SciForum
MOL2NET

Chemometrical analysis of structure-structure and structure-activity trends of cycloartane-based saponins in *Astragalus* genus

SARRAJ-LAABIDI Abir[1, 2], SEMMAR Nabil[1, *]

[1] Université de Tunis El Manar, Institut Pasteur de Tunis, Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), 1002, Tunis, Tunisia; sarraj.abir@live.fr; nabilsemmar@yahoo.fr

[2] Université de Tunis El Manar, Faculté des Sciences de Tunis, Campus Universitaire, 2092 Tunis, Tunisia; sarraj.abir@live.fr

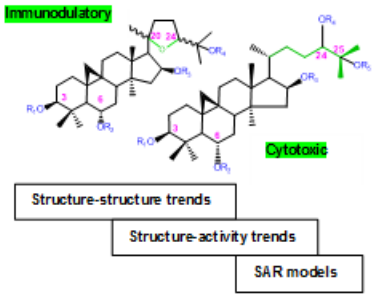
***** Author to whom correspondence should be addressed; nabilsemmar@yahoo.fr;

Abstract: *Astragalus* genus represents the widest terrestrial plant taxon with more than 2200 species of herbs or shrubs. Under phytochemical aspect, this genus was characterized by a high structural diversity of saponins essentially based on cycloartane. The high number of saponins offers a strong basis for analysis of structural properties and metabolic trends governing molecular synthesis and diversity. Such trends can be highlighted from significant correlations between chemical substitution types/positions and aglycone forms. Beside the high number of chemical structures, pharmacological activities of saponins provide another variation aspect which was less invested because of not systematic evaluations of elucidated molecules. Despite this disproportion constraint between pharmacological evaluations and structural elucidations, preliminary significant structure-activity (SA) trends can be highlighted using appropriate statistical tools. This work focused on statistical analysis of structure-structure (SS) and structure-activity (SA) trends in *Astragalus* saponins by a sequential way including detections, significance evaluations and predictions. Dataset concerned 193 cycloartane-based saponins including 35 evaluated ones shared between cytotoxic and immunomodulatory activities (the most published activities in *Astragalus* saponins). SS and SA trends were initially highlighted by correspondence analysis and their significances were evaluated by Fisher's exact test. Results revealed significant affinities between aglycone forms and glycosylation positions. Moreover, both cycloartane forms and glucosylation positions showed significant effects on considered pharmacological activities. Finally, using the significantly influent structural variables, SAR models were developed by logistic regressions. Obtained models showed high sensitivity and specificity in favor of good predictability and distinctness of each separated activity. These results remain preliminary and need more confirmation from more pharmacological data that could be cumulated in the future.

.
.
.

Keywords: *Cytotoxic activity, immunomodulatory activity, structure-activity trends, correspondence analysis, Fisher's exact test, logistic regression*

Graphical Abstract:



.
.

Introduction: Saponins represent a wide family of secondary metabolites showing high structural diversity. Although new structures are subject to continuous elucidations, pharmacological activities are not systematically evaluated leading to much more not evaluated molecules than evaluated ones. Such a state can be represented by two separated subsets concerning saponins with known (evaluated) and unknown (not evaluated) activities, respectively. Despite such a disproportion between pharmacologically evaluated and not evaluated saponins, wide phytochemical data cumulated in literature offer high structural variability that can be statistically analyzed to highlight preliminary significant associations between structural traits and activities.

This work concerned statistical analysis of a wide dataset of *Astragalus* saponins (cumulated in literature) by focusing on link analysis between structural traits and cytotoxic and immunomodulatory activities [1].

Materials and Methods: The aim of the current work was based on the following question: how the small set of known active saponins can be structurally distinguished from the wide set of not evaluated ones? This question found responses through three sequential statistical analyses including (1) structure-activity (SA) trends detections, (2) significance evaluations and (3) SA-predictive models (**Figure 1**).

