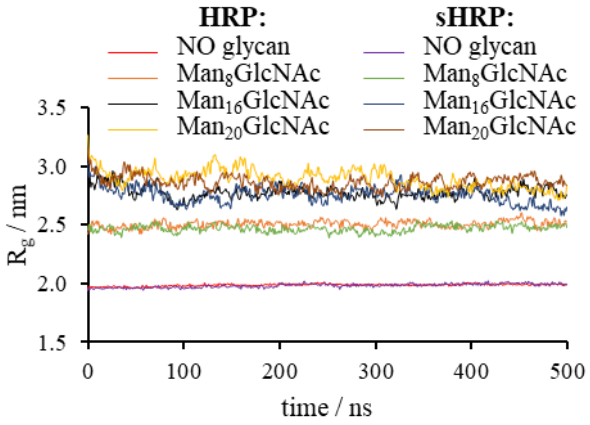


Figure 3. Radius of gyration (R_g) of all eight systems in nm.

The glycoprotein with glycosylation type $Man_{20}GlcNAc_2$, rather than being longer, is more branched that the other two types of glycosylation. Therefore the average R_g of $Man_{16}GlcNAc_2$ (approximately 2.8 nm) and $Man_{20}GlcNAc_2$ (approximately 2.9 nm) glycoproteins are comparable, while for $Man_8GlcNAc_2$ R_g is smaller, approximately 2.5 nm (Table 2 and Figure 3).

Table 2. Averaged values of radius of gyration (R_g) of protein during MD simulations.

System		NO glycan	$Man_8GlcNAc_2$	$Man_{16}GlcNAc_2$	$Man_{20}GlcNAc_2$
HRP	R_g / nm	1.98±0.01	2.51±0.03	2.77±0.06	2.89±0.08
sHRP	R_g / nm	1.98±0.01	2.46±0.03	2.75±0.07	2.86±0.06

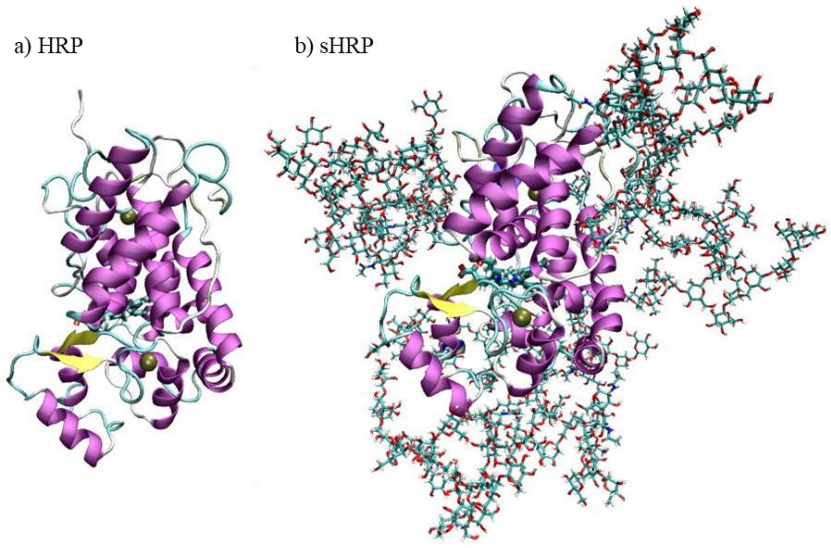


Figure 4. Snapshots of HRP protein taken after 500 ns of MD simulation of: a) HRP and b) sHRP with $Man_{20}GlcNAc_2$ type of branching.

Root mean square deviations (RMSD) of protein backbone during MD simulations show that all systems are equilibrated and stable after 500 ns (Figure 5) with no significant changes in the overall protein structure. In both cases, HRP and sHRP, it was observed that glycan presence does not significantly influence the protein tertiary structure, which is in accordance with literature data according to which *N*-glycosylation does not induce significant changes in the protein tertiary structure. Moreover, the secondary structure of the protein is preserved and dynamically stable regardless of glycosylation (Figure 6).

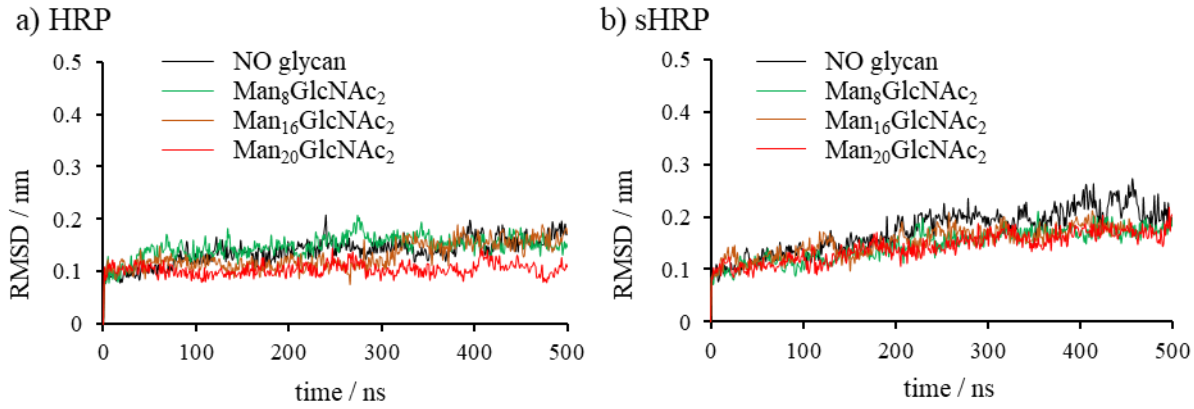


Figure 5. RMSD values of: a) HRP protein and b) sHRP protein. Backbone carbon atoms – (Cα) were considered in calculations.

In order to analyze the structural properties of the investigated systems in more detail, Principal Component Analysis (PCA) of the C_α protein backbone atoms during MD simulations was conducted (Figure 7). From movement along the first principal component (PC1), one can observe that the structure of HRP remains almost fully preserved, regardless of glycan presence, with a total range of about 2 nm along PC1. On the other hand, the volume occupied by PCA projections in the space spanned by the first two principal components (PC) is smallest in systems prepared with the $Man_{20}GlcNAc_2$ branching type for both HRP and sHRP. However, comparison of PCA results for HRP and sHRP shows that structural changes are more pronounced for the split protein (sHRP) (Figure 7) where the first two PCs span a larger area compared to HRP in all systems. The structural changes that cause this in sHRP are caused by the split in the polypeptide chain between residues 213 and 214. Due to the introduced split, sHRP possess two additional fluctuating terminuses compared to HRP. In conclusion, PCA confirmed the observation that *N*-glycosylation does not have a significant influence on HRP nor sHRP structural properties.

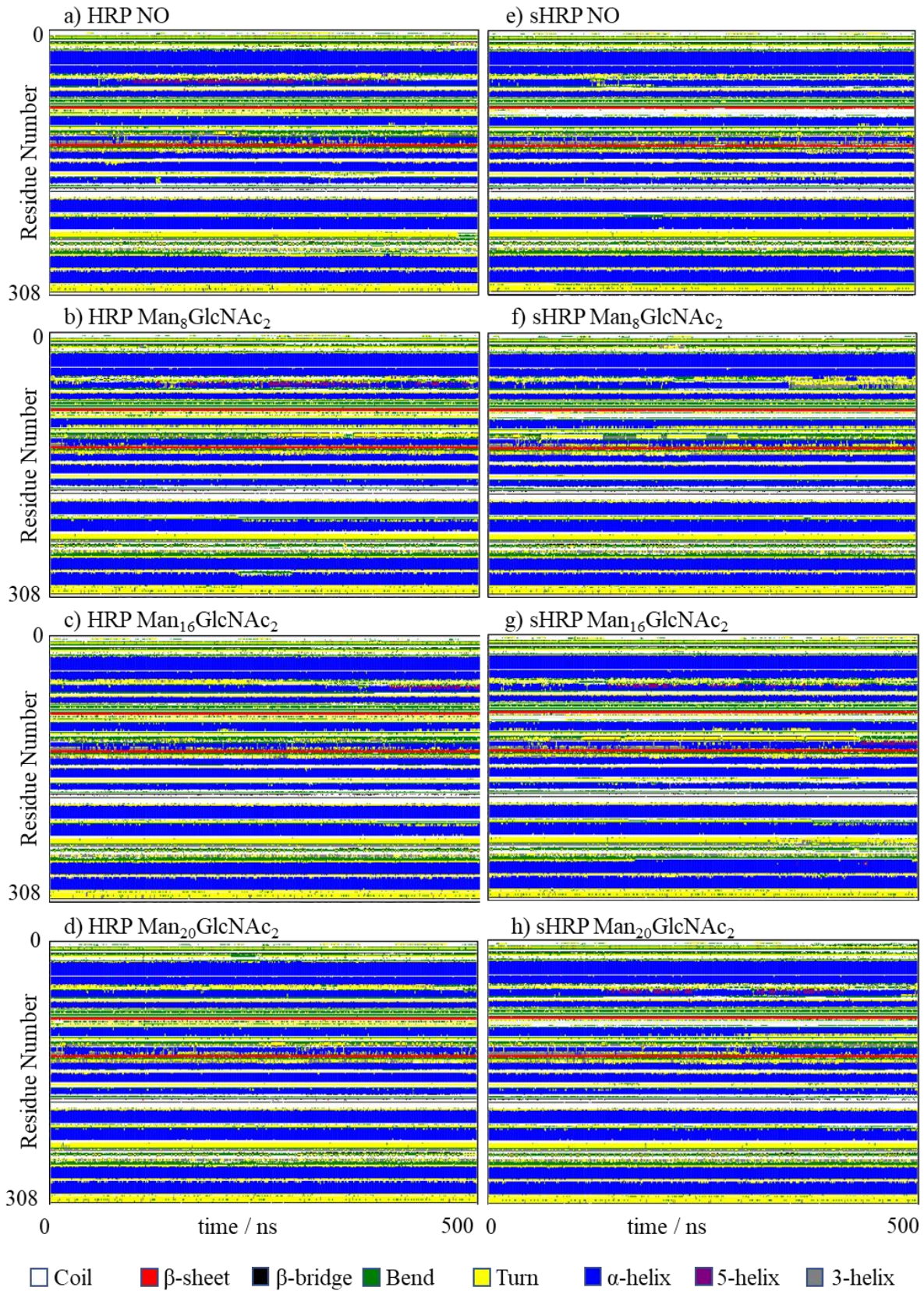


Figure 6. Secondary structure evolution of the HRP and sHRP structures with and without glycans with the $Man_{20}GlcNAc_2$, $Man_{16}GlcNAc_2$ and $Man_8GlcNAc_2$ branching types. Secondary structure of HRP and sHRP is preserved in time for all systems.

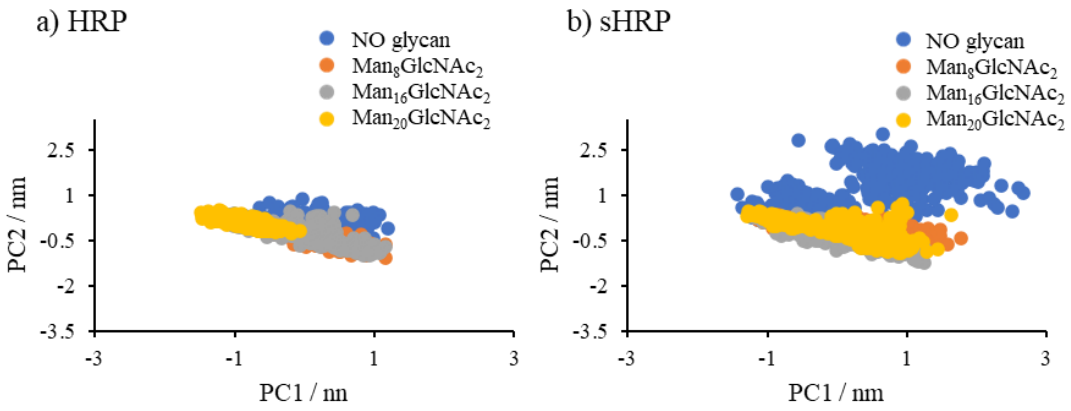


Figure 7. PCA analysis – 2D projection of first two eigenvectors PC1 and PC2 of a) HRP and b) sHRP proteins.

2.2.1.2 Glycan structural and dynamical properties

In glycoproteins, glycans occupy a large part of the space around the protein since they are mobile and fluctuate more than the protein itself (Figure 4 and Figure 8). Even the glycan residue closest to the protein, GlcNAc, fluctuates on average more than the protein - fluctuations of HRP's GlcNAc are 0.09-0.35 nm, and in sHRP's they are 0.08-0.27 nm (Figure 8). On average, oscillations of glycoprotein ends in $Man_{20}GlcNAc_2$ and $Man_{16}GlcNAc_2$ are similar, while $Man_8GlcNAc_2$ ends fluctuate less. This is in accordance with previous results regarding the radius of gyration, which showed that R_g is similar for $Man_{20}GlcNAc_2$ and $Man_{16}GlcNAc_2$ glycoproteins.

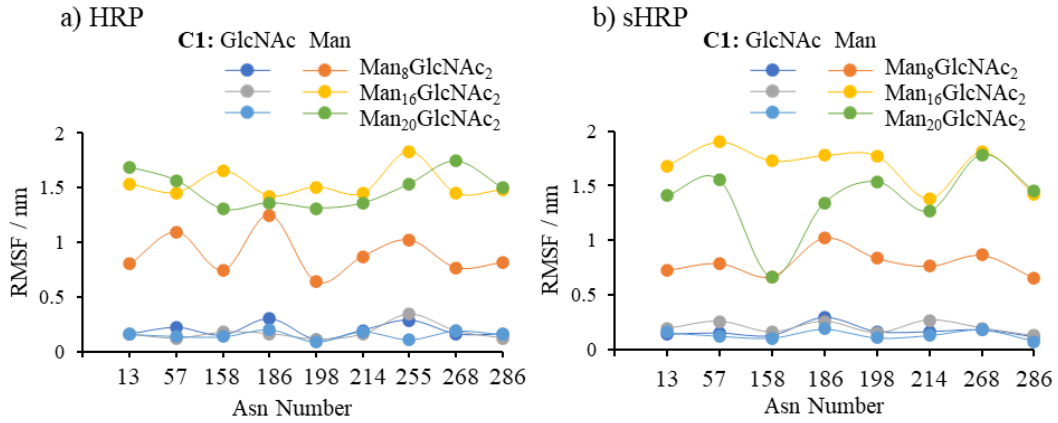


Figure 8. RMSF values of glycans for: a) HRP and b) sHRP protein. The figure shows RMSF of the C1 atom of the *N*-acetylglucosamine (GlcNAc) closest to the protein and the last mannose (Man8, Man16 and Man20) connected to Asn and C1 atom.

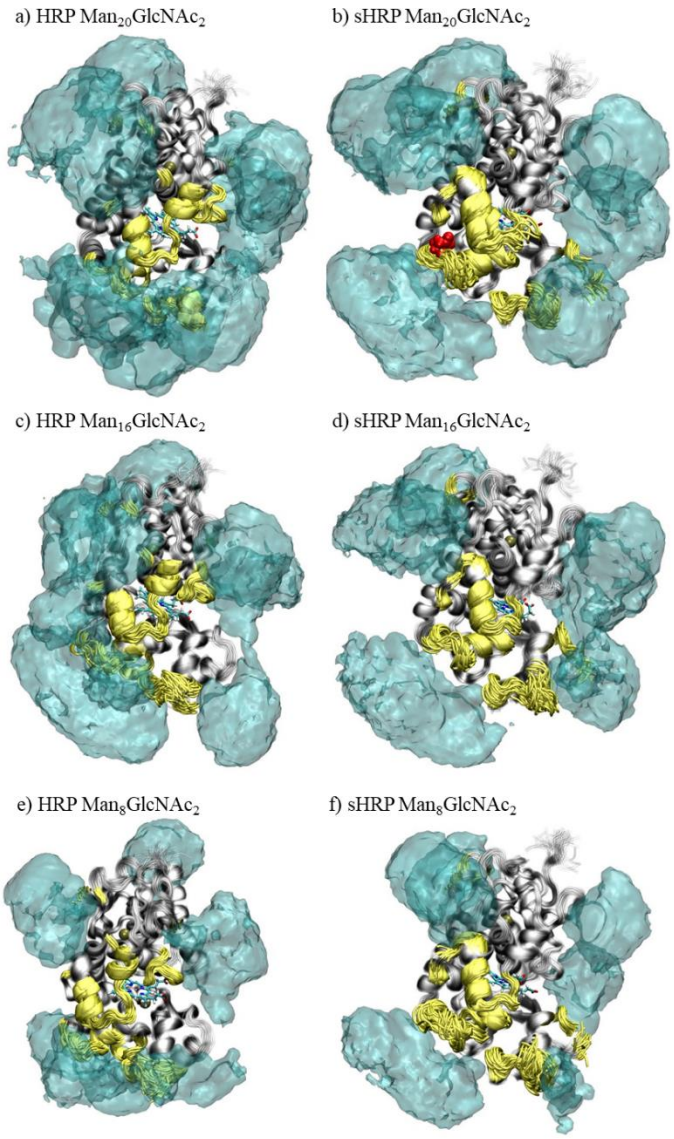


Figure 9. Snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. The average volume map for glycans (isovalue 0.15) is shown in cyan. The mutated Asp255 residue is shown in red.In yellow are colored residues which fluctuations decreased the most in glycoproteins: $Man_{20}GlcNAc_2$, $Man_{16}GlcNAc_2$ and $Man_8GlcNAc_2$ (fluctuation decreased by > 0.03 nm in $Man_{20}GlcNAc_2$).

