Supporting Information

Structural Analyses of *Glycyrrhiza glabra* C-Glycosyltransferase: A Molecular Dynamics Study to Elucidate Catalytically Active Complexes

Gonzalo A. Jaña,[a] Fabiola E. Medina,[b] Francisco Barrios,[c] Jorge Martínez-Araya,[d] and Fernanda Mendoza.[a]*

[a] Departamento de Ciencias Químicas, Facultad de Ciencias Exactas, Universidad Andres Bello, Autopista Concepción-Talcahuano 7100, Talcahuano, Chile.

[b] Departamento de Química, Facultad de Ciencias, Universidad del Bío-Bío, Avenida Collao 1202, Concepción 4051381, Chile.

[c] Centro de Investigación e Innovación Tecnológica (CIITEC), Estado Mayor General, Ejército de Chile, Valenzuela Llanos 643, La Reina, 7860251 Santiago, Chile.

[d] Departamento de Ciencias Químicas, Facultad de Ciencias Exactas, Universidad Andres Bello, Av. República 275, 8370146 Santiago, Chile.

Email: maria.mendoza@unab.cl

TABLE OF CONTENT

1. METHODOLOGY

1.1 Precursor Enzyme-complex model.

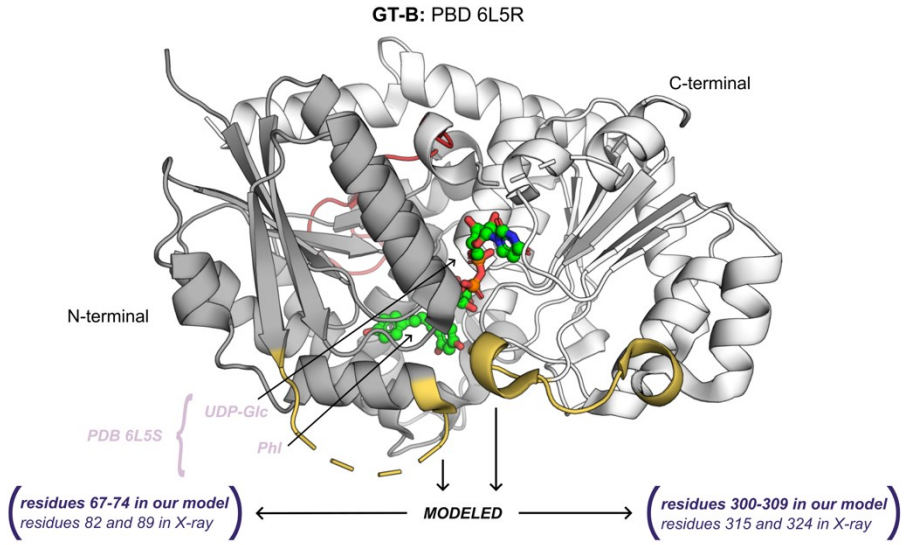

Figure S1. X-ray structure 6L5R (protein and UDP-Glc) combined with the X-ray structure 6L5S (Phl). The yellow loops were modeled by homology.

1.2 Global and local descriptors and indexes

Global reactivity descriptors.

Several reactivity descriptors were proposed from the Conceptual Density Functional Theory, among them:

- **Electronic chemical potential**, $\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)} = -\chi = -0.5(I_1 + A_1)$, which is expressed in $hartree \cdot e^{-1}$,and χ is the electronegativity according to the Mülliken's definition. It measures the escaping tendency of electron from equilibrium where I_1 is the first vertical ionization potential, and A_1 is the first vertical

electron affinity. On the other hand, the Frontier Molecular Orbital Approximation (FMOA), based on Koopmans' theorem, leads to the following working formula $\mu \approx 0.5(\varepsilon_H + \varepsilon_L)$ where ε_H stands for the HOMO energy, and ε_L, for the LUMO energy.

- **The molecular hardness**, $\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)} = I_1 - A_1$ is expressed in $hartree \cdot e^{-2}$ and it corresponds to the resistance to charge transfer; after applying Koopmans' theorem, we obtain $\eta \approx \varepsilon_L - \varepsilon_H$. The global softness is defined as $S = \eta^{-1} = \left(\frac{\partial^2 E}{\partial N^2}\right)^{-1}_{v(r)} = (I_1 - A_1)^{-1}$ and it is expressed in $hartree^{-1} \cdot e^2$; it quantifies the ease to charge transfer of a system. In terms of frontier molecular orbitals energies, we get $S \approx (\varepsilon_L - \varepsilon_H)^{-1}$. It measures the energy change of an electrophile when it becomes saturated with electrons, by considering the case when an electrophilic species is immersed in an idealized zero-temperature free electron sea of zero chemical potential. After applying the FMOA, it turns into $\omega \approx \frac{(\varepsilon_H + \varepsilon_L)^2}{8(\varepsilon_L - \varepsilon_H)}$. To quantify the response to charge donation and charge acceptance, the electron-donating and electron-accepting powers are defined as follows, respectively: $\omega^- = \frac{(\mu^-)^2}{2\eta^-} = \frac{(3I_1 + A_1)^2}{16(I_1 - A_1)}$ and $\omega^+ = \frac{(\mu^+)^2}{2\eta^+} = \frac{(I_1 + 3A_1)^2}{16(I_1 - A_1)}$, a smaller value of ω^- of a system makes it a better electron donor, whereas a larger ω^+ value corresponds to a better capability of accepting charge. When the

FMOA is applied, these formulae turn into $\omega^- \approx \frac{(3\varepsilon_H + \varepsilon_L)^2}{16(\varepsilon_L - \varepsilon_H)}$ and $\omega^+ \approx \frac{(\varepsilon_H + 3\varepsilon_L)^2}{16(\varepsilon_L - \varepsilon_H)}$. In order to quantify these both capabilities in just one term, the net electrophilicity $\Delta\omega^{\pm}$ was proposed and it corresponds to an electron-accepting power relative to electron-donating power; its working formula is given by: $\Delta\omega^{\pm} = \omega^+ - (-\omega^-) = \omega^+ + \omega^-$. Working formulae based on FDA or FMOA, depend on the respective working formulae of ω^+ and ω^- as defined before.