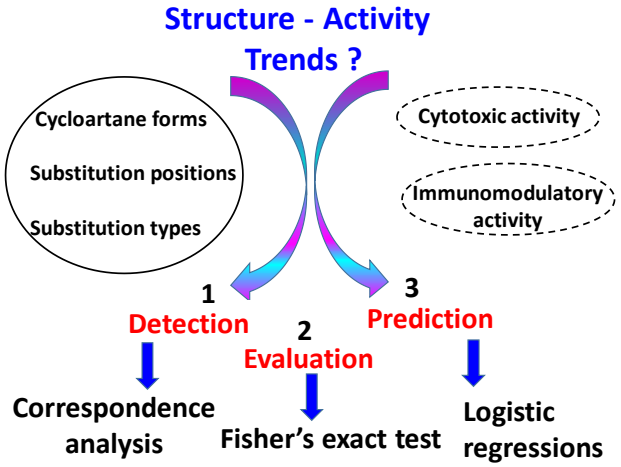


Figure 1. Three methodological steps for statistical highlighting (1), significance evaluations (2) and prediction (3) of structure-activity trends applied for cycloartane-based saponins in *Astragalus* genus.

Detection of SA trends was carried out by correspondence analysis (CA) applied on a dataset containing 178 cycloartane-based saponins in rows and chemical substitutions of carbons in columns [2]:

For rows, saponins were initially identified by their cycloartane forms including 20,24-epoxyxyloartane (*Ep1*), 20,25-epoxcycloartane (*Ep2*) and cycloartane with aliphatic lateral chain (*LCh*) (**Figure 2**).

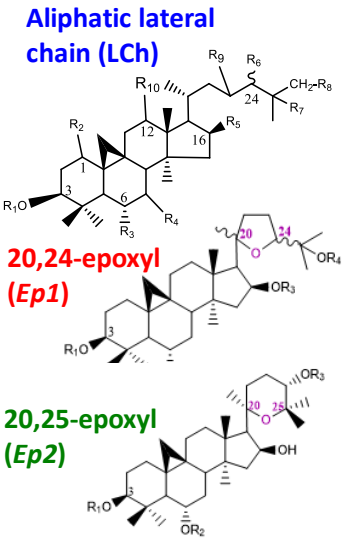


Figure 2. Chemical structures of different cycloartane forms in *Astragalus* saponins.

Also, saponins were characterized by two indicative variables concerning evaluated cytotoxic (*Cyt*) and immunomodulatory (*Imn*) activities, respectively. In all, *Cyt* and *Imn* were represented by 35 molecules among the 178 ones. Chemical substitution concerned carbons C3, C6, C16, C24, C25 susceptible to attach hydroxyl, glycosyl and/or acetyl groups. Glycosyls included glucosyl, xylosyl, rhamnosyl, arabinosyl, apiosyl and glucuronic acid. Other carbons were not considered because of their rare chemical substitutions leading to outlier cases.
SA trends highlighted from factorial plots of CA were statistically evaluated by means of Fisher's exact test (*FET*) [3]. In this link test, well-known evaluated saponins were considered as a target set the characteristics of which were confronted to the global state of random set containing all the not evaluated molecules (**Figure 3**). Randomness was attributed to this second set because it can include active and not active molecules the pharmacology of which remains unknown by waiting confirmative evaluations.

Finally, a synthesis of significant SA trends was carried out by applying two logistic models on the subset of 35 *Cyt* or *Imn* molecules to predict each activity in relation to the most discriminant structures given by *FET* (**Figure 1**) [3].
CA was carried out by ADE statistical software [4]. *FET* and logistic regressions were applied using JMP statistical software [5].

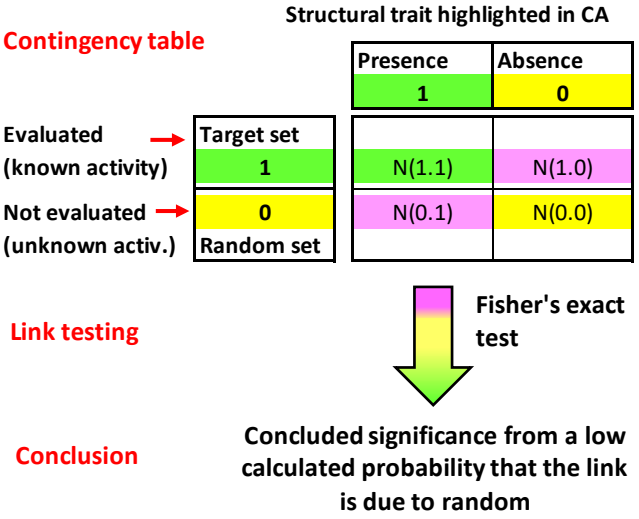


Figure 3. Principle of Fisher's exact test used for evaluating significances of SA trends highlighted in CA.