From the average volume of glycans around the protein presented in Figure 9 it can be concluded that certain parts of the protein surface (those closer to the glycosylated asparagines) are surrounded and to some extent “protected” by the glycans, while the rest of the protein is exposed and more accessible to solvent and substrates. Generally, the peripheral parts of the protein are protected by glycans while the central region (containing the enzyme active site) is exposed directly to solvent (water) and substrate (Figure 9).

This is interesting because the heme cofactor which is crucial for enzyme activity is in this central region. Because of the mutation at site 255 (Asn residue in HRP), sHRP has one less glycosylation site and a smaller glycan volume around the periphery of the protein than HRP. The calculated electrostatic potential shows that glycans are slightly more negative than the protein (Figure 10b). It is especially interesting that the presence of glycans induces polarization of the electrostatic potential of the protein and heme cofactor (Figure 10 and Figure 11).

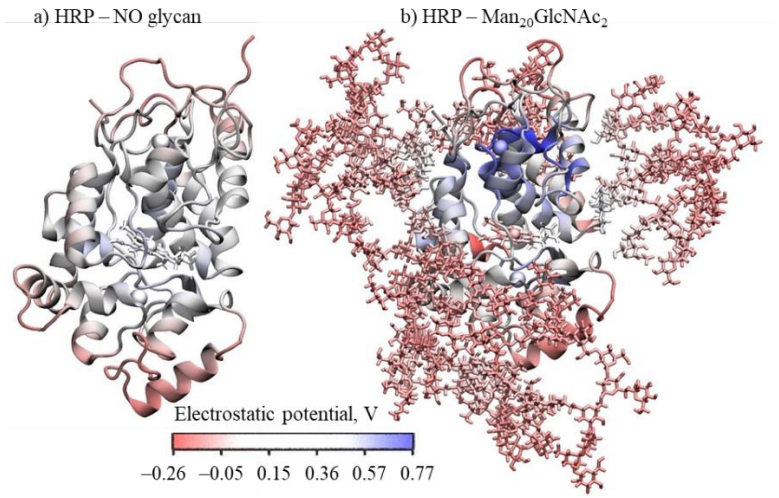


Figure 10. Average electrostatic potential of the (glycol)protein during molecular dynamics trajectory for: a) HRP, b) HRP with $Man_{20}GlcNAc_2$ glycosylation. Total variation of potential is 1.03 V.

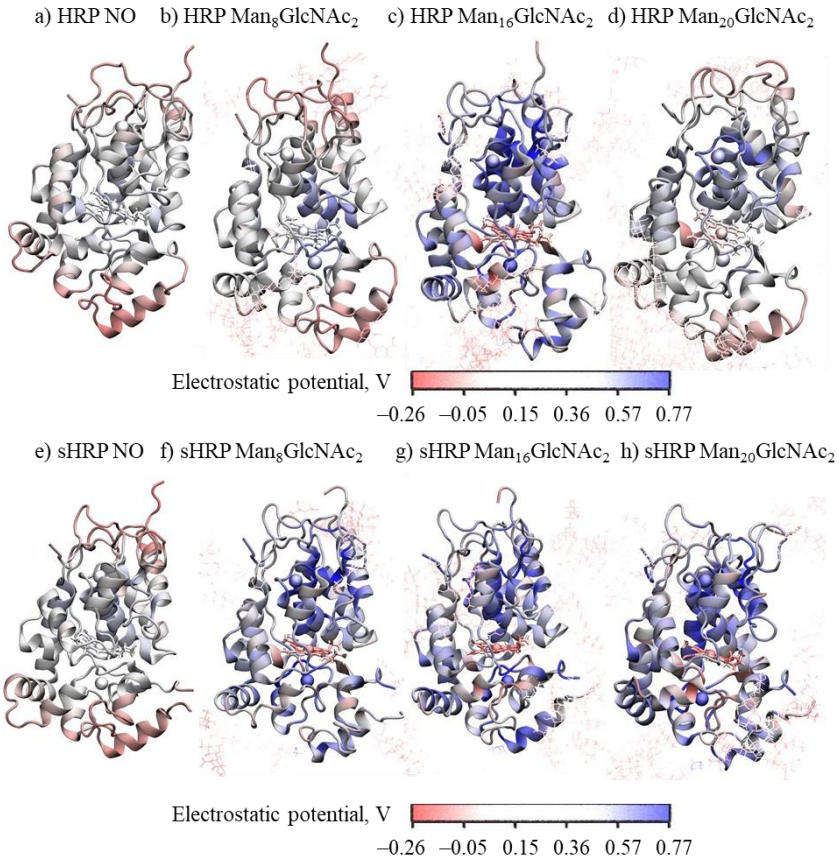


Figure 11. Average electrostatic potential along molecular dynamic trajectories of different systems. Total variation of potential is 1.03 V.

2.2.1.3 Effect of *N*-glycosylation on HRP protein dynamical properties

The influence of glycosylation on protein dynamics was studied using analysis of fluctuations during the trajectories of simulated systems (Figure 12). In general, glycosylation significantly reduces fluctuations along the whole length of the protein (Figure 12, Figures 13 and Figure 14). In addition, for sHRP large fluctuations

were observed in the cut-site region. Since sHRP possesses a cut-site between residues 213 and 214, there is a high peak in fluctuations of the sHRP protein around the mentioned amino acids (green square in Figure 12b) which is not present in HRP (Figure 12a). This peak is the consequence of an additional C- and N-terminus in sHRP because of the cut-site.

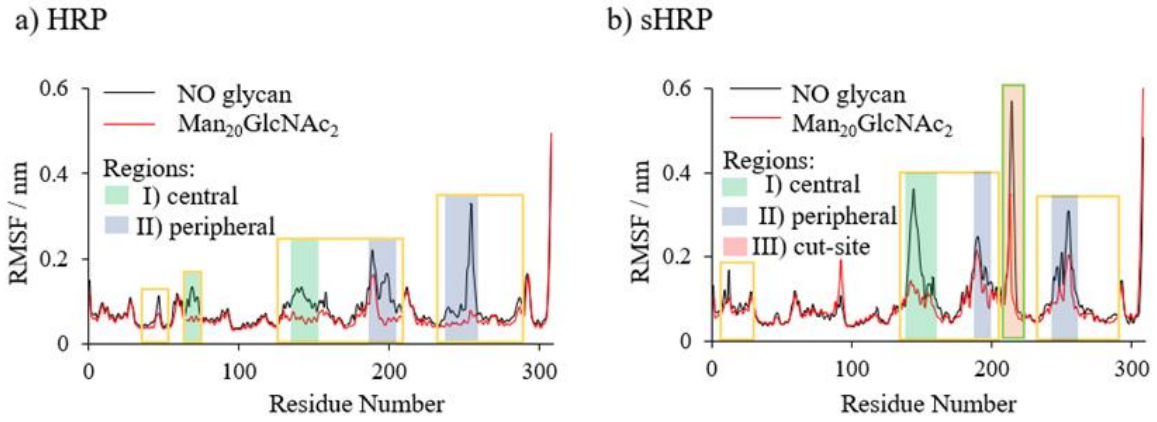


Figure 12. RMSF values of: a) HRP protein and b) sHRP protein. Fragments of the protein where fluctuations are decreased by introducing glycosylation are shown in yellow squares. Backbone carbon atoms – Cα atom of every amino acid from the protein backbone were considered in the calculation. Amino acids (fluctuation decreased by > 0.03 nm) close to each other in tertiary structure forming three regions – I) central, II) peripheral and III) cut-site region are shown shaded in green, blue and red, respectively.

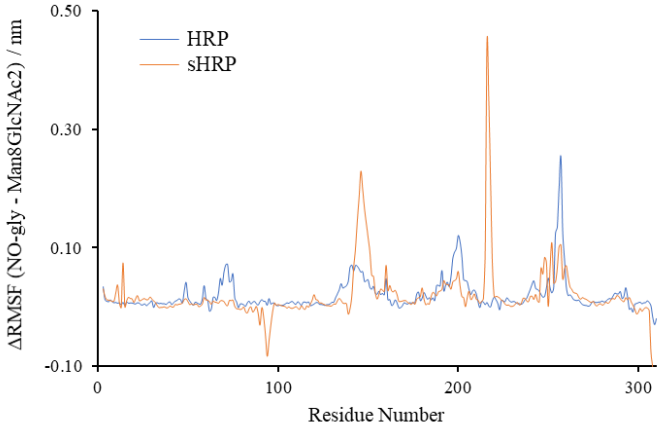


Figure 13. Subtraction of values of fluctuations of non-glycosylated protein and protein with $Man_{20}GlcNAc_2$ type of branching for HRP and sHRP. In HRP, fluctuations are always decreased in the $Man_{20}GlcNAc_2$ glycoprotein compared to the non-glycosylated protein (except for residues 295 and 306-308 where fluctuations are slightly increased). In sHRP, fluctuations are almost always decreased, except in residues 80-81, 84, 87-88, 90-95, 137 and 304-308.

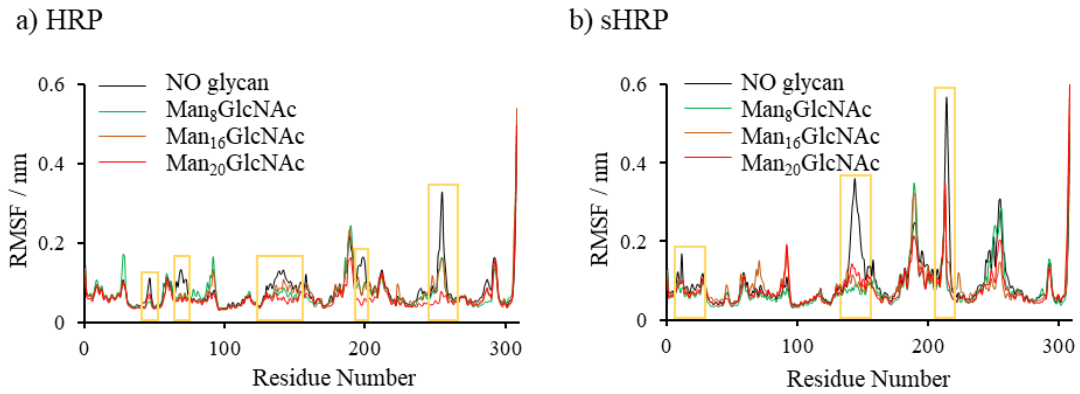


Figure 14. RMSF values of a) HRP protein and b) sHRP protein. Fragments of the protein where fluctuations are decreased by introducing glycosylation are shown in yellow squares. Backbone carbon atoms - Cα of every amino acid from the protein backbone were considered in the calculation.

The comparison of fluctuations in HRP and sHRP shows how important just one glycosylation site can be for protein dynamics. As already mentioned, HRP contains nine and sHRP eight glycosylated asparagine amino acids since in sHRP, Asn255 is mutated to aspartate and it is not glycosylated. Due to this difference, the fluctuations of glycosylated HRP ($Man_{20}GlcNAc_2$) are on average 1.8 $\times 10^{-2}$ nm smaller than in the case of glycosylated sHRP 2.1 $\times 10^{-2}$ nm (Table 3).

Table 3. Average RMSF (without end residues 300-307) and standard deviation for HRP and sHRP for protein without glycan and with all glycan branching types.

System		NO glycan	$Man_8GlcNAc_2$	$Man_{16}GlcNAc_2$	$Man_{20}GlcNAc_2$
HRP	$RMSF_{av}$ $\times 10^{-2}$ / nm	7.7±3.7	6.6±3.1	6.9±2.9	5.7±2.2
sHRP	$RMSF_{av}$ $\times 10^{-2}$ / nm	9.5±7.0	7.6±5.0	7.9±4.1	7.4±3.9

The aforementioned asparagine (Asn255) is the amino acid with the largest average fluctuation during the simulations of non-glycosylated HRP. It also shows the largest decrease in fluctuations upon glycosylation in HRP simulations – its fluctuation is decreased by 0.25 nm in the $Man_{20}GlcNAc_2$ HRP simulation as compared to the non-glycosylated HRP simulation (from 0.33 nm for non-glycosylated to 0.07 nm for HRP with $Man_{20}GlcNAc_2$ branching type). On the other hand, the mutated average fluctuation for Asp255 in sHRP with the $Man_{20}GlcNAc_2$ branching type decreases by only 0.11 nm (from 0.31 nm for non-glycosylated to 0.20 nm for $Man_{20}GlcNAc_2$ branching type). In the case of sHRP, beside the C- and N-termini, the largest decrease of average fluctuation (0.23 nm) due to glycosylation was observed for Thr144 which had average fluctuations of 0.36 nm in simulations of non-glycosylated sHRP (0.36 nm), which decreased by 0.23 nm in simulations of sHRP with the $Man_{20}GlcNAc_2$ branching type. The largest differences in average fluctuations of amino acids between non-glycosylated and glycosylated proteins are marked with yellow squares in Figure 12 and Figure 14. The corresponding changes in tertiary structure are presented in Figure 16 where amino acids for which the fluctuation decreased more than 0.03 nm due to glycosylation are labelled in yellow.

Interestingly, the highest decrease of fluctuations due to glycosylation was observed on one side of protein (Figure 15).

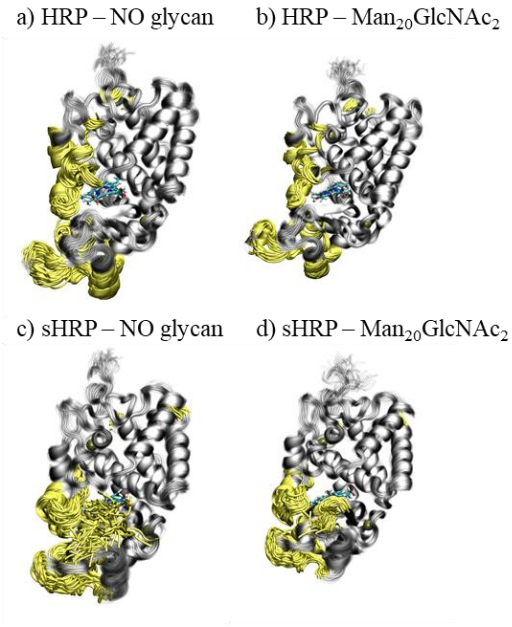


Figure 15. Side view of snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. In yellow are coloured residues which fluctuations decreased the most in glycoprotein $Man_{20}GlcNAc_2$ (fluctuation decreased by > 0.03 nm). The fluctuations mentioned in the text are mostly on one side of protein. Glycans in the glycoprotein (b and d) are not shown for clarity.

The other side is quite rigid even without glycans. Simulations show that the type of glycan branching does not significantly influence the fluctuations (Figure 14 and Table 3). The regions where the presence of glycans significantly decreases fluctuations (more than 0.03 nm) can be grouped into two/three regions within the tertiary structure of the protein:

I) central region (HRP amino acids: 66-73, 133-151, 158, sHRP amino acids: 140-151, 155-158, 160; shaded green in Figure 5 and Figure 6)

II) peripheral region (HRP amino acids: 186-189, 191-204, 238-242, 247-249, 251-258, sHRP amino acids: 189-199, 244-247, 249-261; shaded blue in Figure 5 and Figure 6)

III) cut-site region (sHRP amino acids: 213-217; shaded red in Figure 5 and Figure 6).