Local reactivity: Local Hyper-softness

From the Conceptual Density Functional Theory, a local reactivity descriptor that reveals sites on a molecule that is susceptible to undergo nucleophilic and electrophilic attacks trending to form covalent bonds is the dual descriptor $f^{(2)}(r)$

$$f^{(2)}(r) = \rho_{N+1}(r) - 2\rho_N(r) + \rho_{N-1}(r) \approx |\psi_{LUMO}(r)|^2 - |\psi_{HOMO}(r)|^2$$

where $|\psi_{LUMO}(r)|^2$ is the electron density of LUMO and $|\psi_{HOMO}(r)|^2$ is the electron density of HOMO. These parameters can be plotted as 3D isosurfaces revealing electrophilic and nucleophilic regions on a molecule. It is also known as a second-order Fukui function, has been defined by Christophe Morell et al. Its unit is $e^{-1} \cdot bohr^{-3}$ where e means electron. It ranges from -1 to 1. Nevertheless, dual descriptor cannot be used to compare local reactivities among different molecules because lobes become insignificant as the molecule's size increases. The local hyper-softness (LHS) mends this intrinsic behavior of the dual descriptor; its advantages is explained in J. Math. Chem. 62, 461–475 (2024). LHS is a

local reactivity descriptor that considers the molecular size and whose working formula based on the FMOA is presented as follows:

$$s^{(2)}(r) \approx S^2 f^{(2)}(r) = \left(\varepsilon_{LUMO} - \varepsilon_{HOMO}\right)^{-2} \cdot \left(\left|\psi_{LUMO}\,(r)\right|^2 - \left|\psi_{HOMO}\,(r)\right|^2\right)$$

Where S^2 is the squared global softness. $s^{(2)}(r)$ is defined in terms of LUMO and HOMO energies under the assumption that Koopmans' theorem is satisfied. Its unit is $e^3 \cdot hartree^{-2} \cdot bohr^{-3}$.

This descriptor's advantage is that it allows comparing local reactivities among different molecules; hence, we used it in the present work to assess the local reactivity of molecules under analysis.

2. RESULTS AND DISCUSSION

2.1 Root Mean Square Deviation (RMSD) Analysis for Nine Native Models.

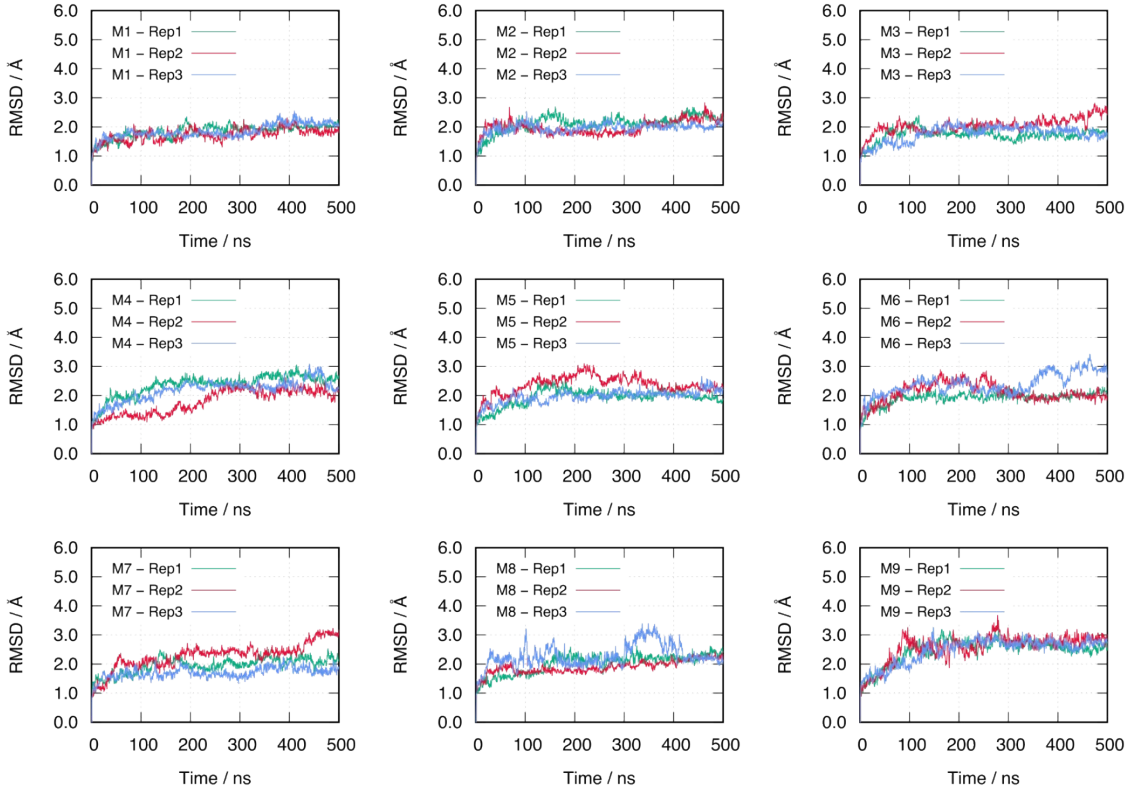

Figure S2. RMSD evolution of backbone atoms of the protein, along the MD simulations for the three replicas, considering models M1 to M9.

To verify the convergence along the 500 ns of Molecular Dynamics production, the RMSD was calculated for individual replicas **(Figure S2),** which not only demonstrates the convergence of the three replicas around 2 Å, but also indicates that the time of simulation is sufficient to sample the conformational landscape. Additionally, the RMSD normal distribution is represented for the backbone atoms of the protein, as well as the heavy atoms of donor and acceptor substrates.