Results and Discussion: Strong structure-structure (SS) associations in *Astragalus* saponins were highlighted by the second factorial plan (*F3F4*) concerning both *LCh* and *Ep1*: massive *LCh* points projected in the same subspace ($F3^+F4^+$) than the 16-*Glc* variable (**Figures 4b, 4c**); this positive association was confirmed by a low *p*-value (=0.002) in *FET*. Concerning *Ep1*, it was relatively more characterized by 6-xylosylation (6-*Xyl*) than *LCh* form as shown in the upper left quadrant $F3^-F4^+$ **Figures 4b, 4c**; such a metabolic affinity between *Ep1* and *6-Xyl* was confirmed by low *p*-value in *FET* ($p = 0.01$). Apart from SS associations, different SA trends were highlighted by CA. Concerning cytotoxicity, CA highlighted topological proximity between *Cyt* and C3-glycosylation (3-*Glc*) points (plan *F1F2*) (**Figure 4a**) and superimposition between *Cyt* and *LCh* (plot *F7F8*) (**Figure 4d, e**). This indicated some positive trends between *Cyt* and both 3-*Glc* and *LCh*. Along the eighth principal component (*F8*), *Cyt-LCh* association showed topological opposition to *Imn-Ep1* one; this later indicated some positive trend between immunomodulatory activity and the 20,24-epoxycycloartane (**Figure 4d, e**). This was also confirmed by projections of *Imn* and *Ep1* points in a same subspace in *F3F4* plot (**Figure 4b, c**). Moreover, *Imn* projected close to 6-*Glc* in *F3F4* indicating some association between C6-glucosylation and this activity (**Figure 4b**).

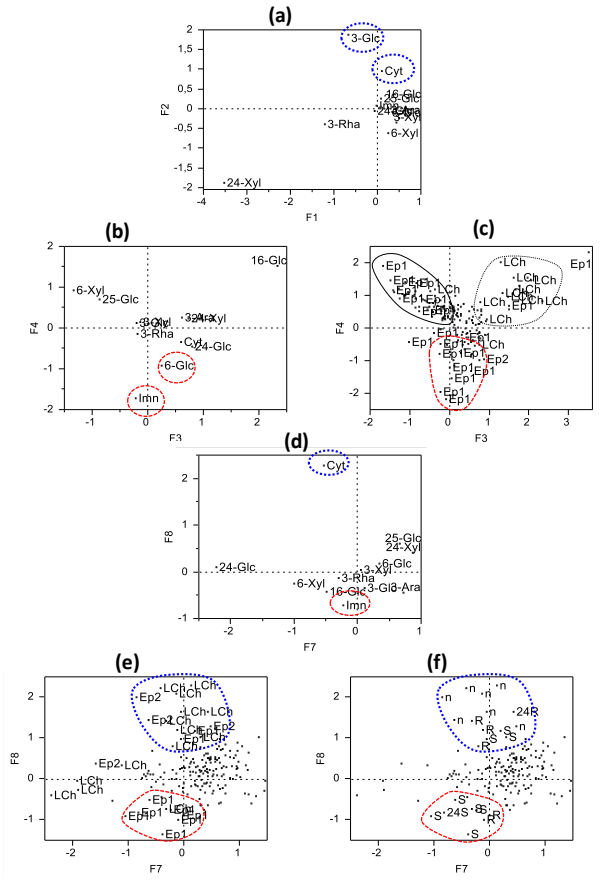


Figure 4. Factorial plots given by correspondence analysis and highlighting trends between *Cyt* or *Imn* activities and structural traits of *Astragalus* saponins.