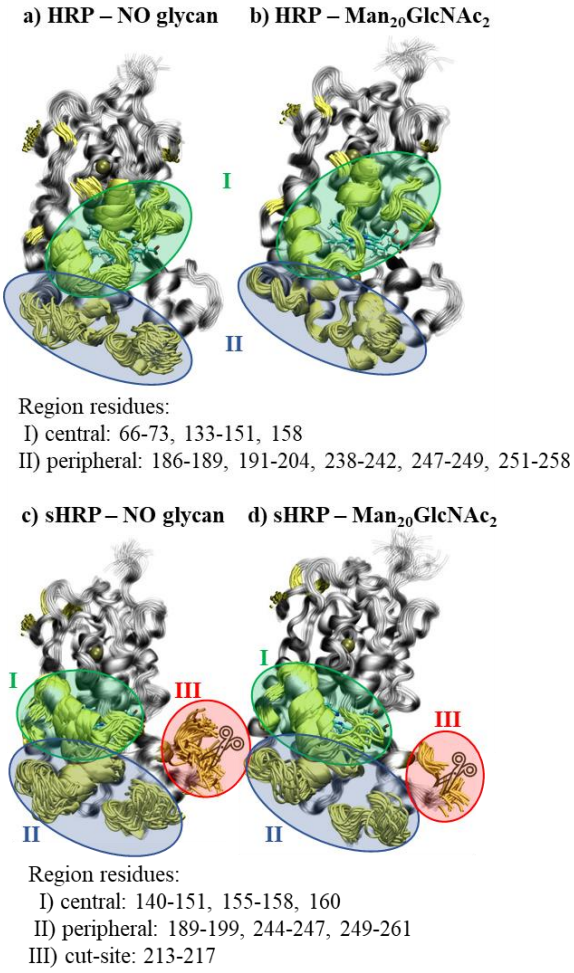


Figure 16. Snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. In yellow are colored residues which fluctuations decreased the most in glycoprotein $Man_{20}GlcNAc_2$ (fluctuation decreased by > 0.03 nm). Glycans in glycoprotein are omitted for clarity.

By visualizing the results of principal components analysis (PCA) performed on the Cα protein backbone atom coordinates of all trajectories (HRP and sHRP, non-glycosylated and all three glycan branching types), we can see that the regions with the most fluctuations along PC1 correspond well to the aforementioned central, peripheral and cut-site regions of the protein (Figure 17). Therefore, for HRP and sHRP together, the most fluctuating regions of the protein are the central region I – residues 140-155, the peripheral region II – residues 245-255 and the cut-site region III – 185-212. These regions, derived from PC1, correspond to regions determined by RMSF analysis – they differ only in these regions derived using PCA contain fewer amino acids. Central region I is close to the protein core and heme cofactor, but it has pronounced oscillations (Figure 16). Similar results are obtained when applying principal component analysis to all simulated systems individually (Figure 18).

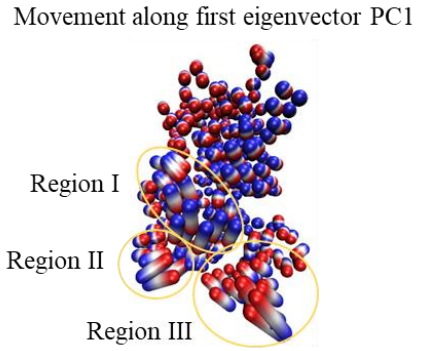


Figure 17. Movements along the first eigenvector PC1 of protein backbone carbon atoms of all systems together. Cα atoms in two structures with extreme values of PC1 are shown in red and blue, while intermediary structures are shown in between.

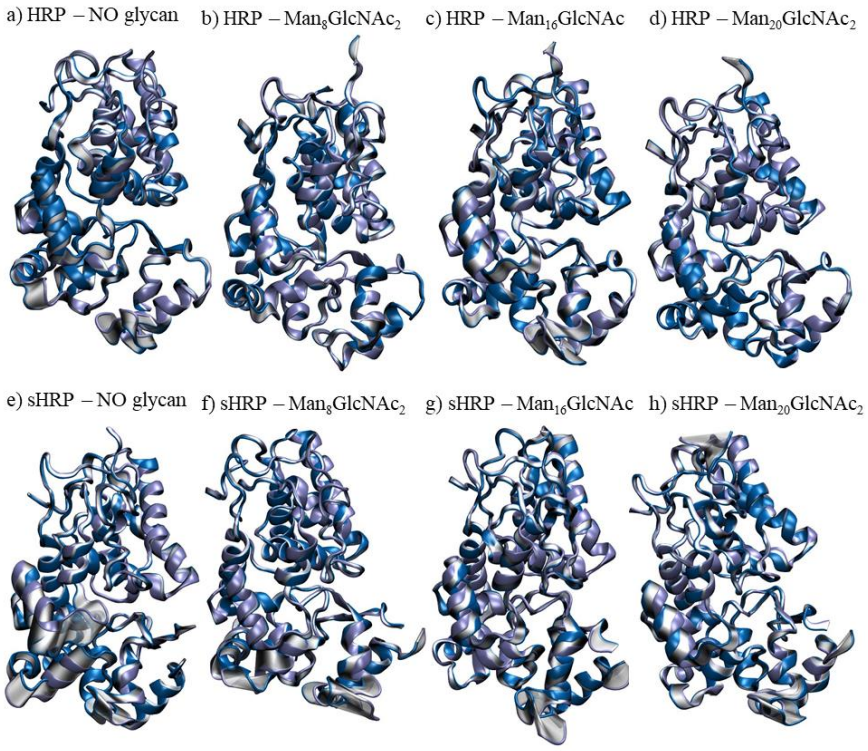


Figure 18. Movements along the first eigenvector PC1 of PCA performed for all proteins separately. Blue and ice blue are two extreme structures, the structures in-between are colored gray.

2.2.1.4 Propagated effect of *N*-glycosylation on HRP protein dynamical properties

Interestingly, the influence of glycosylation on protein flexibility is highly pronounced for some protein regions which are not in the vicinity of glycans. The observed effect of glycosylation on protein dynamics was the most pronounced for two regions, the central region I and the peripheral region II placed at the proximal site of heme, for both, HRP and sHRP (Figure 16). The cut-site region III as a region of increased fluctuability is only present in the sHRP protein. This is interesting, because the largest difference in fluctuations between non-glycosylated and glycosylated protein (yellow alpha coils in Figure 19) is not concentrated only at the protein glycosylation site regions but is spread to other areas of the protein as well, especially to central region

I (Figure 19 and Figure 9).

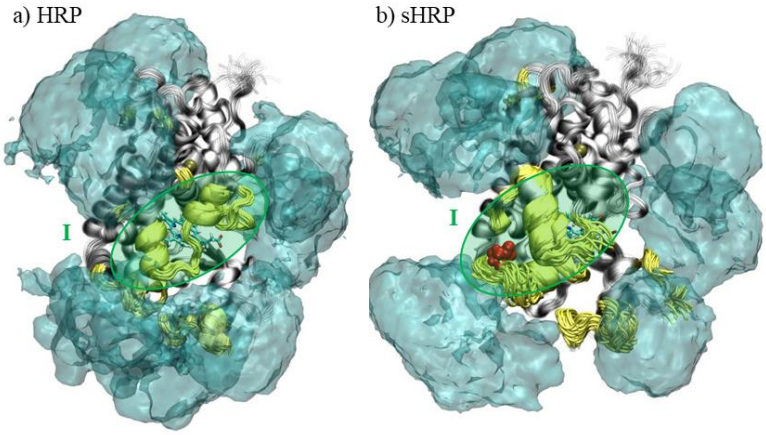


Figure 19. Snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. Aligned snapshots of (glyco)protein every 1000th frame from trajectory. In yellow are colored residues which fluctuations decreased the most in glycoprotein $Man_{20}GlcNAc_2$ (fluctuation decreased by > 0.03 nm). The average volume map for glycans (isovalue 0.15) is shown in cyan. The mutated Asp255 residue is shown in red.

This phenomenon was also observed in different types of glycoproteins and our results support these findings.[14] Peripheral region II is directly affected and protected by glycans, while central region I is not covered by glycans and is easily accessible for water or other important molecules, such as substrates. However, what is remarkable is that even though not directly protected by glycans, this region I (which is also close to the core of protein) is stabilized by glycans. This is the regions which contains the catalytic site of the protein, including the heme cofactor and two calcium ions (Ca^{2+}). Figure 20 shows the deviation of the calcium ions and heme cofactor from their initial positions during 500 ns of MD simulation for non-glycosylated and glycosylated systems. It is apparent that glycosylation stabilized the active site cofactors, possibly enhancing enzyme activity. This will be further studied using QM/MM simulations. In sHRP, this effect is not observed and oscillations of Ca^{2+} ions are not decreased as a result of glycosylation. The likely reasons for this are the six introduced mutations and one less glycosylation site (Asn255).

It is interesting that the electrostatic potential of the protein, especially the core of the protein including the heme cofactor and Ca^{2+} ions, is more polarized in glycoproteins compared to their non-glycosylated counterparts (Figure 11). The heme cofactor has a more negative and the Ca^{2+} ions a more positive electrostatic potential in glycoproteins compared to non-glycosylated proteins. This polarization of electrostatic potential in glycoproteins is more pronounced in sHRP (Figure 11 e)-h)).

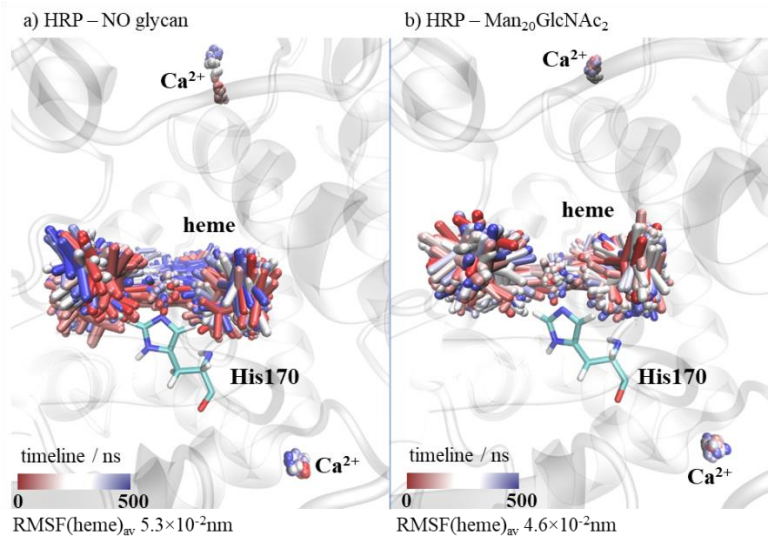


Figure 20. Snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. Shown are the heme cofactor and calcium ions for HRP: a) without glycans and b) with $Man_{20}GlcNAc_2$ branching type.

In summary, glycans influence HRP and sHRP fluctuations, but the tertiary structure of the protein is conserved even in the absence of glycans. This is in agreement with literature data which stated that HRP is still active and functional even in the absence of glycans,[15] but glycans decrease dynamic fluctuations of the protein structure[16] and the process of protein unfolding is 2-3 times faster in the absence of glycans.[17] Glycans oscillate more than the protein and in that way decrease protein fluctuations. All of the results shown here lead to the conclusion that the stability HRP and sHRP protein is increased when glycosylation is introduced.

2.2.2 Cysteine bridge Cys97-Cys301 effect on sHRP structure

2.2.2.1 Mutations of Cys97-Cys301

In an early stage of the project, one of the ideas of the whole project team was to try to replace the cysteine disulfide bridge that is in case of sHRP between subunits A and B (Cys97-Cys-301) with the strong hydrogen bond (H-bond). In order to achieve that, those two Cys were mutated to different amino acids that are capable of forming strong H-bonds. Following mutants were prepared *in silico* and subjected to MD simulations: C97S-C301S, C97D-C301N, C97N-C301D, C97D-C301R and C301R-C97D.

Simulations of each mutant were analyzed using visualization and standard analyzing tools (Figures 21-28). Analyses pointed to the C97S-C301S mutant (Figure 23) as the best candidate since a stable hydrogen bond between Asp97 and Asn301 was formed and remained stable through whole simulation (Figure 24).

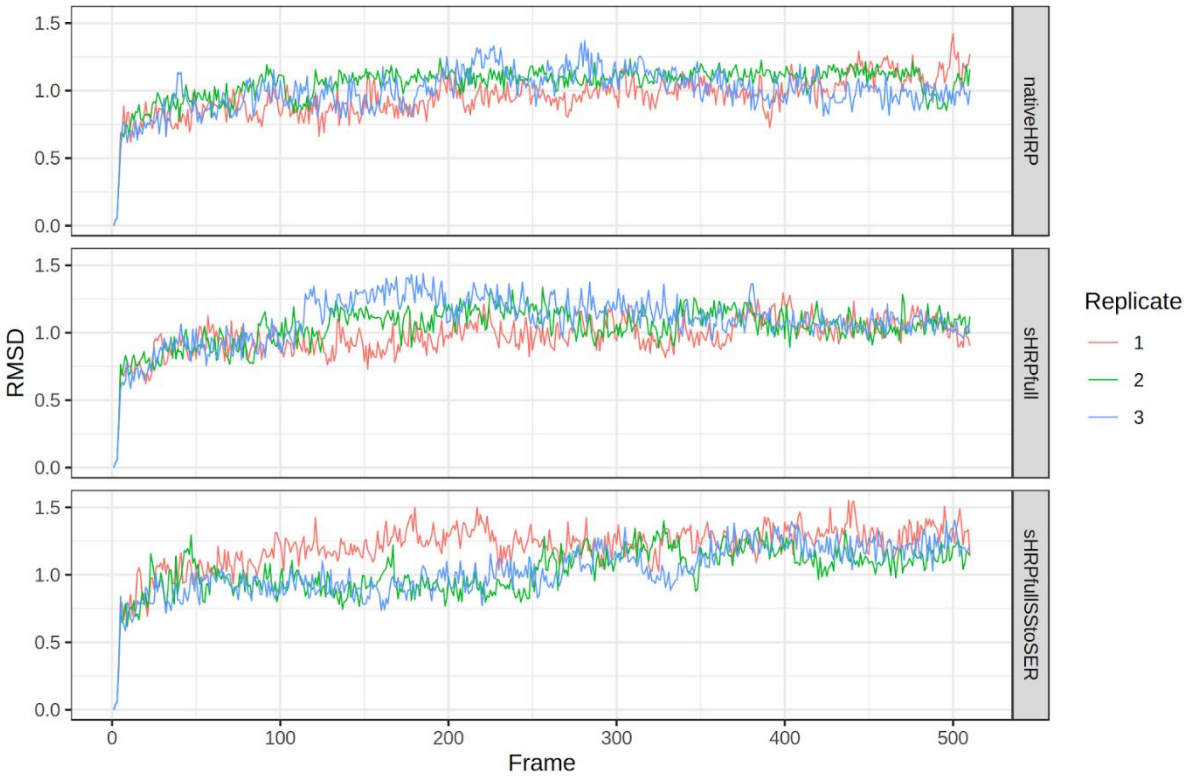
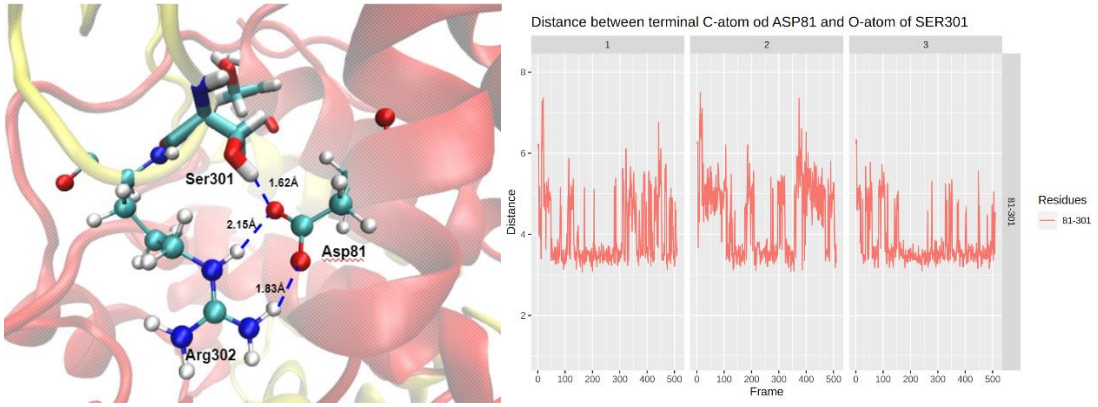


Figure 21. RMSD values of the HRP protein (up), the sHRP protein (middle) and the C97S-C301S mutant during 50 ns of MD simulation.