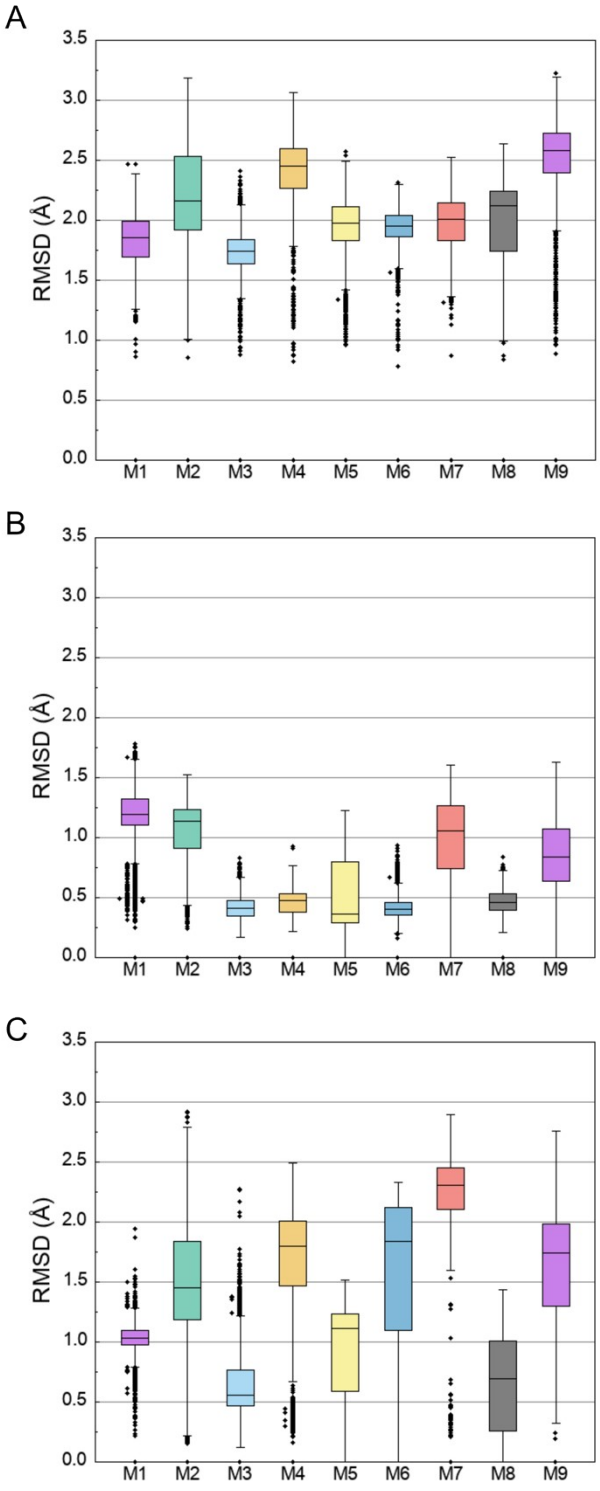

Figure S3. The normal distributions of RMSD are represented by using box plots for mutant models associated to models M1 to M9. **A.** Normal distribution of protein backbone; **B.**

Normal distribution of donor substrate (UDP-Glc) and; **C.** Normal distribution of acceptor substrate (Phl).

As can be observed, the medians of the RSMD backbone values oscillate around 1.6 Å in models M1 and M3, whereas the values of the rest of the models oscillate between 2.0-2.5 Å.

2.2 Root Mean Square Deviation (RMSD) Analysis for Mutant Models.

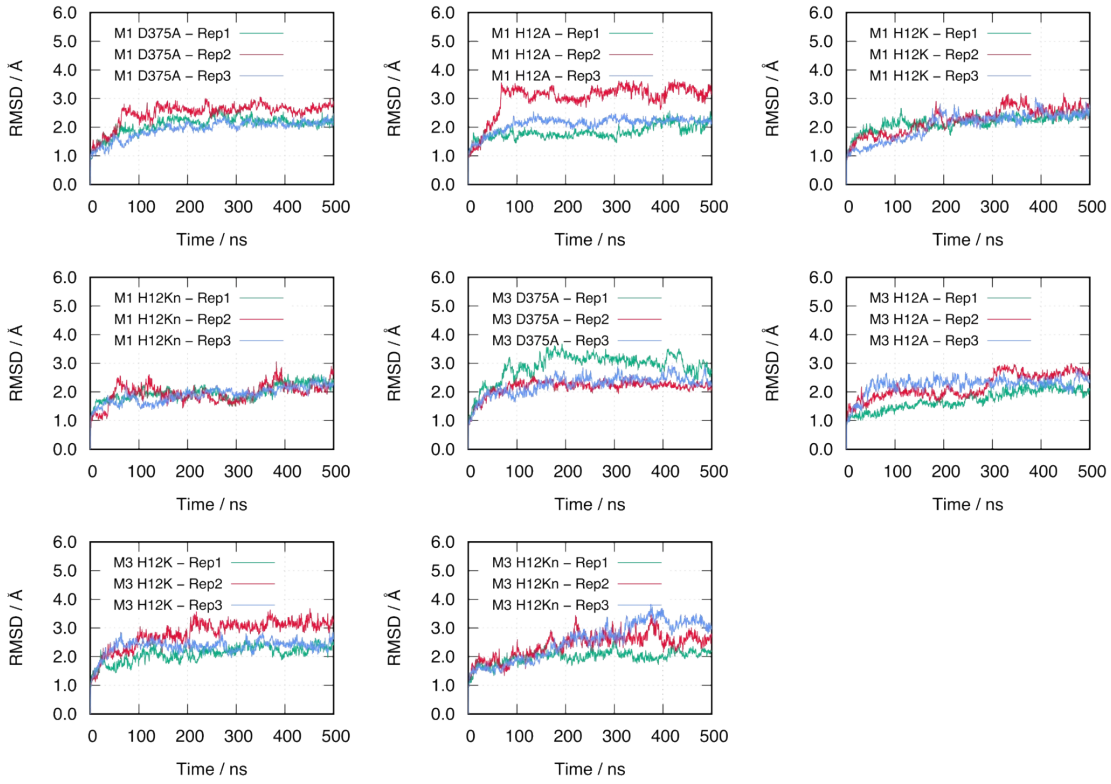

Figure S4. RMSD evolution of backbone atoms of the protein, along the MD simulations for the three replicas, considering all mutant models for models M1 and M3.

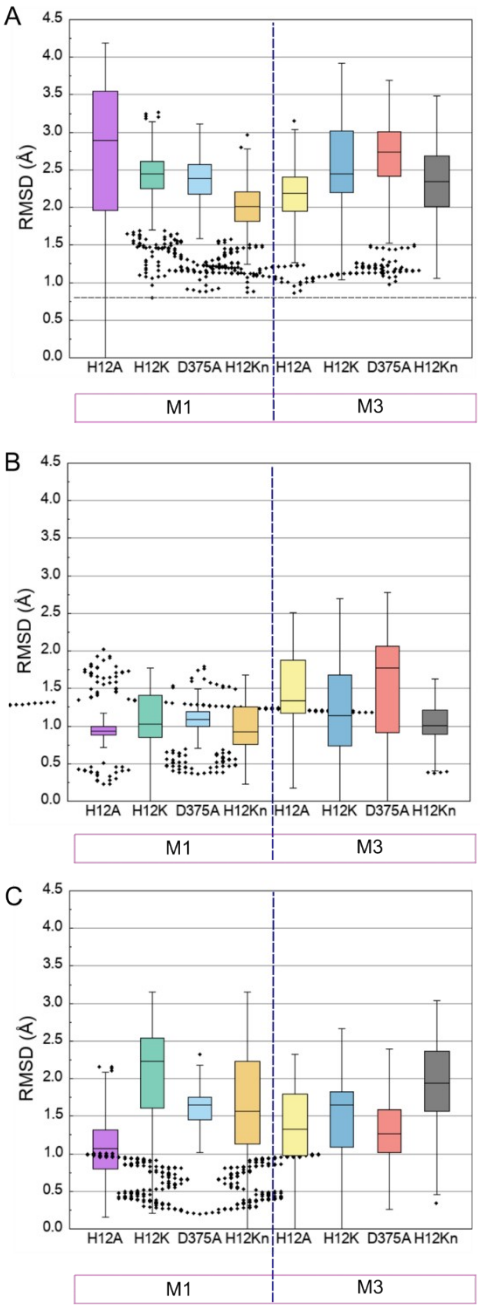

Figure S5. The normal distributions of RMSD are represented by using box plots for mutant models associated to M1 to M3. **A.** Normal distribution of protein backbone; **B.** Normal distribution of donor substrate (UDP-Glc) and; **C.** Normal distribution of acceptor substrate (Phl).

