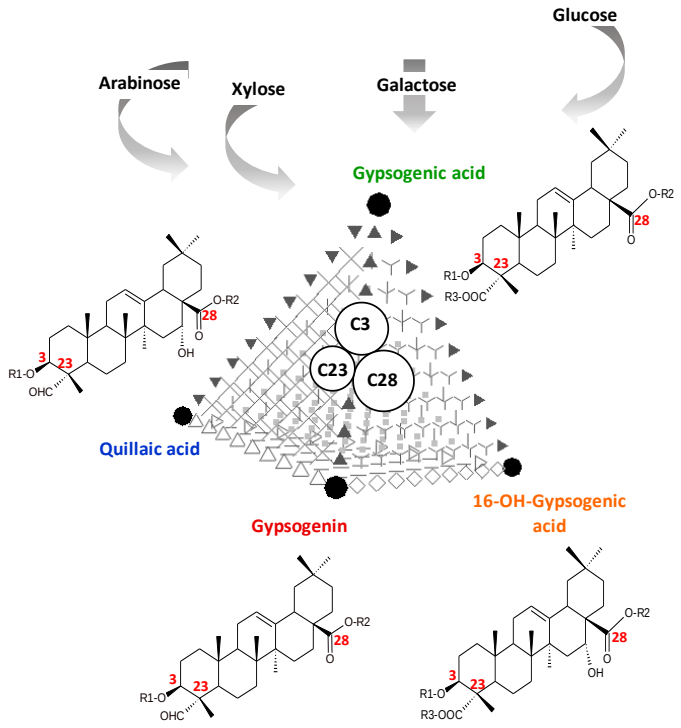
MOL2NET, International Conference Series on Multidisciplinary Sciences
Insert the title of the workshop here

Highlighting glycosylation ways in Caryophyllaceae saponins by simplex simulation approach

Hammami Asma [a], *Farman Muhammad* [b], *Semmar Nabil* [c*]

[a] *Institut National des Sciences Appliquées et de Technologies (INSAT), Université de Carthage, Tunisia*
[b] *Quaid-i-Azam University, Department of Chemistry, Islamabad 45320, Pakistan*
[c] *Laboratory of BioInformatics, bioMathematics and bioStatistics (BIMS), Institut Pasteur de Tunis, 1002, Université de Tunis El Manar, Tunis, Tunisia*
[*] *For correspondence: nabilsemmar5@gmail.com*

Graphical Abstract



Abstract. Glycosylation mechanisms in saponins of Caryophyllaceae plant family were subjected to simulation by statistically exploring variability of 231 chemical structures belonging to four different aglycones: gypsogenin (*Gyp*), quillaic acid (*QA*), gypsogenic acid (*GA*), 16-OH-gypsogenic acid (16-OH-*GA*). Saponins based on different aglycones were initially characterized by relative glycosylation levels of different carbons. Simulation was initialized by combining the four saponin groups using Scheffé's mixture design which provides a complete set of *N* gradual weightings of groups. Combined saponins were randomly and iteratively sampled from different groups by bootstrap technique. For a same combination, saponins were averaged leading to barycentric glycosylation profile. Iterations of the *N* barycentric profiles and averaging provided a final response matrix of *N* smoothed glycosylation profiles from which regulation mechanisms of carbons were highlighted in different aglycone-based saponins. Glucose (*Glc*) was revealed to be widely favored in *GA* and 16-OH-*GA* with more target aspect of 28-*Glc* in 16-OH-*GA* and relatively shared distribution between C28 (mainly) C3 and C23 in

	GA. Strong competition for galactose (*Gal*) was highlighted between C3 and C28 with target aspects to 28-*Gal* in *GA* and 3-*Gal* in (*Gyp*, *QA*). *Gyp* and *QA* showed higher regulations of pentoses (xylose, *Xyl*; arabinose, *Ara*) with more affinity of *GA* for (3-*Ara*, 28-*Xyl*) and 16-OH-*GA* for (3-*Xyl*, 28-*Ara*). These results call for further investments in simulations of glycosylation mechanisms helping for better understanding metabolic aspects of saponins, and encouraging future analytic experiments in the field.