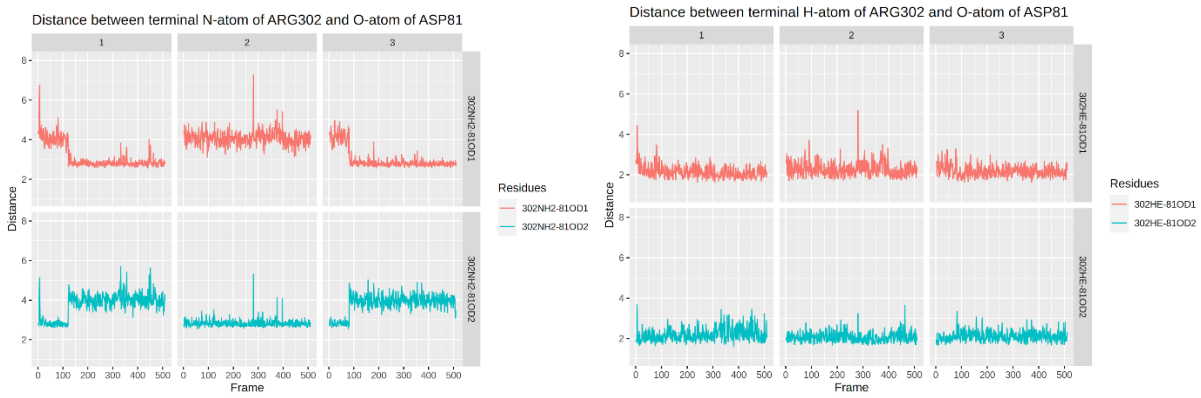


Figure 22. The hydrogen bond network Ser301 forms during simulations of the C97S-C301S mutant. Shown are the key interactions and their fluctuations during the simulation.

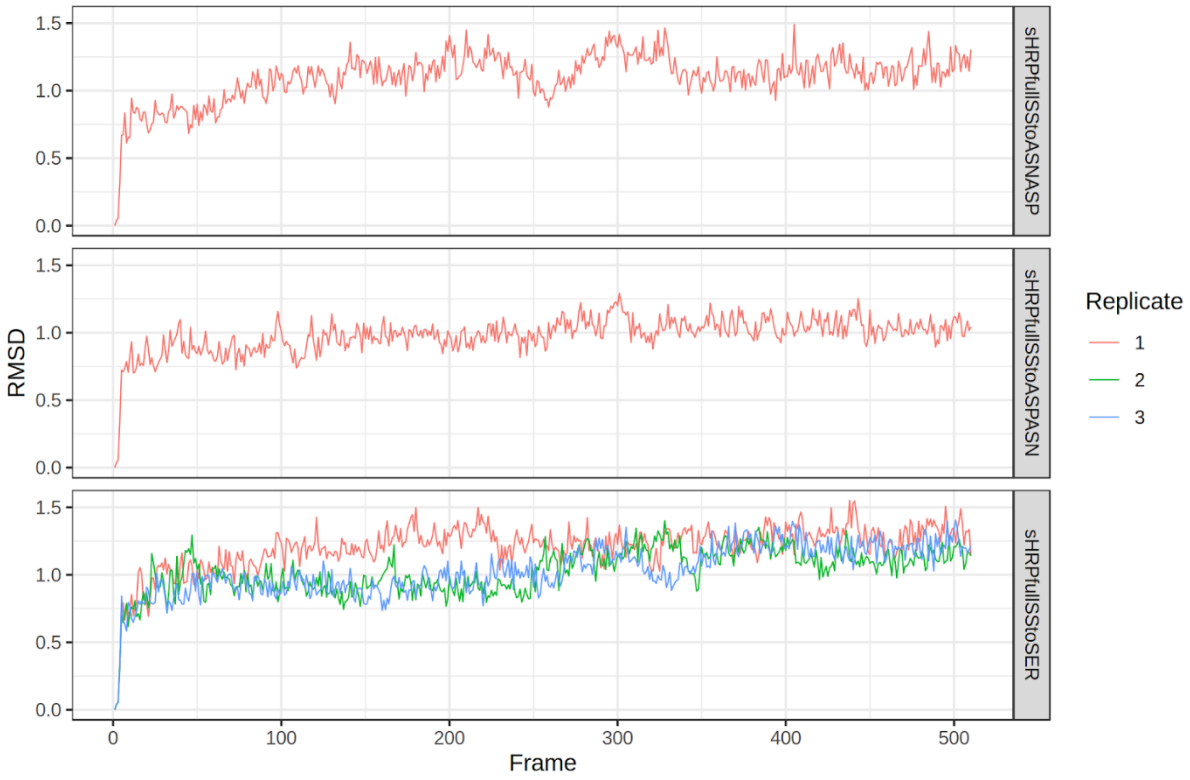
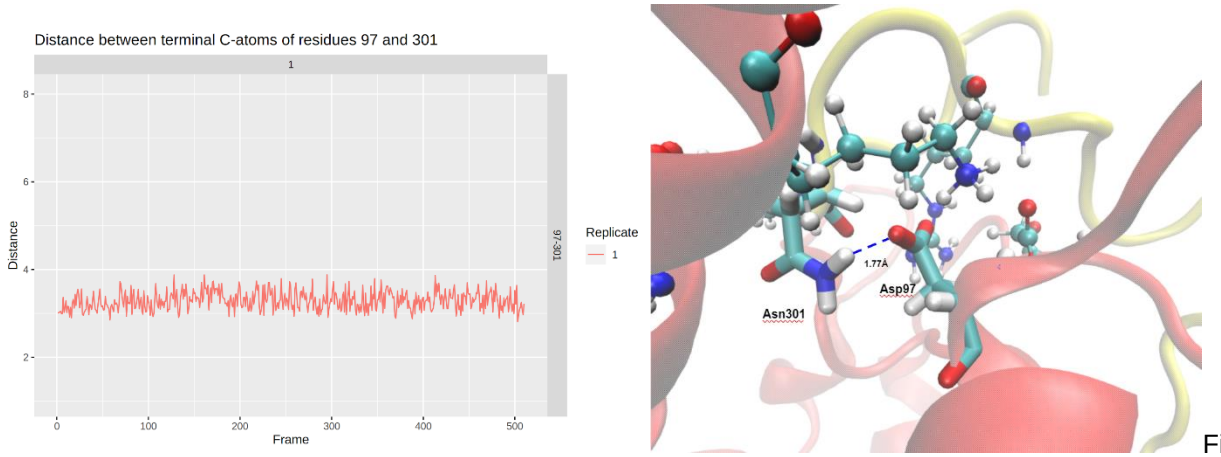


Figure 23. RMSD values of the C97N-C301D (up), C97D-C301N (middle) and C97S-C301S (down) mutants during 50 ns of MD simulation.



Figure

Figure 24. The hydrogen bonds between Asn301 and Asp97 during 50 ns of simulation of the C97D-C301N mutant and at the end of the simulation.

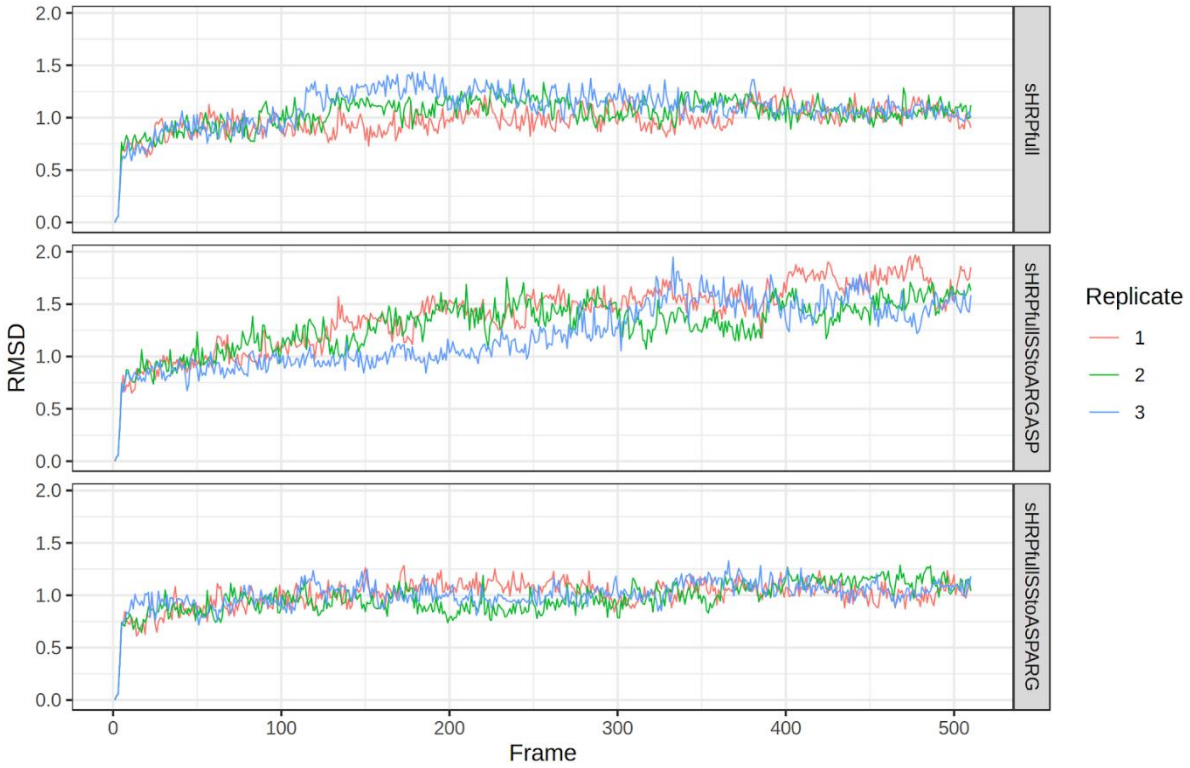


Figure 25. RMSD values of the split HRP protein (up), the C97R-C301D (middle) and C97D-C301R (down) mutants during 50 ns of MD simulation.

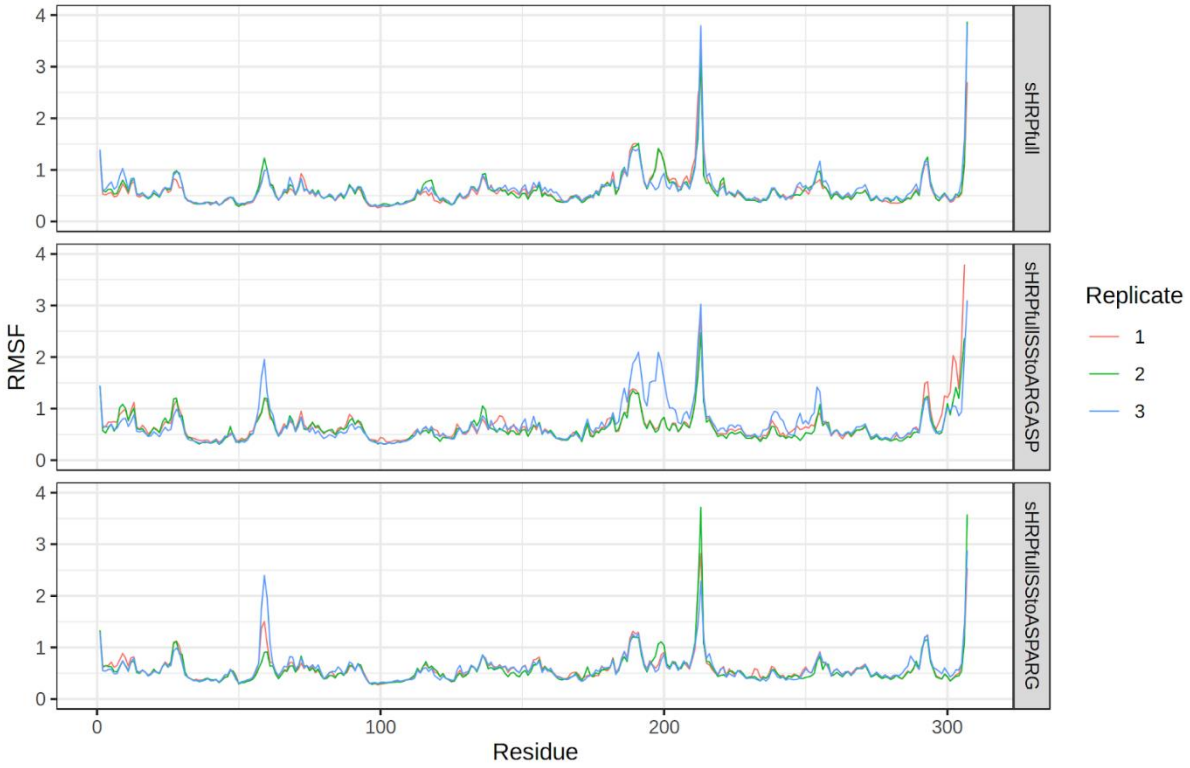


Figure 26. RMSF values of the split HRP protein (up), the C97R-C301D (middle) and C97D-C301R (down)

mutants during 50 ns of MD simulation.

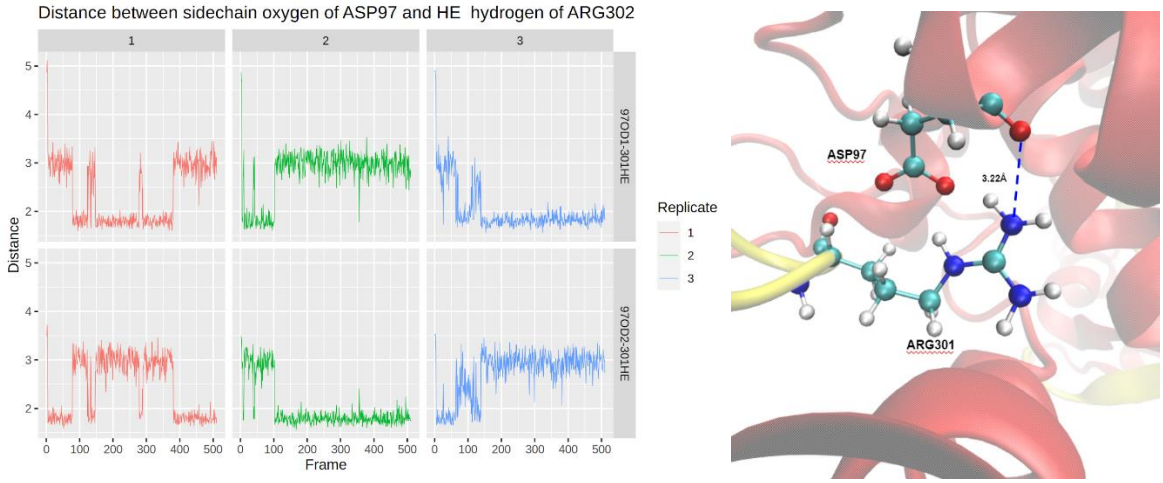
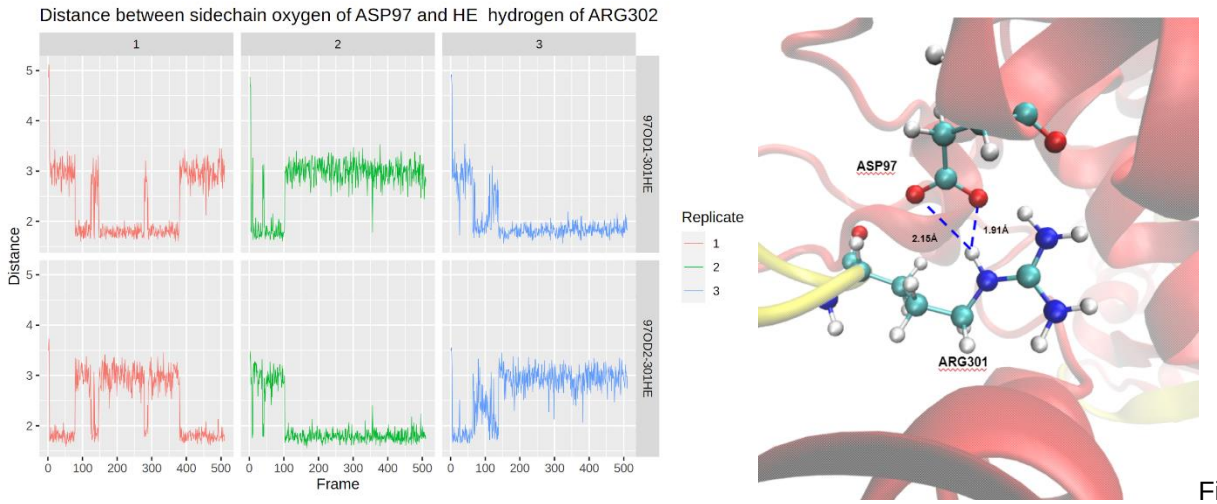


Figure 27. The hydrogen bond between the backbone oxygen of Asp97 and the sidechain of Arg301 during 50 ns of simulation of the C97D-C301R mutant and at the end of the simulation.