2.3 RMSF and Secondary Structure Analyses for the nine native models studied.

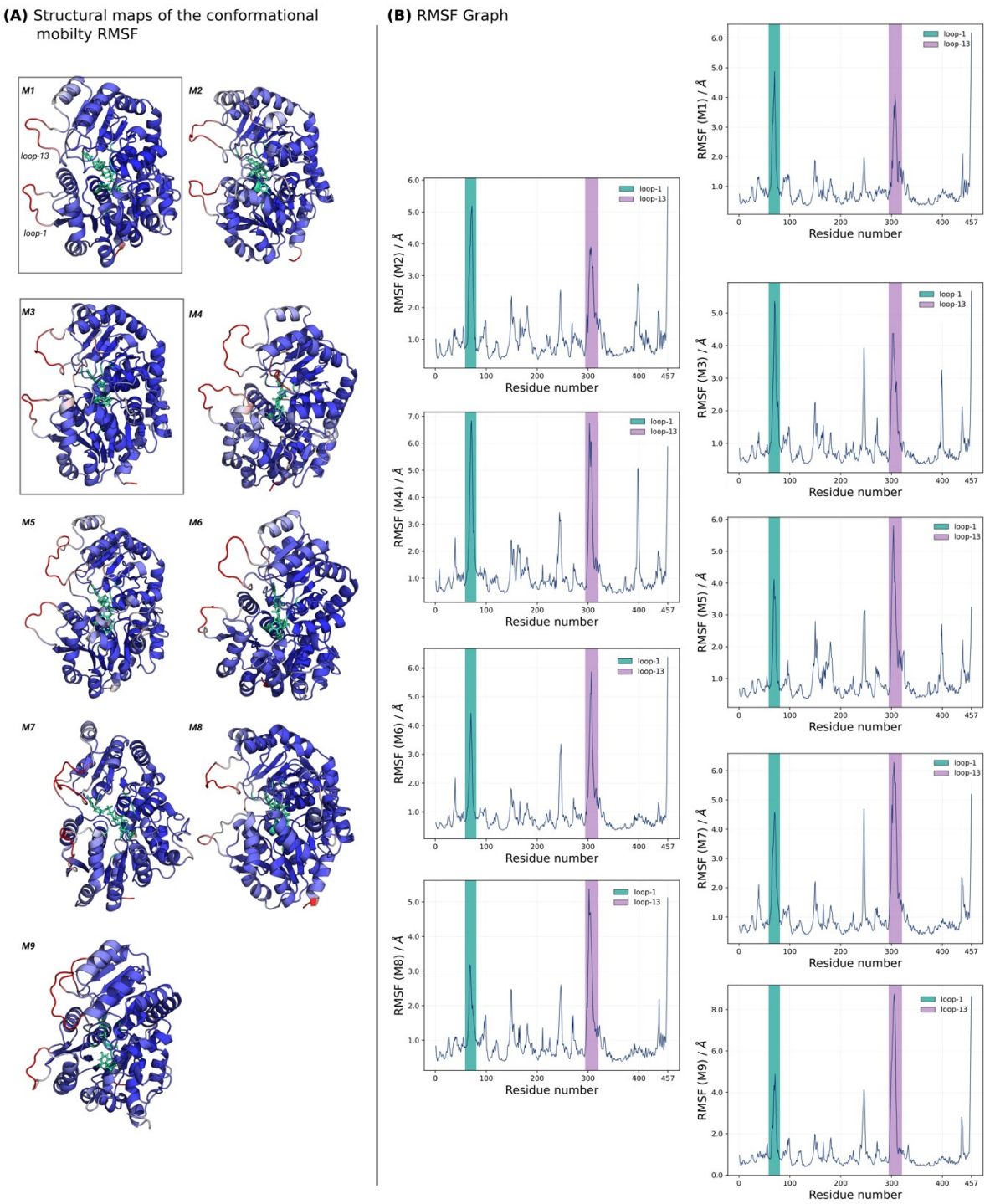

Figure S6. (A) 3D-represenattion of RMSF analysis of the models studied. The most flexible loops are highlighted in red, whereas the less flexible sequences are represented in blue color. **(B)** RMSF graph analysis for the nine models. The most flexible loops are also highligthed.

Loop-1 (Hie63-Phe77) and the loop-13 (Val296-Phe315) turned out to be exposed to the solvent and are included within each one of the Rossmann fold domains separated by the cleft that accommodates the active site. Consequently, it is expected that they exhibit the most flexible behavior along the MD simulations.

2.4 Secondary structure analysis for nine native models

In this section, we examine the secondary structural elements present in nine native C-glycosyltransferase (C-GT) models, focusing on how α-helices, β-sheets, and loop regions contribute to enzyme stability and substrate binding. Validation of these structural components confirmed that the secondary structure remained largely unchanged throughout the molecular dynamics, with observed changes primarily involving direct interactions with the substrate.