Introduction. Saponin synthesis is essentially based on glycosylation of sapogenin at different candidate carbons by different saccharide types with different levels. Regulation mechanisms of saponin synthesis remain limited because of the few investments in analytical experiments concerning enzymatic action modes, steric effects and regioselectivity (**Tiwari et al., 2016**; **Moses et al., 2013**; **Meesapyodsuk et al., 2007**; **Haralampidis et al., 2002**). Such mechanisms could be alternatively approached by computational way consisting in statistically exploring structural variability of saponins to highlight regulation trends between glycosylated carbons (**Sarraj-Laabidi et al., 2018**). For that aim, simulation simplex approach was applied to a bibliographic set of 231 Caryophyllaceae saponins to highlight aglycone- and carbon-dependent glycosylation mechanism for different saccharides.

Materials and Method. A whole set of 231 saponins was initially organized into 4 aglycone subsets: *Gyp*, *QA*, *GA*, 16-OH-*GA* (79, 80, 52, 20 saponins, respectively) (**Cheikh Ali et al., 2019**). Saponins were characterized by their glycosylation (*Gly*) regulation profiles containing relative occurrence degrees of different saccharides at different carbons (**Sarraj et al., 2018**). *Gly* profiles of saponins were iteratively combined by applying Scheffé's mixture design giving a complete set of *N* gradual weights for the *q* (=4) groups (**Scheffé, 1958**). At the output of N iterated combinations, *N* barycentric *Gly*-profiles were calculated resulting in smoothed data that were used for graphical analysis of regulation trends between different saccharides at different carbons under different aglycones (**Fig. 1**).

Results. *Glc* strongly characterized *GA* and 16-OH-*GA* saponins whereas *Gyp* and *QA* showed higher diversification of glycosylation at the expense of *Glc*. *GA* and 16-OH-*GA* were functionally differentiated by distributional and target aspects of *Glc*, respectively (**Fig. 1**). In *GA*, *Glc* was shared between C28 (60%), C3 (12%) and C23 (5%), whereas in 16-OH-*GA*, *Glc* was mainly target to C28 (80%). *Gyp* and *QA* were less concerned with *Glc* with low regulation levels (< 3%). This could indicate interactions between aglycone type and carbon position on regulation levels of *Glc*. Concerning *Gal*, *GA* showed relatively higher trend of 28-*Gal* whereas 16-OH-*GA* did not favor galactosylation at all the carbons. However, *Gyp* and *QA* showed target mechanism of 3-*Gal* at the expense of 28-*Gal*. Pentoses (*Xyl*, *Ara*) manifested in *Gyp* and *QA* vs negligible levels in *GA* and 16-OH-*GA*. *Gyp* and *QA* were differentiated by glycosylation positions: 28-*Xyl* and 3-*Ara* were relatively more favored in *Gyp*, vs 28-*Ara* and 3-*Xyl* in QA.