Apart from cycloartane forms and glucosylation positions, spatial configuration *R* and *S* of cycloartane showed opposite projections in subspaces occupied by *Cyt* and *Imn*, respectively (**Figure 4f**). This could indicate some implication of aglycone configuration in pharmacological activity.

For synthesis, relative occurrences of different structural traits were calculated for the subsets of *Cyt* and *Imn* by reference the whole set of all the saponins: structural profiles of *Cyt* and *Imn* showed well-distinct even opposite aspects (**Figure 5**).

After FET application on all SA trends highlighted in CA, the lowest *p*-values of positive effects on *Cyt* concerned interaction between *LCh* and 3-*Glc* ($p = 5.10^{-4}$) (**Table 1**). Concerning *Imn*, the most significant positive effect resulted from interaction between *Ep1* and 6-*Glc* ($p = 6.10^{-4}$) (**Table 1**).

Using these four most significant and interactive variables given by *FET* (*LCh*, *Ep1*, 3-*Glc*, 6-*Glc*), logistic regressions were applied to develop SA models predicting *Cyt* and *Imn* activities. Both models showed high sensitivity (*Ss*) and specificity (*Sp*) (**Figure 6**): *Ss* = 85.7% for *Cyt* vs 90.0% for *Imn*; *Sp* = 76.5% for *Cyt* vs 81.8% for *Imn*. These results were in favor of good predictive and distinctive ability of both models of different structure-activity subsets.

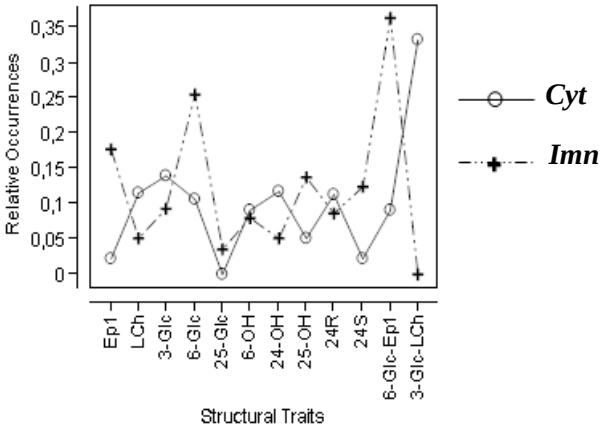


Figure 5. Two profiles showing relative occurrences of different chemical substitutions in saponins showing *Cyt* and *Imn* activities.

The synthesis of these preliminary results highlighted effects of aglycone form and glycosylation position on *Cyt* and *Imn*. For *Cyt*, Verotta et al. (2001) [6] evoked not significant cytotoxic activity in 20,24-epoxyxloartane (*Ep1*). Interaction effect of *LCh* and 3-*Glc* was in agreement with other works on saponins of not *Astragalus* species revealing key roles of aglycone and glycosylation in cytotoxic activity [7]. For *Imn*, previous works on some *Astragalus* saponins evoked positive implications of 20,24-epoxycycloartane and *6-Glc* in *Imn* activity compared to *LCh* and not glucosylated C6, respectively [8, 9].

Conclusions: This work concerned a preliminary analysis of SA trends from updated data of *Astragalus* cycloartane-based saponins cumulated in literature. Although, the current results remain preliminary because of the limited number of evaluated molecules, the method provided a sequential statistical way to extract significant information on SA trends despite sparse states of phytochemical-pharmacological data. Interaction

between aglycone (cycloartane) form and glycosylation (glucosylation) position seemed to be crucial for saponins' activities (**Figure 7**).

This three steps-method can be applied to larger datasets (with more available pharmacological evaluations) to confirm and/or improve knowledge on SA links of saponins and other metabolic families.

Table 1. Results of Fisher's exact test concerning significance evaluations of structure-activity trends previously highlighted by correspondence analysis. Legend: - effect, + effect: negative and positive effects, respectively.