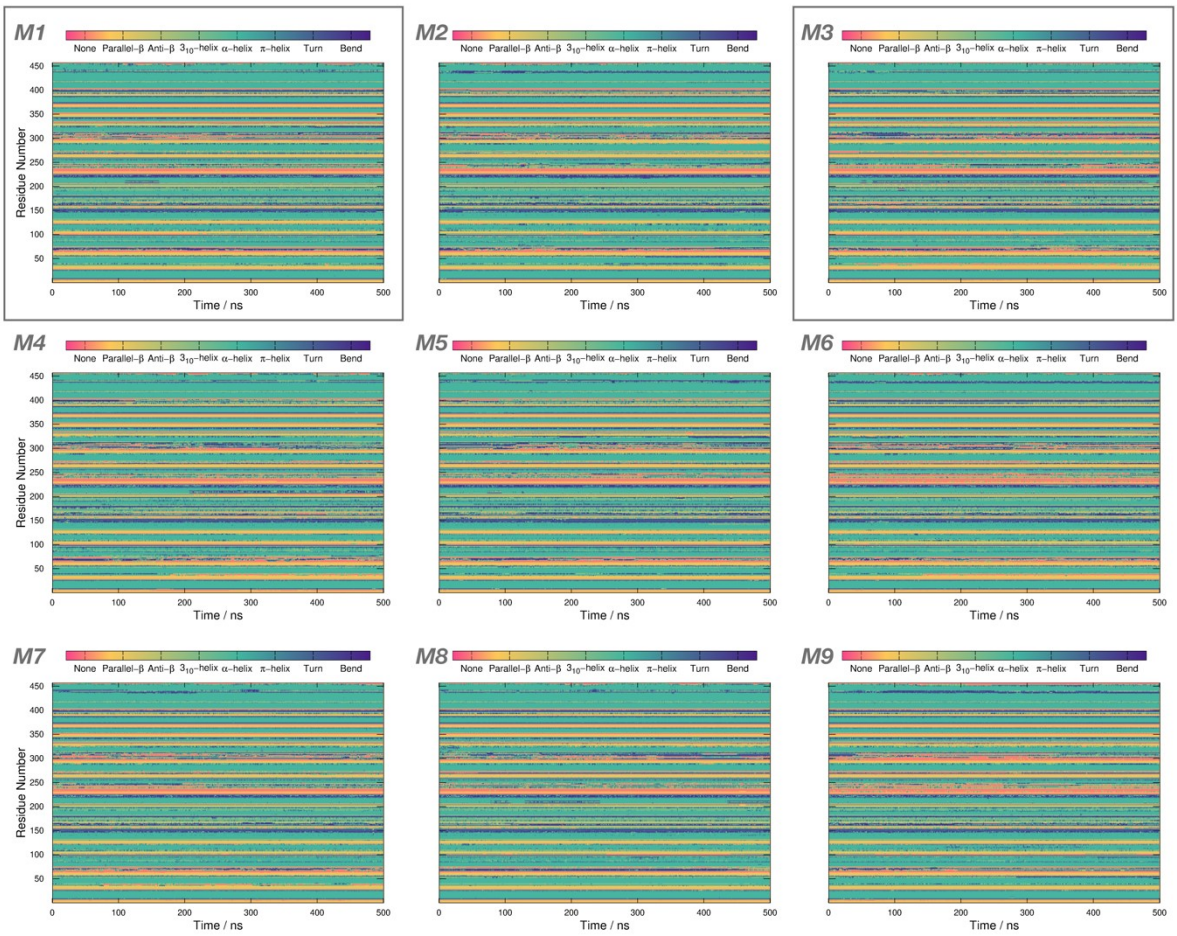

Figure S7. Full protein secondary structure analysis of the nine native models studied.

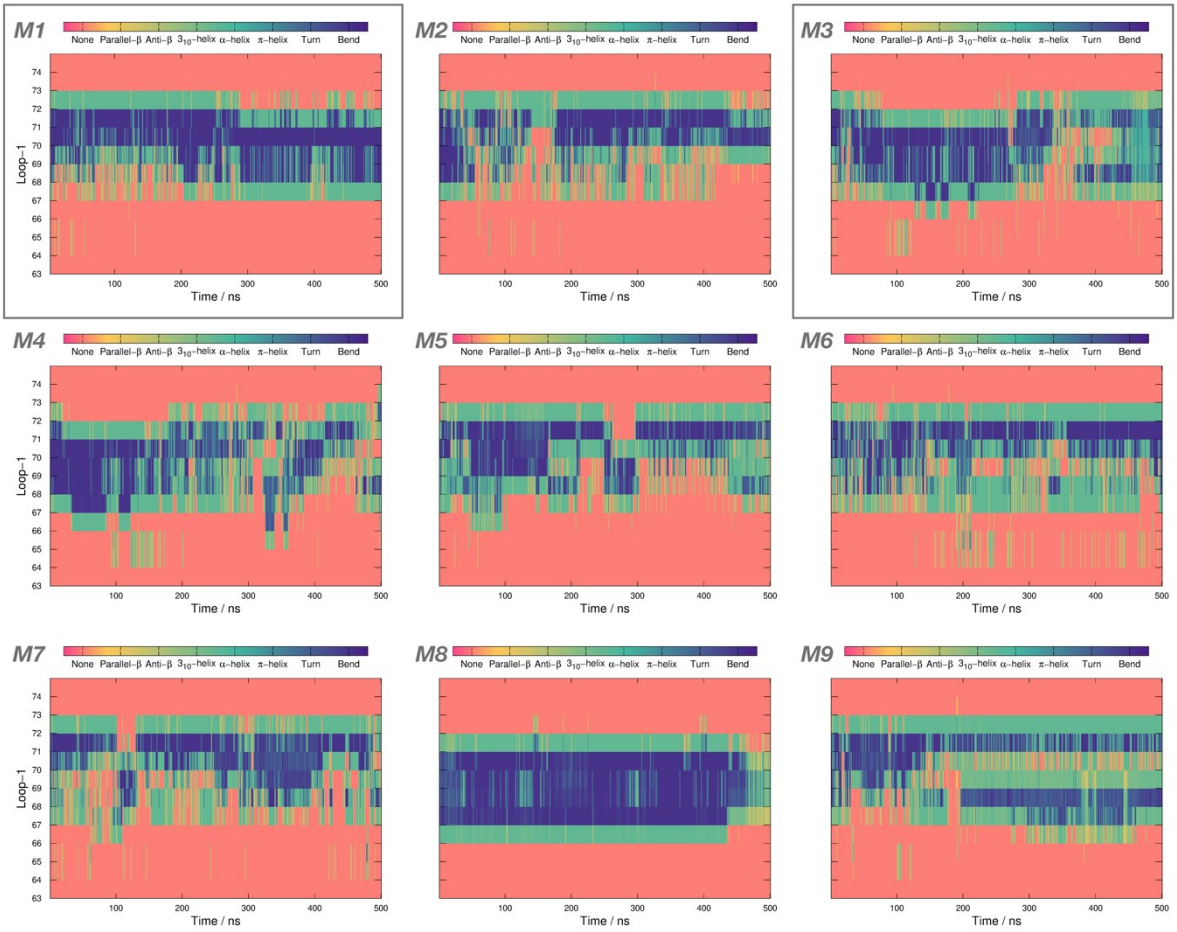

Figure S8. Secondary structure analysis of **loop-1** of the models studied.