Figure

Figure 28. The hydrogen bonds between the sidechain oxygen atoms of Asp97 and the sidechain of Arg301 during 50 ns of simulation of the C97D-C301R mutant and at the end of the simulation.

2.2.2.2 Detailed analysis of mutation of Cys97-Cys301 into Ser97-Ser301

In order to provide molecular interpretation of experimental results which showed that the HRP protein is inactive when Cys97-Cys301 is mutated into Ser97-Ser301, MD simulations were performed. Since significant changes in the protein structure of the Ser97-Ser301 mutant were not noticed during the classical MD simulations (0.5 μs), enhanced MD simulation pulling technique was applied in an attempt to capture events that might happen on a longer timescale. After 0.5 μs of MD simulations with and without mutation, two amino acids in two systems (Cys97-Cys301 (no disulfide bond) and Ser97-Ser301) were pulled (Figure 29). The distance was increased from approximately 0.6 nm to approximately 1.0 nm with a force of 1000 kJ mol^{-1} nm^{-2} and velocity of 0.001 nm ps^{-1}. Each newly prepared system was simulated for additional 0.5 μs and compared with the system before pulling.

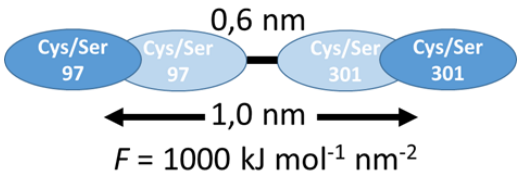


Figure 29. Schematic representation of pulling Cys97-Cys301 and Ser97-Ser301.

The distance between Cys97-Cys301 and Ser97-Ser301 after pulling did not return to the distance before pulling (Figure 30). The position 301 in the protein, Cys301 or Ser301, is close to the fluctuable C-terminal end of the protein (Figure 31 and Figure 32) whose fluctuations were additionally increased after the pulling. Fluctuations along the rest of protein are only slightly changed (Figure 31, Figure 35a). This is also noticeable in the RMSD plot from which the C-terminal amino acids are excluded. The global structure of the protein does not change and it was stable during the simulations (Figure 33a). The secondary structure changes only slightly at the C-terminal end (Figure 34).

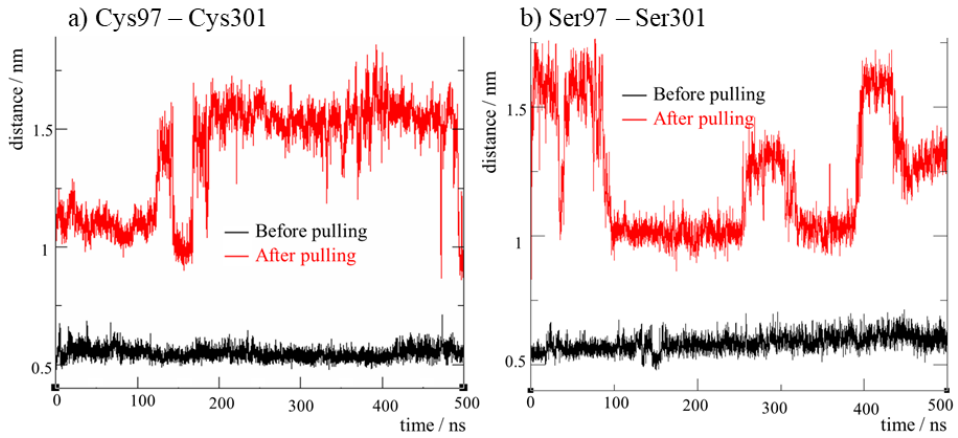


Figure 30. Distance between Cys97-Cys301 and mutated Ser97-Ser301, before and after pulling.

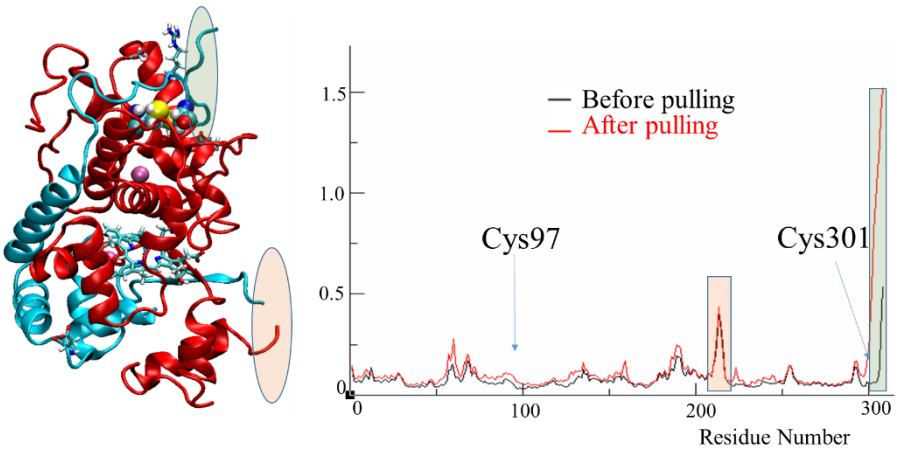


Figure 31. Left: structure of the protein before pulling after 500 ns of simulation. Right: RMSF of system with Cys97-Cys301 before and after pulling.

On the other hand, during the simulation of mutated Ser97-Ser301 pulling caused incensement of fluctuations along the whole protein (Figure 32 and Figure 35b). Even the structure of the protein was not stable during the simulation (Figure 33b). Analysis of secondary structure shows that changes in the overall structure, not only at the C-terminal end as in the non-mutated system.

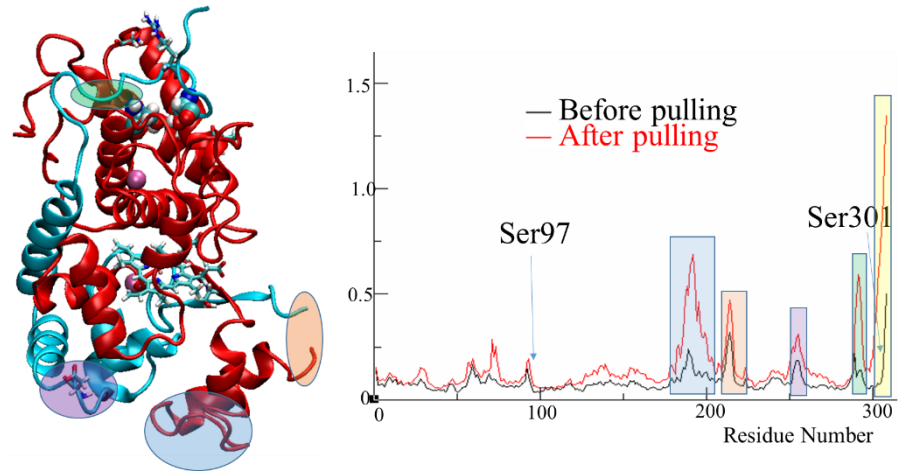


Figure 32. Left: structure of the protein before pulling after 500 ns of simulation. Right: RMSF of system with Ser97-Ser301 before and after pulling.

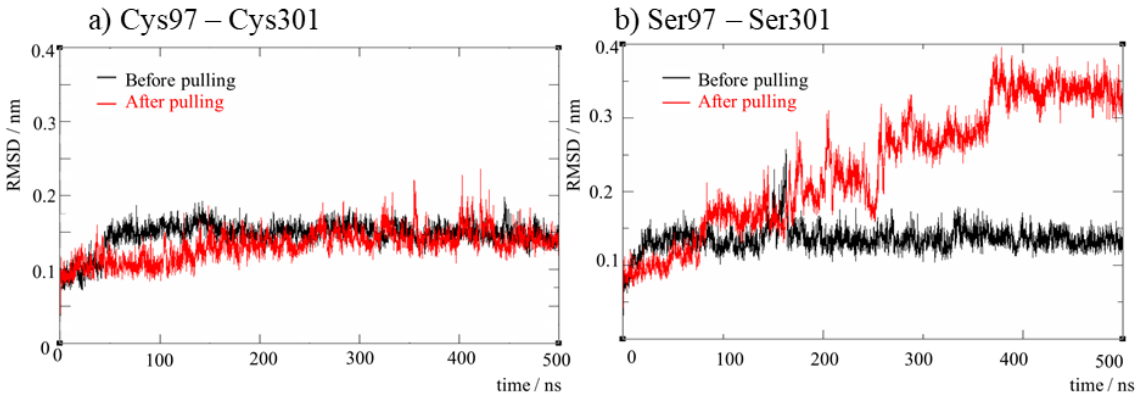


Figure 33. RMSD of systems with a) Cys97-Cys301 and b) mutated Ser97-Ser301, before and after pulling. Amino acids 1-300 were considered because C-terminal ends (amino acids 301-308) are extremely fluctuable and not included in calculation of RMSD.

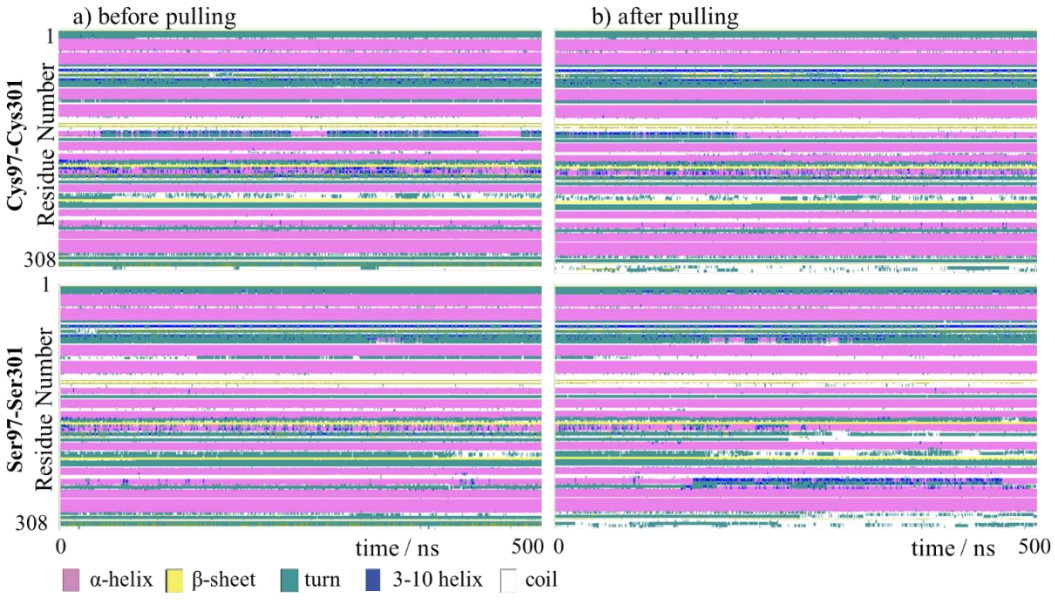


Figure 34. Secondary structure evolution of the HRP with Cys97-Cys301 (upper) and Ser97-Ser301 (down) a) before and b) after pulling.

The results show that the structure of the non-mutated Cys97-Cys301 system is significantly more stable than the structure of the Ser97-Ser301 mutant. It can be concluded that the disulfide bridge formed by residues Cys97 and Cys301 is important for the stability of sHRP structure which is in accordance with experimental results.

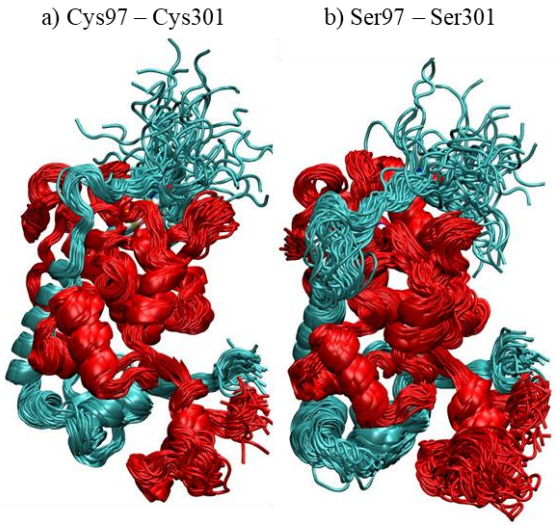


Figure 35. Snapshots taken every 1000th frame of MD simulations of (glyco)proteins aligned by the backbone Cα atom. Subunit A (residues 1-213) is shown in red and subunit B (residues 214-308) is shown in cyan.

2.2.3 Influence of short and long His-tag presence on: mHRP, sHRP, sHRP-A and sHRP-B

In one phase of the project conduction, introduction of polyhistidine-tag(s) (His-tag(s)) on C- and/or N-terminus of HRP was considered by experimentalist project collaborators. Such His-tag(s) were considered as a possible target(s) for oligonucleotide binding. In order to support the experimentalists and provide atomistic/molecular insights into the effects that introduction of His-tag(s) would have on structural and dynamical properties of different forms of HRP, MD simulations of HRP protein with His-tag(s) at C- and/or N-terminal were prepared, conducted and analyzed.

In order to study the effect of the length of His-tag(s) on HRP protein properties, two different types of His-tags were considered and added to the C- and/or N-terminal ends of: mHRP, sHRP, sHRP-A and sHRP-B structures. A short His-tags containing five amino acids: (i) HHHGS- for N-terminal end of protein and (ii) -GSHHH for C-terminal end of protein were computationally investigated. A long His-tag containing twelve amino acids: (i) HHHHHHGSGAGS- for N-terminal end and (ii) -ASGAGSHHHHHH for C-terminal end of protein were computationally investigated. Since it was shown that glycosylation is important for protein function, the protein was glycosylated with $Man_{16}GlcNAc_2$ branching type in all systems. In mHRP and sHRP, short and long His-tags were added to both terminal ends. Two additional simulations were performed with mHRP with long His-tag on either C- or N- terminal. In addition, MD simulations of sHRP-A with either short or long His-tags on N-terminal and of sHRP-B with either short or long His-tags on C-terminal were performed (Figure 36).

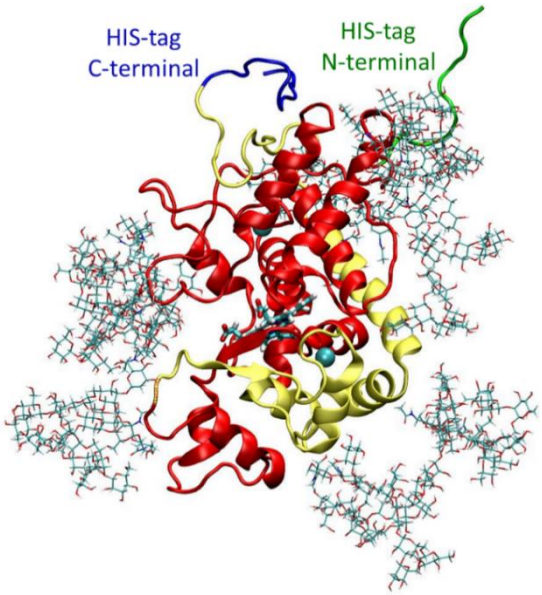


Figure 36: Long His-tag added to N-terminal and C-terminal ends of protein colored in green and blue, respectively.

2.2.3.1 Short His-tag simulations

Short His-tag simulations were performed in order to study the behavior of the tagged protein to try to determine whether the protein termini were accessible for reactions. Both the N- and C-terminal His-tags, are extremely flexible (Figure 37) in mHRP and sHRP.