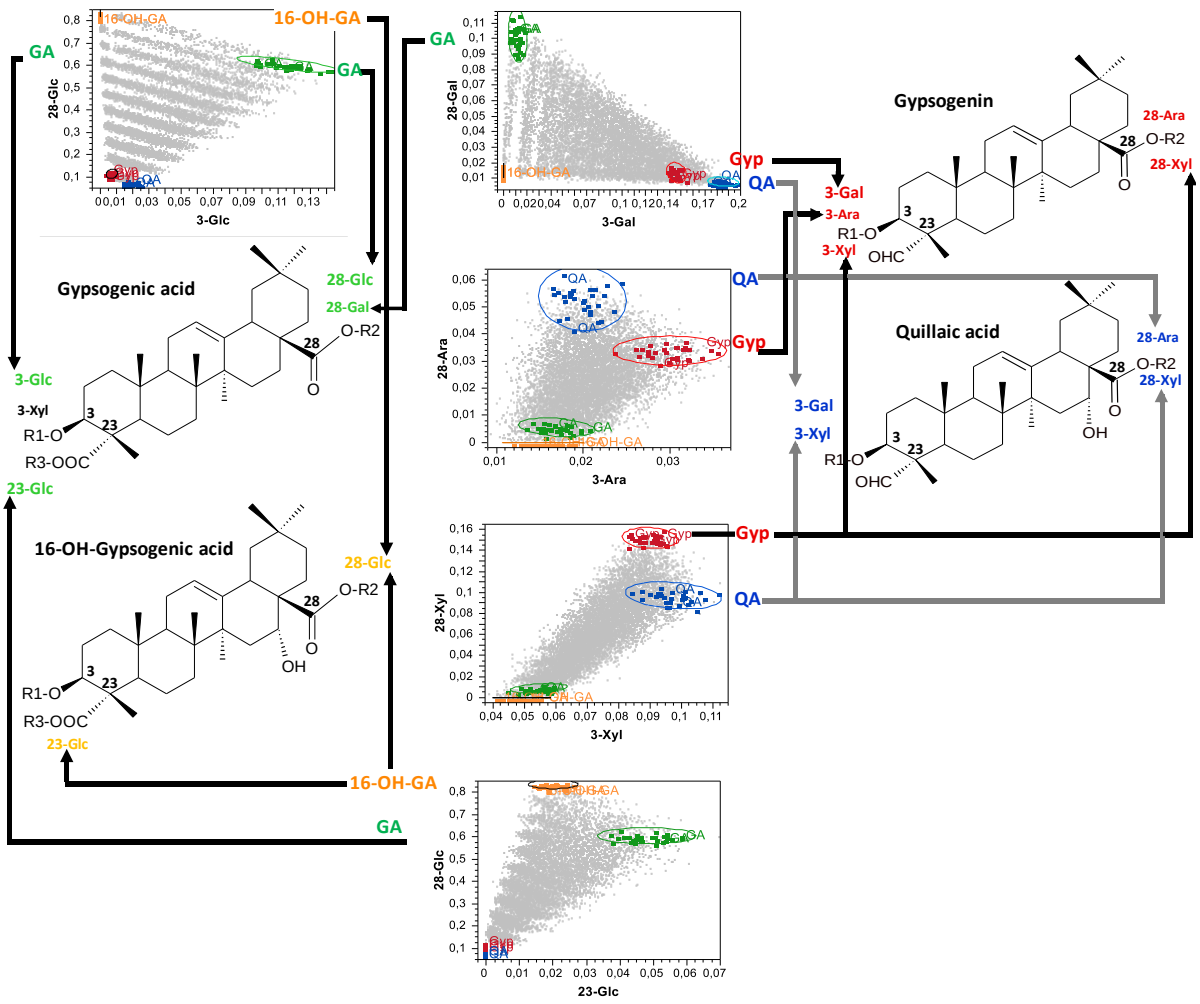


Figure 1. Smoothed plots by simplex approach highlighting regulation trends of different saccharides between different carbons in different aglycones (gypsogenin, quiallic acid, gypsogenic acid, 16-OH-gypsogenic acid) of Caryophyllaceae saponins.

Conclusion. Simulation results could be indicative of differential regioselectivities concerning different saccharide types varying at both inter-molecular (between sapogenins) and intra-molecular (between carbons) scales.

References

Cheikh Ali, S., Farman, M., Lacaille-Dubois M.A., Semmar, N., **2019**. Structural organization of saponins in Caryophyllaceae. *Phytochemistry Reviews* 18(2), 405-441.

Haralampidis, K. Trojanowska, M., Osbourn, A.E., **2002**. *Advances in Biotechnological Engineering & Biotechnology*, 75, 31-49.

Meesapyodsuk, D., Balsevich, J., Reed, D.W., Covello, P.S., **2007**. Saponin Biosynthesis in Saponaria vaccaria. cDNAs Encoding b-Amyrin Synthase and a Triterpene Carboxylic Acid Glucosyltransferase. *Plant Physiology*, 143, 959–969.

Moses, T., Pollier, J., Thevelin, J.M., Goossens, A., **2013**. Bioenggineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytologist* 200, 27-43.

Sarraj-Laabidi, A., Messai, H., Hammami-Semmar, A., Semmar, N., **2017**. Chemometric analysis of inter- and intra-molecular diversification factors by a machine learning simplex approach. A review and research on Astragalus saponins. *Current Topics in Medicinal Chemistry*, 17(25), pp 2820-2848.

Scheffé, H. 1958. Experiments with mixtures. J. R. Stat. Soc. B, 20, 344–360.

Tiwari, P., Sangwana, R.S., Sangwan, N.S., **2016**. Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnology Advances* 34(5), 714-739.