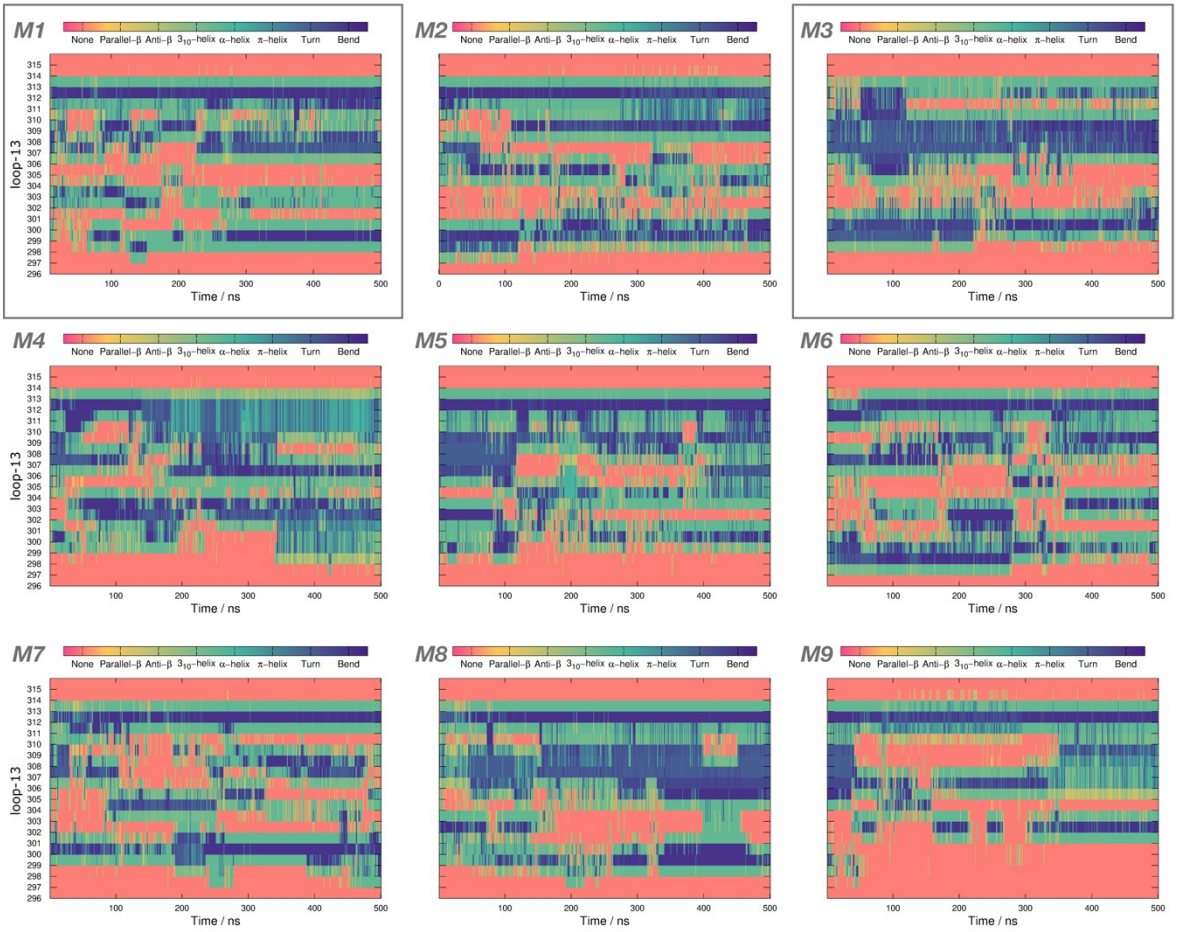

Figure S9. Secondary structure analysis of **loop-13** of the models studied.

2.5 Analyses of Phloretin conformations for nine native models

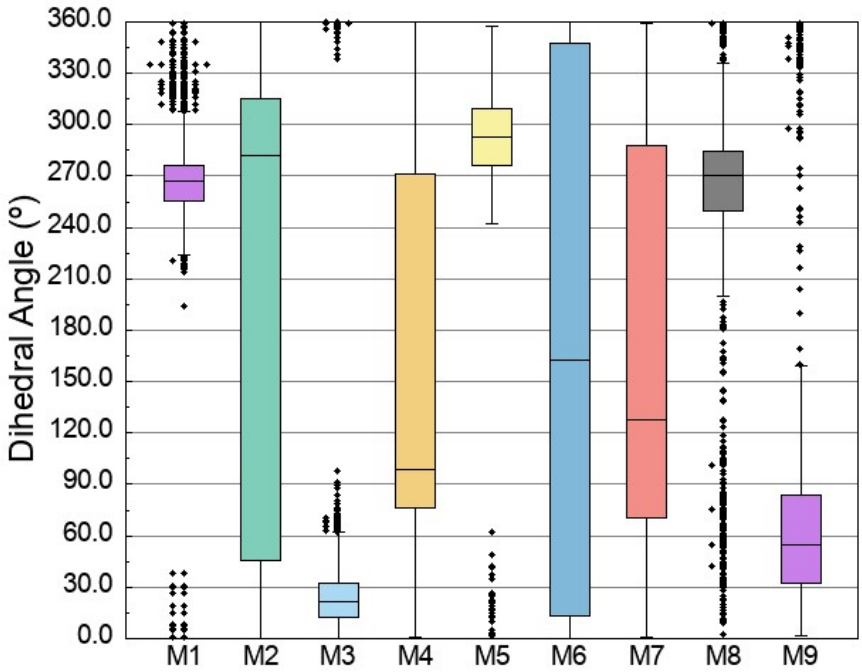

Figure S10. Dihedral angle *abcd* (θ) distribution for nine native models. The θ angle has been represented for the acceptor substrate (Phl) along the 500 ns of MD production.

2.6 Global reactivity indexes.

Table S1 summarizes the results of local and global descriptors on the neutral and anionic forms of Phl in model M1 (extended conformation). The level of theory used is M06-2X/6-31G(d).

Table S1. Some global reactivity descriptors of neutral and anionic forms of Phl in model M1 at the M06-2X/6-31G(d) level of theory. The respective units are defined as follows: I_1, A_1, ω, ω^-, ω^+ and Δω (in $hartree$); μ (in $hartree \cdot e^{-1}$); η (in $hartree \cdot e^{-2}$); *S* (in $hartree^{-1} \cdot e^2$). FMOA provides values of I_1 and A_1 based on Koopmans' theorem.