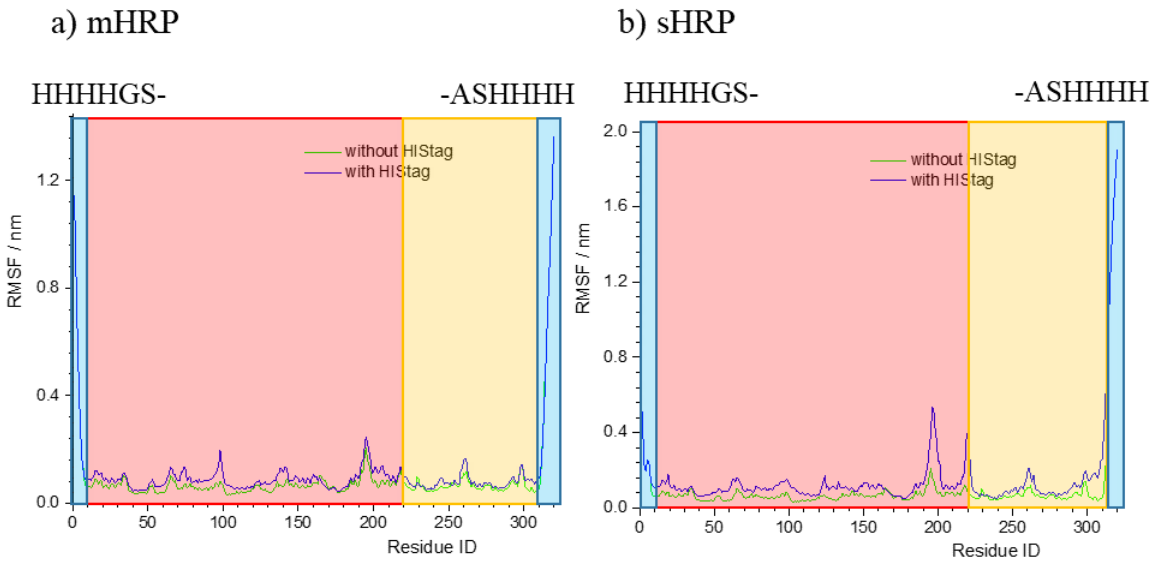


Figure 37: RMSF values of: a) mHRP protein and b) sHRP protein with and without short His-tags.

All of simulated His-tags interact with glycans (Figure 38) where the N-terminus in the mHRP structure is partly covered by glycans. In sHRP the C-terminus is partly covered by glycans. We concluded that it is possible that the glycans limit the availability of the short His-tags for reaction and prepared the simulations with longer His-tags.

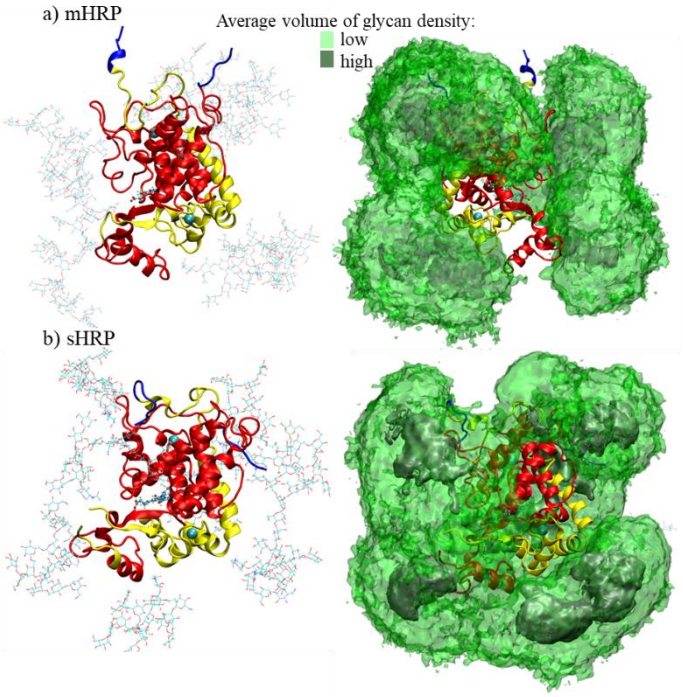


Figure 38: Left: snapshots of protein with attached short His-tags presented in blue. Right: figure in left with the average volumes of glycans around protein shown in green.

2.2.3.2 Long His-tag simulations

The mHRP, sHRP, sHRP-A and sHRP-B systems were prepared and run with long His-tags (defined in the previous section). As in the simulations with short His-tags, in these simulations the His-tags at the N- and C-terminal ends of the protein are extremely flexible (Figure 39). Glycans occupy a large area around the protein and surround the His-tags (Figure 40). From the averaged volume maps for both, the long His-tags and the glycans (Figure 41), it is difficult to have the reliable conclusions about N- and/or C-terminal exposion to the solvent and, therefore, availability for reaction (attachment of oligonucleotide).

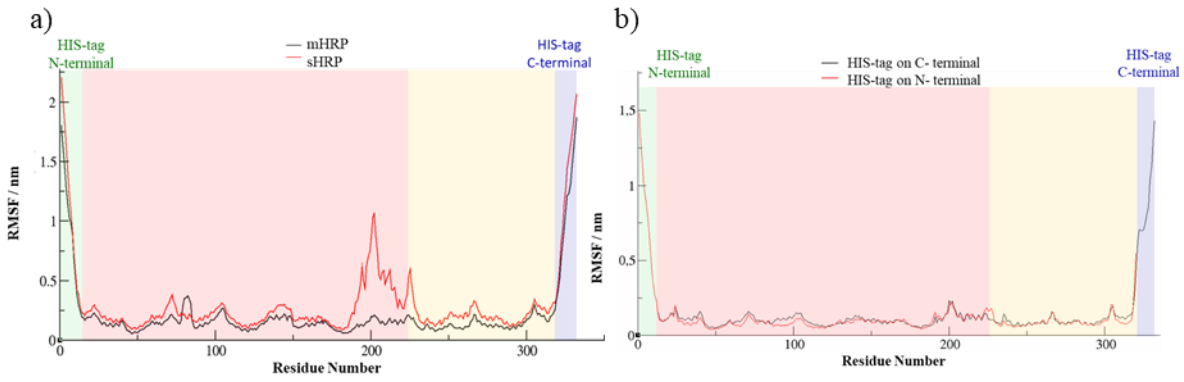


Figure 39: a) mHRP (black) and sHRP (red) with long His-tags attached to both protein termini. b) mHRP with His-tag either on C-terminus (black) or N- terminus (red).

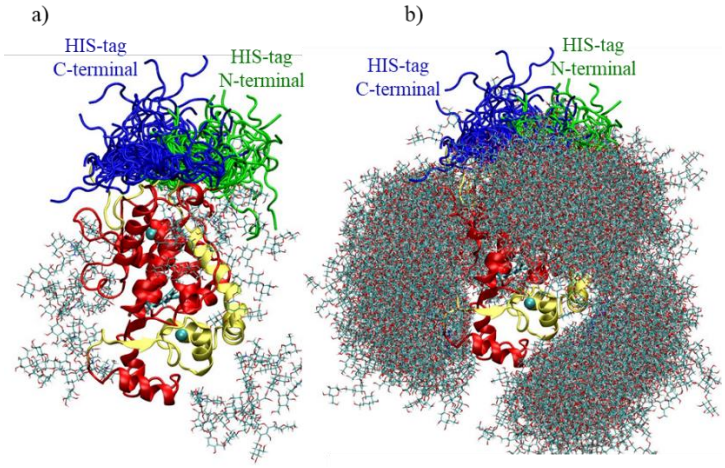


Figure 40: Snapshots taken every 1000th frame of MD simulations of (glyco)proteins with long His-tags aligned by the backbone Cα atom.

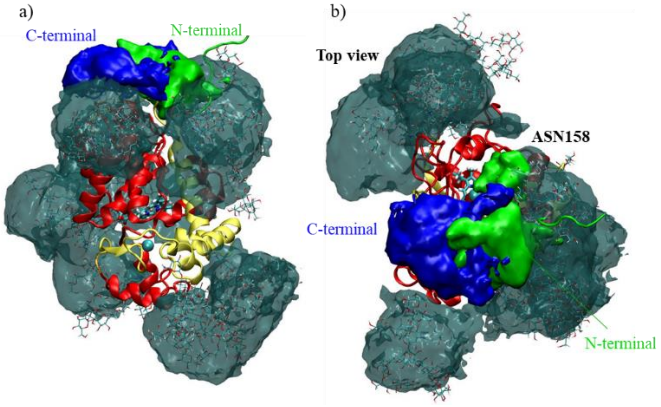


Figure 41. Availability of long His-tags added to sHRP shown from two different perspectives.

The similar situation is with simulations in which His-tag is present only on the N- or the C-terminus of the protein (Figure 42).

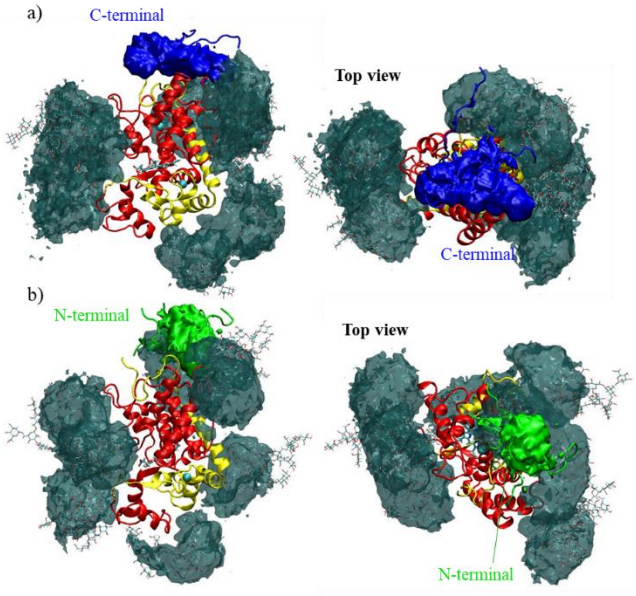


Figure 42. Availability of long His-tags added to sHRP a) only at C-terminus and b) only on N-terminus.

2.2.4 Availability of Lys for reaction (attachment of oligonucleotide)

As colleagues from AIT managed to express HRP structure with the $Man_5GlcNAc_2$ glycan (Figure 43a) additional systems with this specific glycosylation pattern have been simulated. The $Man_5GlcNAc_2$ glycan is smaller than previously investigated glycans (Section 1) and therefore covers a smaller area of the HRP enzyme (Figure 43b). The influence of $Man_5GlcNAc_2$ glycans on the structure of the mHRP and sHRP enzyme has been tested, as well as its steric influence on the potential single strand DNA (ssDNA) binding sites: primarily the N- and C- terminus, but also lysines were studied as well.

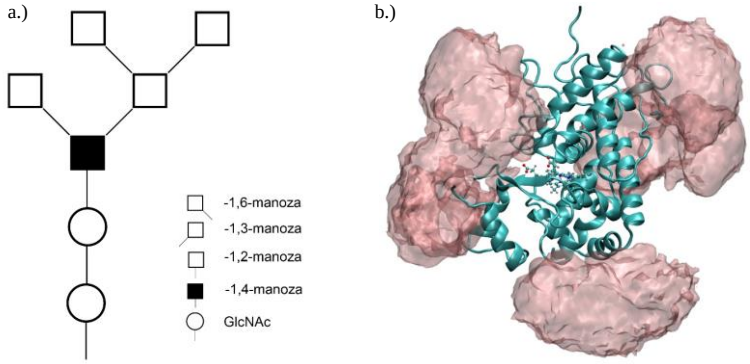


Figure 43. a) Schematic representation of the $Man_5GlcNAc_2$ glycan and b) averaged volume which this glycan covers during simulation of sHRP without a His-tag.

To determine the effect of the $Man_5GlcNAc_2$ glycan presence on stability of the mHRP and sHRP, heatmaps describing atom fluctuations were generated (Figure 44). These heatmaps describe the deviation of the C_α backbone carbon atom (y-axis) during the trajectory (x-axis) relative to the position of the same atom at the start of the simulation (colour).

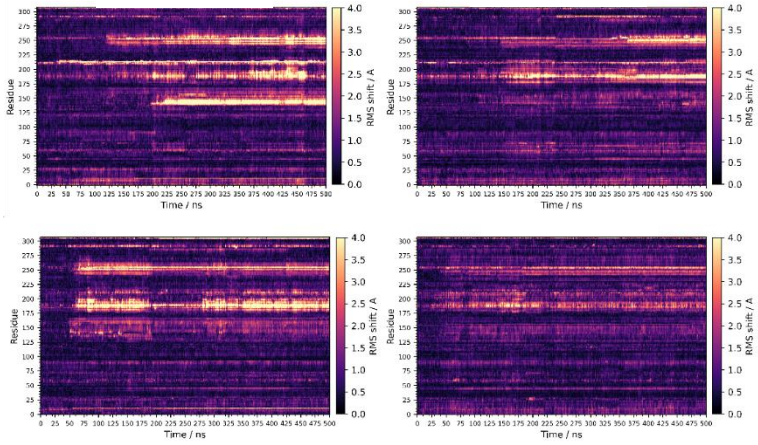


Figure 44. Heatmaps for different simulated systems. sHRP (top left) and mHRP (bottom left) without glycans; sHRP (top right) and mHRP (bottom right) with $Man_5GlcNAc_2$ glycan present. Colour describes the deviation of the C_α backbone carbon atom relative to the position of the same atom at the start of the simulation.

By subtracting the values of this deviation for simulations of the non-glycosylated protein from the $Man_5GlcNAc_2$ glycosylation type and assigning a color scale to the obtained difference one can obtain a single heatmap diagram that describes the influence of the $Man_5GlcNAc_2$ glycans on the HRP backbone carbon atom fluctuation (Figure 45). Red colour on the obtained heatmap (positive values) indicates residues which are more stable in glycosylated HRP, while blue colour (negative values) indicates residues which are more stable

in non-glycosylated HRP. White colour (neutral values) simply means that there is no significant difference in fluctuations between glycosylated and non-glycosylated form.

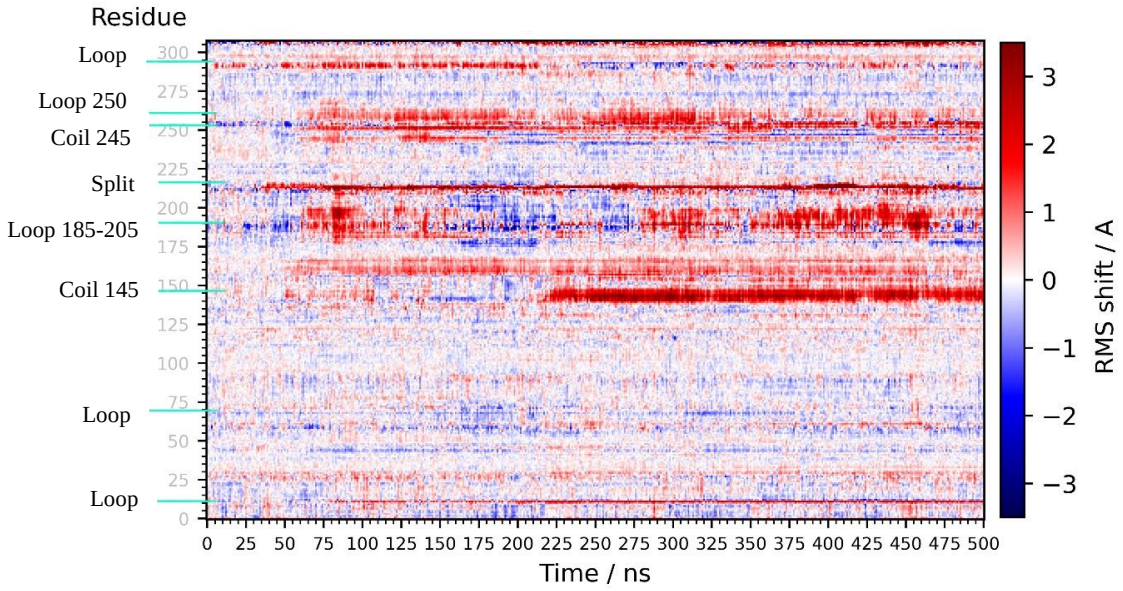


Figure 45. Heatmap describing influence of the $Man_5GlcNAc_2$ glycans on the movement of the C_α backbone carbon atom.

In general, the glycosylated forms of the mHRP and sHRP are more stable than non-glycosylated forms. This is especially evident in case of the area around residue 145 which experiences large shifts in positions of the α-coil during simulation (Figure 46). This change happens only during simulations of the non-glycosylated sHRP and as a result there is an additional opening through which water molecules can approach the active site. Because this effect didn't appear during non-glycosylated mHRP simulation, there isn't enough supporting evidence to link this behaviour (or lack of it) with glycosylation. Instead, it is likely that this change represents an additional conformational minimum the sHRP occupied during the MD simulations.