Global Reactivity Descriptors		**FDA**		**FMOA**	
Name	**Symbol**	**Neutral**	**Anionic**	**Neutral**	**Anionic**
First Vertical Ionization Potential	I_1	*0.29197*	*0.10985*	*0.26082*	*0.07941*
First Vertical Electron Affinity	A_1	*-0.03691*	*-0.15307*	*-0.00746*	*-0.11672*
Electronic Chemical Potential	μ	*-0.12753*	*0.02161*	*-0.12668*	*0.01866*
Molecular Hardness	η	*0.32889*	*0.26292*	*0.26828*	*0.19613*
Global Softness	***S***	***3.04055***	***3.80338***	***3.72745***	***5.09866***
Electrophilicity	ω	*0.02472*	*0.00089*	*0.02991*	*0.00089*
Electron-donating Power	ω^-	*0.13377*	*0.00740*	*0.13992*	*0.00471*
Electron-accepting Power	ω^+	*0.00624*	*0.02901*	*0.01324*	*0.02336*
Net Electrophilicity	**Δω**	***0.14001***	***0.03642***	***0.15317***	***0.02807***

Notice that I_1 and A_1 in the FMOA columns correspond to the frontier molecular orbitals energies under the assumption the Koopmans' theorem is satisfied.

Table S2 summarizes the results of local and global descriptors on the neutral and anionic forms of Phl in model M1 (extended conformation). The level of theory used is M06-2X/6-31G+(d).

Table S2. Some global reactivity descriptors of neutral and anionic forms of Phl in model M1 at the M06-2X/6-31G+(d) level of theory. The respective units are defined as follows: I_1, A_1, ω, ω^-, ω^+ and $\Delta\omega$ (in $hartree$); μ (in $hartree \cdot e^{-1}$); η (in $hartree \cdot e^{-2}$); S (in $hartree^{-1} \cdot e^2$). FMOA provides values of I_1 and A_1 based on Koopmans' theorem.

Global Reactivity Descriptors		FDA		FMOA	
Name	**Symbol**	**Neutral**	**Anionic**	**Neutral**	**Anionic**
First Vertical Ionization Potential	I_1	*0.30030*	*0.12655*	*0.27125*	*0.10020*
First Vertical Electron Affinity	A_1	*-0.01807*	*-0.10394*	*0.01253*	*-0.07003*
Electronic Chemical Potential	μ	*-0.14112*	*-0.01131*	*-0.14189*	*-0.01509*
Molecular Hardness	η	*0.31837*	*0.23050*	*0.25872*	*0.17023*
Global Softness	***S***	***3.14101***	***4.33849***	***3.86518***	***5.87441***
Electrophilicity	ω	*0.03127*	*0.00028*	*0.03891*	*0.00067*
Electron-donating Power	ω^-	*0.15301*	*0.02061*	*0.16493*	*0.01952*
Electron-accepting Power	ω^+	*0.01189*	*0.00931*	*0.02304*	*0.00443*
Net Electrophilicity	$\Delta\omega$	***0.16490***	***0.02992***	***0.18797***	***0.02395***

Table S3 summarizes the results of local and global descriptors on the neutral and anionic forms of Phl in model M3 (packed conformation). The level of theory used is M06-2X/6-31G(d).

Table S3. Some global reactivity descriptors of anionic and neutral conformation of Phl in model M3 at the M06-2X/6-31G(d) level of theory. The respective units are defined as follows: I_1, A_1, ω, ω^-, ω^+ and $\Delta\omega$ (in $hartree$); μ (in $hartree \cdot e^{-1}$); η (in $hartree \cdot e^{-2}$); S (in $hartree^{-1} \cdot e^2$). FMOA provides values of I_1 and A_1 based on Koopmans' theorem.

Global Reactivity Descriptors		FDA		FMOA	
Name	**Symbol**	**Neutral**	**Anionic**	**Neutral**	**Anionic**
First Vertical Ionization Potential	I_1	*0.28514*	*0.11580*	*0.25635*	*0.08702*
First Vertical Electron Affinity	A_1	*-0.04419*	*-0.17463*	*-0.01480*	*-0.14517*
Electronic Chemical Potential	μ	*-0.12047*	*0.02941*	*-0.12078*	*0.02908*
Molecular Hardness	η	*0.32933*	*0.29044*	*0.27115*	*0.23219*
Global Softness	***S***	***3.03647***	***3.44307***	***3.68800***	***4.30682***
Electrophilicity	ω	*0.02203*	*0.00149*	*0.02690*	*0.00182*
Electron-donating Power	ω^-	*0.12489*	*0.00642*	*0.13113*	*0.00362*
Electron-accepting Power	ω^+	*0.00442*	*0.03584*	*0.01035*	*0.03269*
Net Electrophilicity	$\Delta\omega$	***0.12930***	***0.04226***	***0.14148***	***0.03631***

Table S4 summarizes the results of local and global descriptors on the neutral and anionic forms of Phl in model M3 (packed conformation). The level of theory used is M06-2X/6-31G+(d).

Table S4. Some global reactivity descriptors of the anionic and neutral forms of Phl in model M3 at the M06-2X/6-31+G(d) level of theory. The respective units are defined as follows: I_1, A_1, ω, ω^-, ω^+ and Δω (in $hartree$); μ (in $hartree \cdot e^{-1}$); η (in $hartree \cdot e^{-2}$); S (in $hartree^{-1} \cdot e^2$). FMOA provides values of I_1 and A_1 based on Koopmans' theorem.

Global Reactivity Descriptors		**FDA**		**FMOA**	
Name	**Symbol**	**Neutral**	**Anionic**	**Neutral**	**Anionic**
First Vertical Ionization Potential	I_1	*0.29390*	*0.13156*	*0.26713*	*0.10665*
First Vertical Electron Affinity	A_1	*-0.02321*	*-0.12403*	*0.01196*	*-0.09669*
Electronic Chemical Potential	μ	*-0.13535*	*-0.00376*	*-0.13955*	*-0.00498*
Molecular Hardness	η	*0.31710*	*0.25559*	*0.25517*	*0.20334*
Global Softness	***S***	***3.15354***	***3.91253***	***3.91896***	***4.91787***
Electrophilicity	ω	*0.02888*	*0.00003*	*0.03816*	*0.00006*
Electron-donating Power	ω^-	*0.14526*	*0.01791*	*0.16203*	*0.01532*
Electron-accepting Power	ω^+	*0.00991*	*0.01415*	*0.02249*	*0.01034*
Net Electrophilicity	**Δω**	***0.15517***	***0.03206***	***0.18452***	***0.02566***

Global softness (S) calculations suggest that the basic extended form (S = 4.339) is approximately 11% more reactive than the basic packed form (S = 3.913), revealing a greater concentration of electron density in the extended conformation. This greater electron distribution could potentially facilitate the nucleophilic character of Phl, favoring the catalytic process. Both basis sets, 6-31G+(d) and 6-31G(d), show the same tendency, but values discussed above refer to the M06-2X/6-31G+(d) level of theory.

The most important conclusion is that both conformers exhibit reactivity toward electrophilic agents. Therefore, based on these observations, the extended (M1) and packed (M3) conformations could be candidates to efficient C-glycosylation.