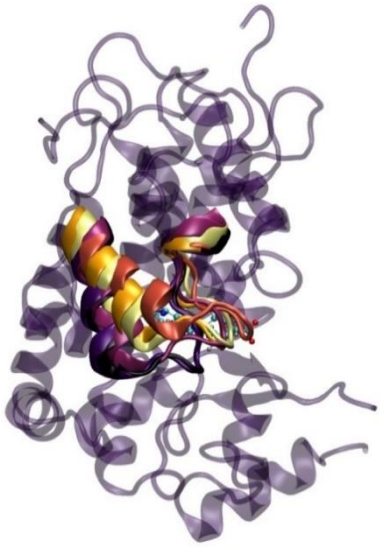


Figure 46. Structure of the non-glycosylated sHRP. Positional change of the α-coil in the region around residue 145 is shown in time. Dark colours represent structures at the start and bright colours at the end of the simulation.

Further, simulations pointed to the differences in the region around residues 213-214 where the split is

introduced in the sHRP structure. This region is much more stable in case of the glycosylated sHRP than of the non-glycosylated sHRP. A possible reason for this is the Asn255 glycosylation site. Addition of the $Man_5GlcNAc_2$ glycan at this split site drastically reduces its fluctuation.

RMSD data further confirms that the HRP protein is more stable when it is glycolyzed with $Man_5GlcNAc_2$ glycan, both in case of the mHRP and sHRP structure (Figure 47). This data presented here is completely in agreement with previously reported data on the simulations with glycolisation achived by $Man_8GlcNAc_2$, $Man_{16}GlcNAc_2$ and $Man_{20}GlcNAc_2$ glycans (Section 1).

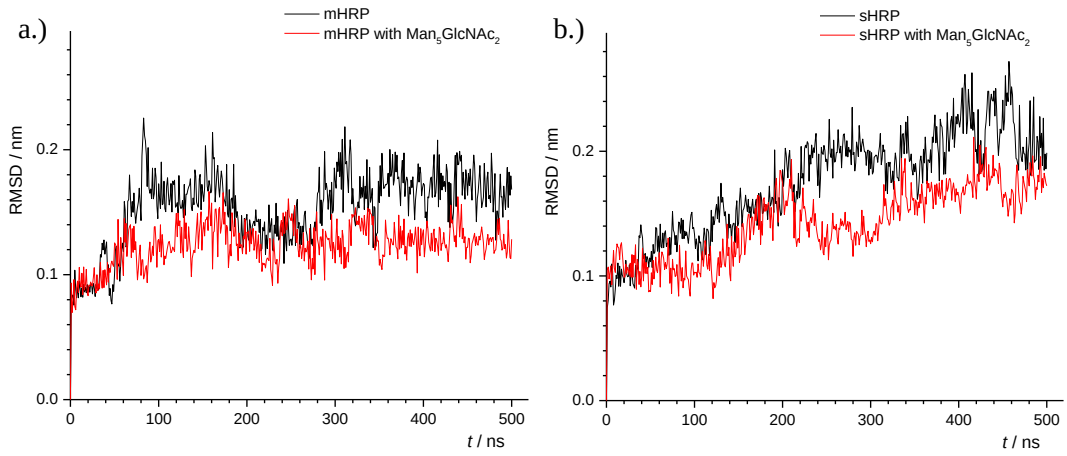


Figure 47. RMSD of backbone carbon atoms (C_α) of the non-glycolyzed (black) and $Man_5GlcNAc_2$ glycolyzed (red) HRP protein during the MD simulation in the case of a) mHRP and b) sHRP.

The residues Arg38, His42 and His170 are in the active site of the enzyme. In Figure 45, the mentioned region around these residues is primarily white indicating that there is no difference in fluctuations between non-glycosylated and glycosylated mHRP and sHRP. However, this merely indicates that enzyme is stable in the mentioned region, but $Man_5GlcNAc_2$ glycans could still reduce activity of the HRP through steric effects by blocking the entrance to the active site. Therefore, visual inspection was conducted followed by calculation of the average volume area that glycans and water molecules occupy around the active site region during the MD simulations (Figure 48). This visualisation shows that $Man_5GlcNAc_2$ glycans are too small to cover the entrance into the enzyme active site region.

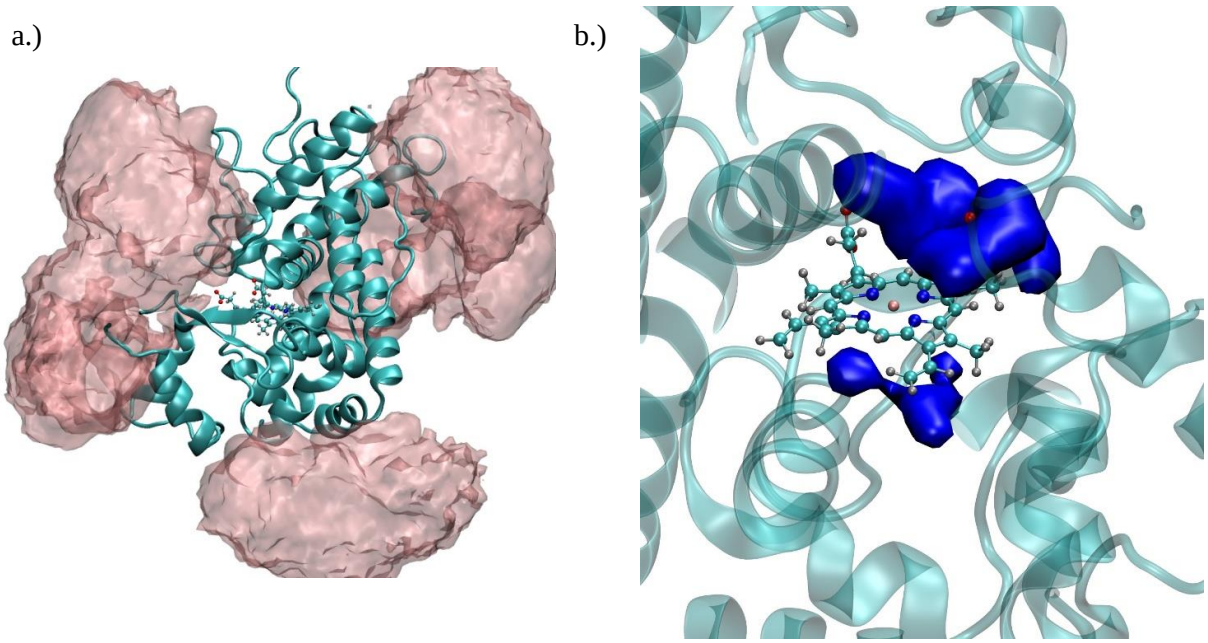


Figure 48. sHRP glycosylated with $Man_5GlcNAc_2$ glycans. a) Volume map of the glycans around the entrance to the active region of the sHRP. b) Volume map of the water molecules around heme cofactor.

A quantitative comparison of the availability of the active site between glycosylated and non-glycosylated forms of the mHRP and sHRP was conducted by counting the number of water molecules within 0.5 nm of the heme cofactor and Arg38 residue during the MD simulations (Figure 49). The number of water molecules in the investigated cases was almost constant during the 500 ns simulations with, on average, nine water molecules in all simulations, regardless of glycosylation. These further confirms the conclusion that $Man_5GlcNAc_2$ glycans don't interfere with the HRP active site by lowering its accessibility toward molecules from the bulk of the solution.

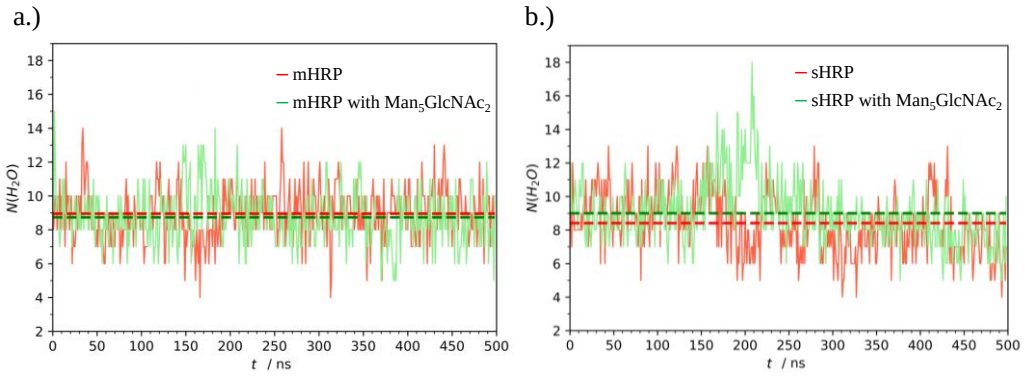


Figure 49. Number of water molecules around heme cofactor and Arg38 residue during MD simulation in the case of a) mHRP and b) sHRP. Red indicates water molecules in the case of non-glycosylated HRP while green indicates water molecules when $Man_5GlcNAc_2$ glycans are present.

In general, we can conclude that, although there are some changes in the backbone carbon atom (C_α) fluctuation, the general structure of the HRP enzyme and its enzyme activity are retained upon addition of the $Man_5GlcNAc_2$ glycans. The focal point of this investigation was also the steric influence of the $Man_5GlcNAc_2$ glycans on the availability of the C- and N-terminus, as well as lysine side chains, for ssDNA binding.

Availability of the C- and N-terminus is also interesting for purification efforts conducted by colleagues from AIT, as it relies on histidine binding to Ni-NTA columns. Figure 50 shows a comparison of C- and N-terminus availability between non-glycosylated (a) and glycosylated (b) HRP. When glycans aren't present, both terminal groups are free in solvent and available for chemical reactions (attachment of oligonucleotide). However, when $Man_5GlcNAc_2$ glycans are present only the C-terminal residues are free, while the N-terminal residues are partly covered by neighbouring glycans. It is important to note that glycans fluctuate a lot during MD simulations, so this does not mean that N-terminal residues are always buried under glycans and completely unavailable. It merely means that the C-terminal residues are more easily approached by reactant molecules then N-terminal residues and, therefore, any chemical modification will be easier on the C-terminal group. Chemical reactions on the N-terminal group are sterically more hindered, but not necessarily impossible.

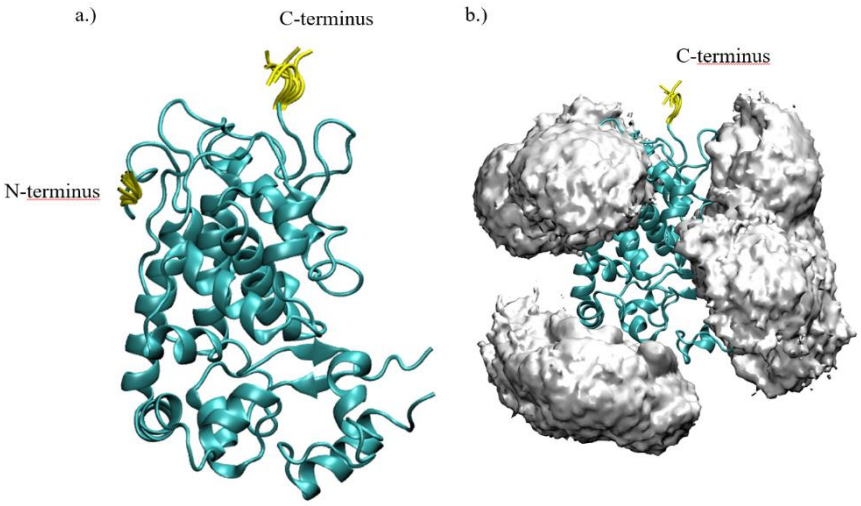


Figure 50. Fluctuations of N and C terminal groups (yellow) during the MD simulations in case of the a) non-glycosylated and b) glycosylated sHRP. White clouds represent average volume that glycans occupy during the MD simulation.

HRP enzyme contains six lysine groups which are located at residues 65, 84, 149, 174, 232, 241. Most of these lysine amino acids are oriented towards the solution with the exception of lysine residues 65 and 84 which are buried inside the HRP structure. Quantitative analysis was conducted by calculating the number of water molecules inside a 0.5 nm radius around the lysine sidechain nitrogen atom. The average number of these water molecules with standard deviation and median value are reported in Figure 51. The data presented here confirms that Lys65 and Lys84 are buried inside the HRP protein structure with a very small number of water molecules around the nitrogen atom on lysine sidechain (five atoms in the case of Lys65 and only one atom in the case of Lys84). The rest of the lysine residues have a higher median value of water molecules surrounding them during the simulation, which ranges from 8 – 11 water molecules. This confirms that these lysine residues are oriented toward the bulk of the solute, and we hypothesise that these can be used for ssDNA tagging.

$Man_5GlcNAc_2$ glycans have a steric influence on water accessibility of two lysine residues: Lys149 and Lys241. In both cases the number of water molecules around the sidechain nitrogen atom is lower. In the case of Lys149 this effect is mild with around 20 % lower median value of water molecules, but in the case of Lys241 this effect is much more expressed with 40 % lower median value of the water molecules.

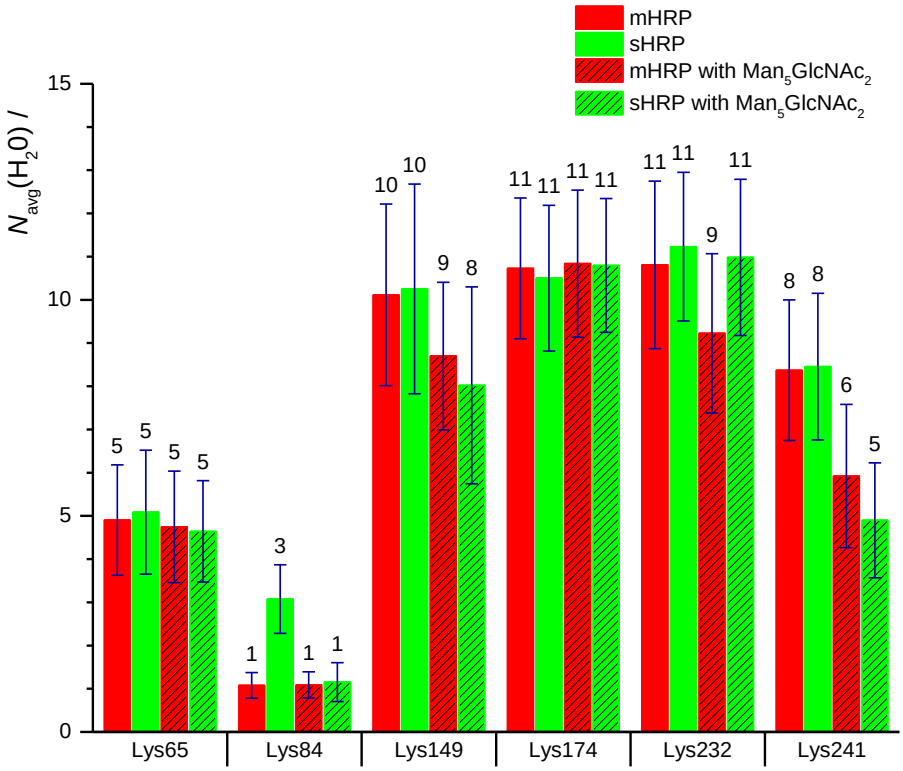


Figure 51. Average number of water molecules around lysines during MD simulation of non-glycosylated (no pattern) and glycosylated (with pattern) mHRP (red) and sHRP (green). Median value is shown by number above each bar, while standard deviation is indicated with blue error bars.

Based on this, we can conclude that, although $Man_5GlcNAc_2$ glycans are relatively small and cover a small area of the HRP protein, they still have a steric influence on lysine accessibility toward ssDNA tagging. In particular, ssDNA tagging of the glycosylated HRP should be a more specific reaction than in the case of the non-glycosylated HRP.

2.2.5 Effect of mutations present in mHRP and sHRP structures

Effect of the Martell mutations (Figure 11) introduced into the HRP was investigated on a molecular level using MD simulations. Simulations were run in two types of systems: HRP and sHRP. Both systems were non-glycosylated. Simulations were 500 ns long (Figure 52)

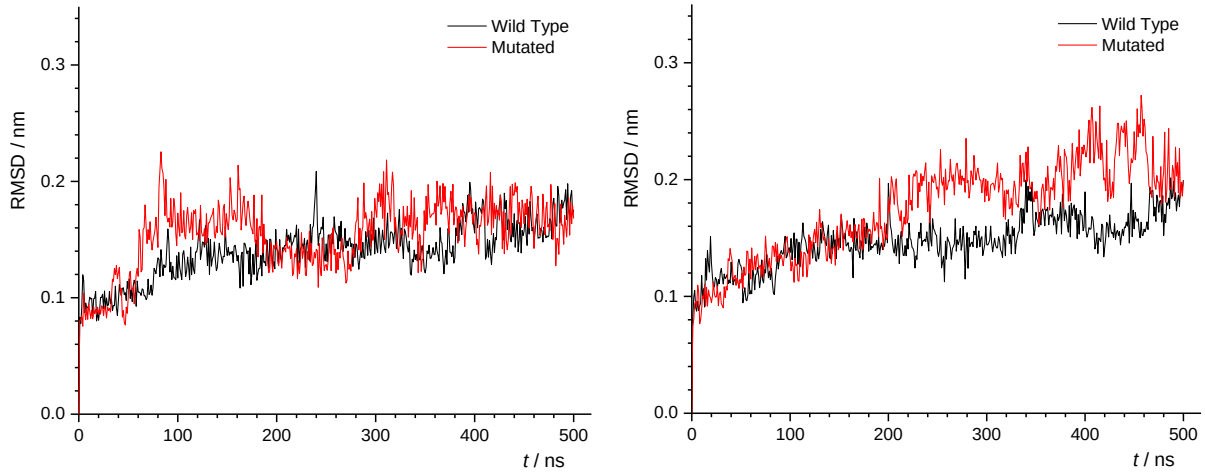


Figure 52. RMSD values of wild type (black) and mutated HRP (red) in case of HRP (left) and sHRP (right). Backbone carbon atoms (C_α) were considered in calculations.

The RMSF analysis pointed to the difference in fluctuations N255D mutation and regions around it. Detailed analysis showed that difference in fluctuations is due to the movement of heme which causes a positional shift of the whole α-coil region. This movement of heme is caused by the N175S mutation which creates a new hydrogen bond between heme and serine oxygens (Figure 54).

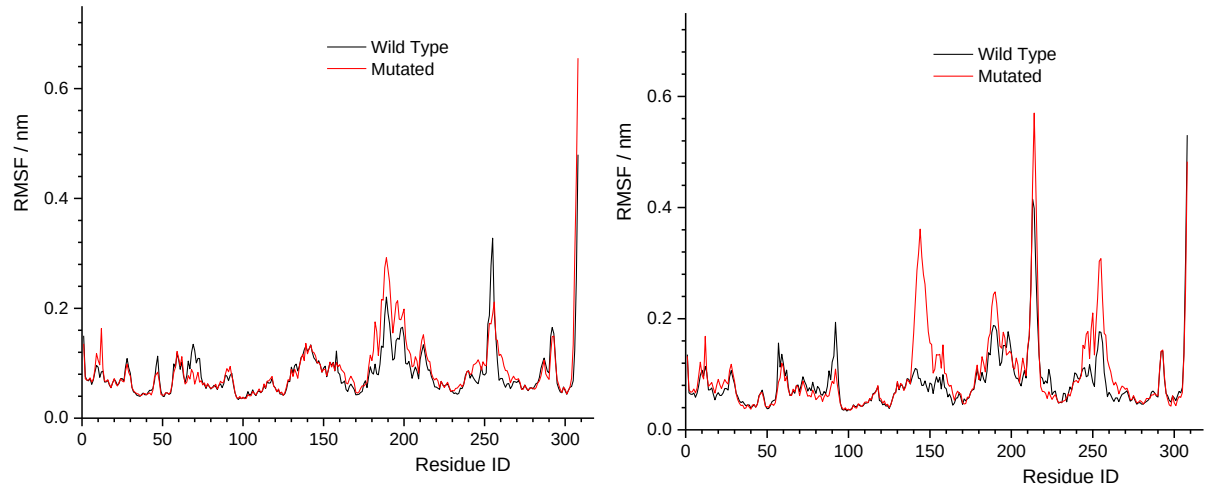


Figure 53. RMSF values of wild type (black) and mutated HRP (red) in case of HRP (left) and sHRP (right). Backbone carbon atoms (C_α) were considered in calculations.

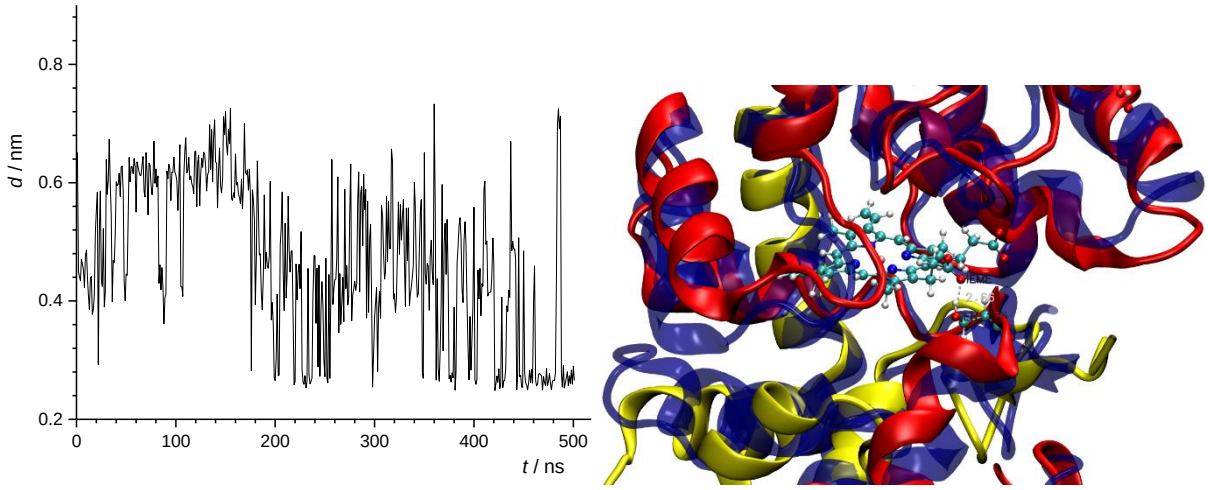


Figure 54. Newly established hydrogen bond between carboxyl atom of heme cofactor and SER175 oxygen.

2.2.6 Simulations without heme cofactor

The influence of the heme cofactor on protein stability was investigated in the case of sHRP and sHRP-A. To include the above-mentioned stabilization effect of glycans on the structure of the HRP, holo and apo structures of sHRP and sHRP-A regarding the heme presence were simulated in non-glycolised and glycolised ($Man_5GlcNAc_2$) forms. Both the holo and apo sHRP structure show similar stability during simulations with and without $Man_5GlcNAc_2$ (Figure 55). This shows that the sHRP fold is structurally stable and removal of the heme cofactor has no significant influence on protein structure stability. There are no significant changes in positions of the α-coils or β-sheets (Figure 56).

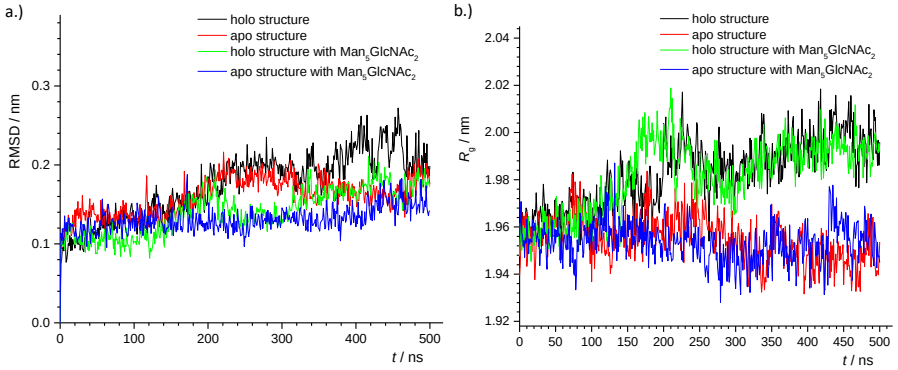


Figure 55. Comparison of the HRP holo and apo structure during the MD simulation. a) RMSD of backbone carbon atoms (C_α), b) gyration radius of the HRP protein.

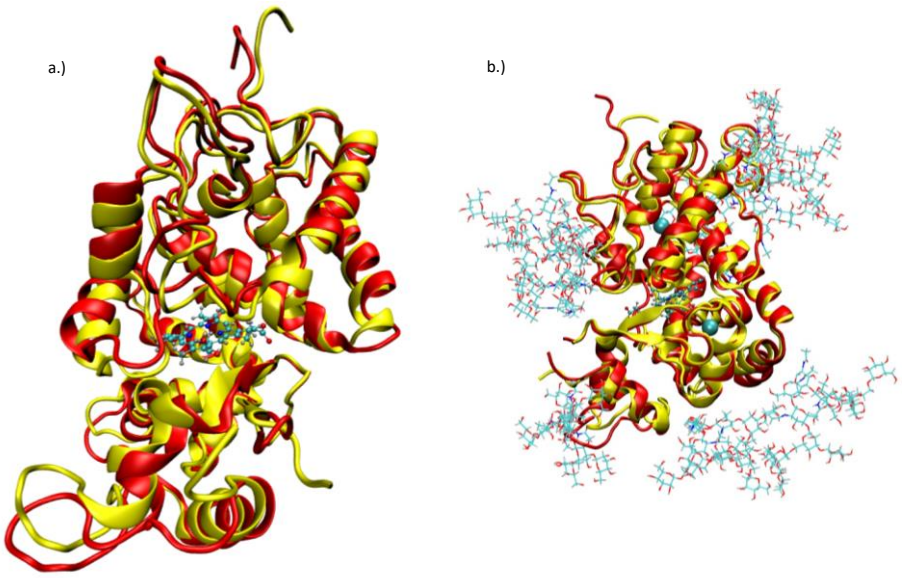


Figure 56. Comparison of the HRP holo (red) and apo (yellow) structure at the end of the MD simulation in the case: a) without glycan and b) with $Man_5GlcNAc_2$ glycan.

In the case of the sHRP-A removal of the heme cofactor has influence on its structure (Figure 57). In both simulations, with and without the $Man_5GlcNAc_2$ glycan, the holo structure is more stable.

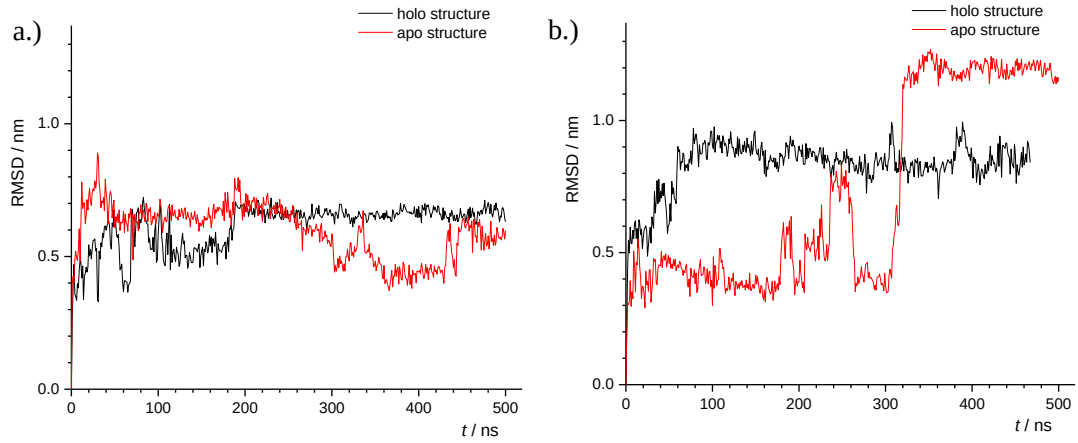


Figure 57. RMSD of backbone carbon atoms (C_α) of the sHRP-A holo (black) and apo (red) structure during the MD simulation in the case without glycan (a) and with $Man_5GlcNAc_2$ glycan (b).

In the case of the apo structure sHRP-A, a creation of a hydrophobic pocket involving eight amino acids: Val23, Leu26, Leu111, Phe179, Ile180, Leu184, Leu205 and Leu208 (Figure 58b) was noticed. This hydrophobic pocket is stable during the simulations and, once sHRP-A attains this conformation, it remains in it for the rest of MD simulation.

In the case of the non-glycosylated form of the sHRP-A this conformation, with a hydrophobic pocket, was not noticed. Instead, only two smaller hydrophobic pockets are formed involving only four out of the eight mentioned amino acids (Figure 58a). These smaller hydrophobic pockets are less stable, and their structure varies through the simulation. The most likely explanation for this difference in behaviour between non-glycosylated and glycosylated sHRP-A, is related to the previously described effect of glycolysation on protein

structure (Section 1).

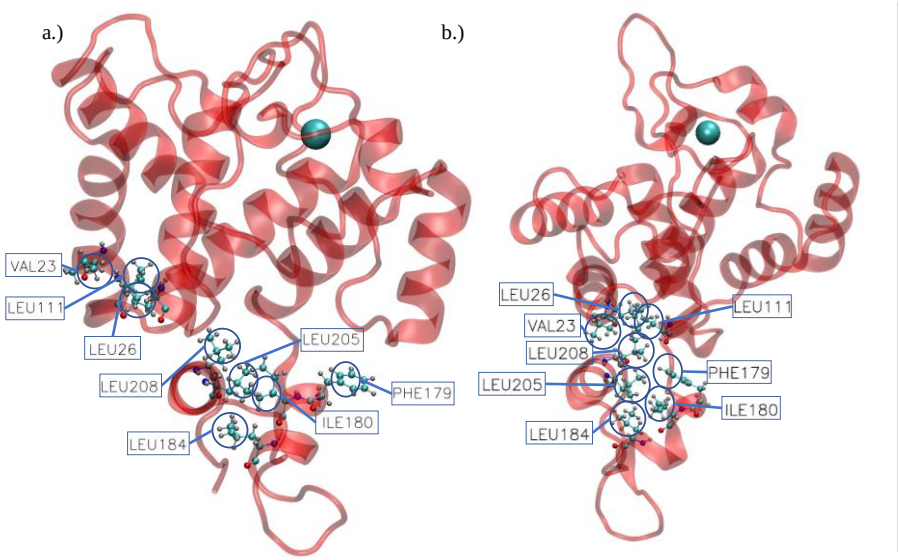


Figure 58. Relaxed conformations of sHRP-A apo structure without glycan (a) and with $Man_5GlcNAc_2$ glycan.

3. Conclusion

The results of the MD simulations show that *N*-glycosylation has a significant effect on the dynamical properties of HRP and sHRP. Fluctuations of amino acids were decreased in glycoproteins compared to non-glycosylated proteins. A decrease in fluctuations was especially pronounced in the central and peripheral regions of both proteins, HRP and sHRP. Beside these two regions, in the case of the sHRP protein, the fluctuations of the cut-site region are also significantly affected by glycosylation. Interestingly, the effect on the decrease of fluctuations is not only localized to the area around glycans, but it also propagated to the central part of the protein including the active site. Further, *N*-glycosylation influences the electrostatic potential of the protein. Induction of polarization of electrostatic potential due to glycans presence was observed, and it affects even the heme cofactor. Altogether, simulations show that *N*-glycosylation is important for enzyme proper function and stability, especially in the case of sHRP.

In the frame of the project, computational simulations often served as a tool that enable a deeper interpretation of experimental results and helps in planning novel experiments. Computational results provided deeper interpretation experimental results by providing the explanation at molecular level for unsuccessful attempts of replacing the cysteine disulfide bridge between the two subunits of sHRP with strong H-bond. Further, computational simulations were applied to support experimentalists to find the most suitable protein group for attachment of the oligonucleotide. For that purpose, simulations of different forms of HRP with short and long His-Tags attached to the C- and/or N-terminal were performed. Further, the availability of lysine residues for the attachment of the oligonucleotide were studied and the most promising ones were identified in case of various HRP forms.

Taken altogether, the main purpose of computational simulations within the project was to provide insight into the molecular properties of the studied systems. The computational results were always compared and validated with experimental results, often help finding the most promising directions of experimental research.

4. References

1. GI, B. *et al.* The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417**, 463–468 (2002).
2. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).
3. Jo, S., Kim, T. & Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2**, e880 (2007).
4. Wu, E. L. *et al.* CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).
5. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
6. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
7. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
8. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
9. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
10. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
11. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).
12. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–8, 27–8 (1996).
13. Krainer, F. W. *et al.* Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in Pichia pastoris. (2013) doi:10.1038/srep03279.
14. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Reports 2015 51* **5**, 1–7 (2015).
15. Smith, A. T. *et al.* Expression of a synthetic gene for horseradish peroxidase C in Escherichia coli and folding and activation of the recombinant enzyme with Ca2+ and heme. *J. Biol. Chem.* **265**, 13335–13343 (1990).
16. Azevedo, A. M. *et al.* Horseradish peroxidase: A valuable tool in biotechnology. *Biotechnol. Annu. Rev.* **9**, 199–247 (2003).
17. Tams, J. W. & Welinder, K. G. Glycosylation and thermodynamic versus kinetic stability of horseradish peroxidase. *FEBS Letters* **421,** 234-236 (1998).