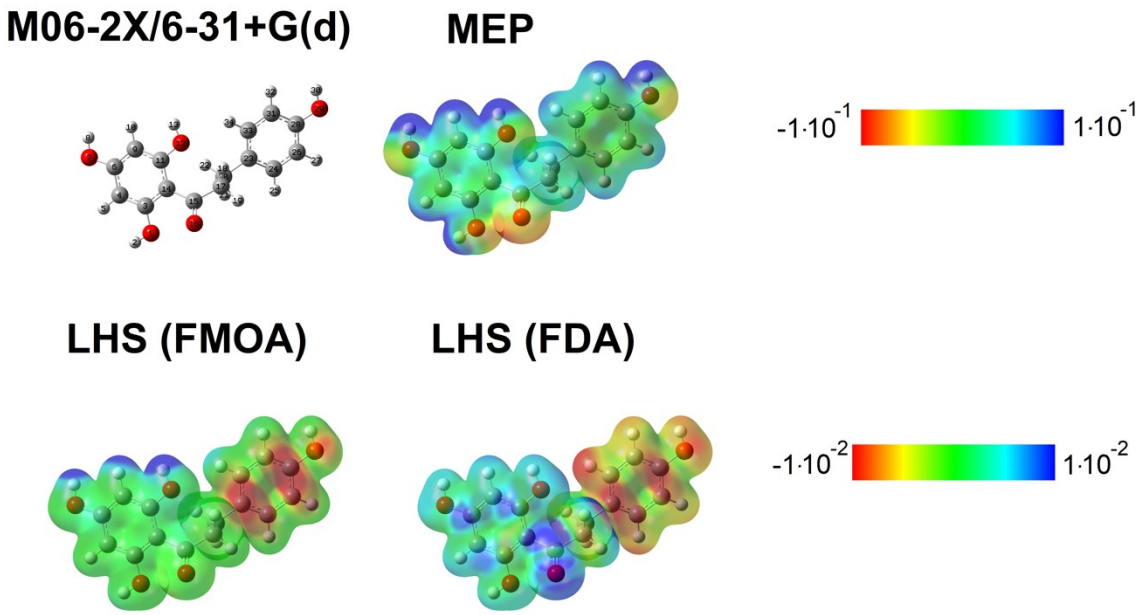

Figure S11. Representations of MEP and LHS descriptors at the M06-2X/6-31+G(d) level of theory for the neutral extended form of Phl in model M1. LHS was computed by two methods: Finite difference approximation (FDA) and frontier molecular orbital approximation (FMOA). MEP is expressed in atomic units ($hartree \cdot e^{-1}$) ranging from $-1 \cdot 10^{-1}$ to $1 \cdot 10^{-1}\ hartree \cdot e^{-1}$; LHS is expressed in atomic units ($hartree^{-2} \cdot e^{3} \cdot bohr^{-3}$) ranging from $-1 \cdot 10^{-2}$ to $1 \cdot 10^{-2}\ hartree^{-2} \cdot e^{3} \cdot bohr^{-3}$.

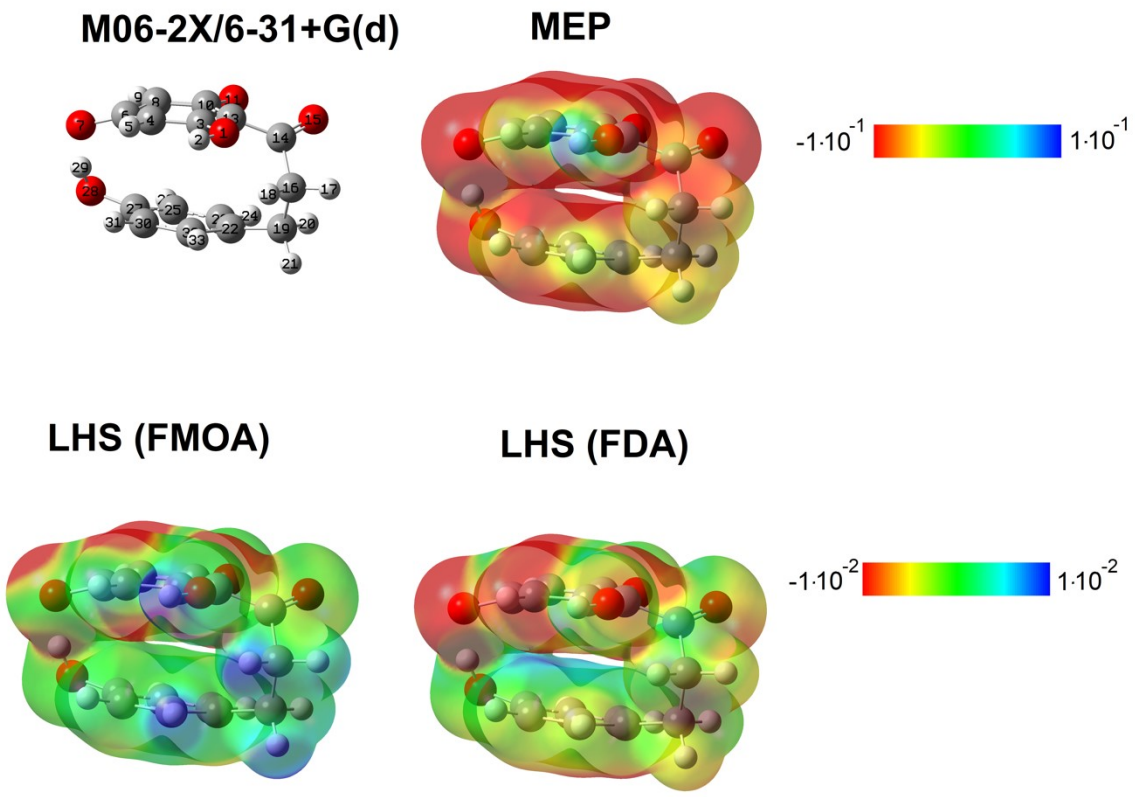

Figure S12. Representations of MEP and LHS descriptors at the M06-2X/6-31G+(d) level of theory for the packed neutral form of Phl in model M3. LHS was computed by two methods: Finite difference approximation (FDA) and frontier molecular orbital approximation (FMOA). MEP is expressed in atomic units ($hartree \cdot e^{-1}$) ranging from $-1 \cdot 10^{-1}$ to $1 \cdot 10^{-1}\ hartree \cdot e^{-1}$; LHS is expressed in atomic units ($hartree^{-2} \cdot e^{3} \cdot bohr^{-3}$) ranging from $-1 \cdot 10^{-2}$ to $1 \cdot 10^{-2}\ hartree^{-2} \cdot e^{3} \cdot bohr^{-3}$.