Activity	Influencing structural traits				*p* -values	
					- effect	+ effect
Cytotoxic	*LCh*					0.002
	Ep1				0.01	
		3-Glc				0.063
			24-OH			0.094
				24R		0.033
				24S	0.033	
	LCh	***3-Glc***				$\mathbf{5.10^{-4}}$
			24-OH	*24R*		0.004
		3-Glc		*24R*		0.036
	LCh			*24R*		0.013
	LCh	*3-Glc*		*24R*		0.020
Immuno-modulatory	*Ep1*					0.004
		6-Glc				$\mathbf{6.10^{-4}}$
	Ep1	***6-Glc***				$\mathbf{6.10^{-4}}$
		6-Glc	*25OH*			0.006
	Ep1	*6-Glc*	*25-OH*			0.001

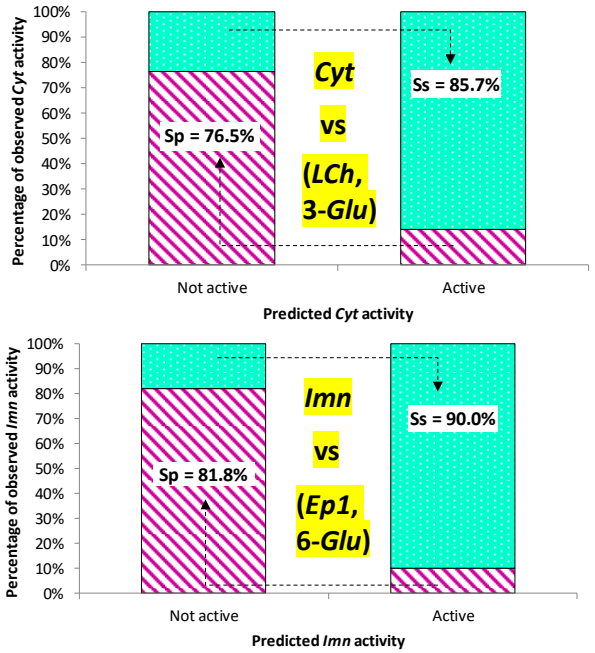


Figure 6. Results given by logistic regression models which were sapplied to predict *Cyt* and *Imn* activities in relation to cycloartane forms (*LCh*, *Ep1*) and glucosylation positions (3-*Glc*, 6-*Glc*). Legend: *Ss*, sensitivity; *Sp*, specificity.

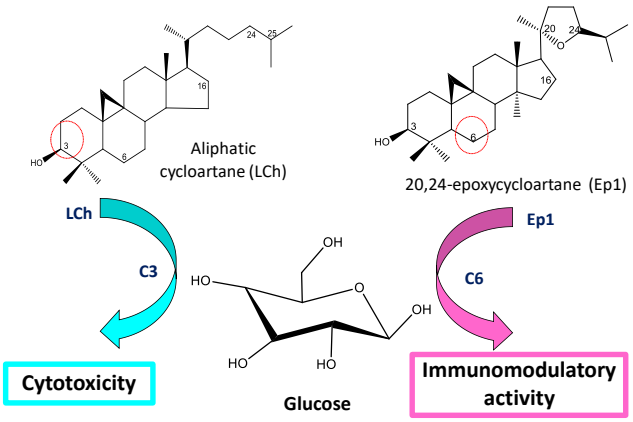


Figure 7. Preliminary results concluded from the three statistical steps-based method consisting of detection, evaluation and prediction of structure-activity trends applied to *Astragalus* saponins.

.

References:

1. Ionkova, I.; Shkondrov, A.; Krasteva, I.; Ionkov, T. Recent progress in phytochemistry, pharmacology and biotechnology of *Astragalus* saponins. *Phytochem. Rev.* **2014**, *13*, 343–374.
2. Greenacre, M. *Correspondence analysis in practice*; CRC press, 2007.
3. Selvin S. *Practical Biostatistical Methods*. Duxbury Press, Belmont, 1995.
4. Chessel, D.; Dolédec, S.; Thioulouse; J., Olivier, J.M. ADE-4 Software, University of Lyon I, Lyon, **1998**.
5. SAS Institute Inc, JMP 8, Cary, NC, USA: SAS Institute, **2008**.

6. Verotta, L.; El-Sebakhy, N.A. Cycloartane and oleanane saponins from *Astragalus* sp. In: *Studies in Natural Products Chemistry 25*, Atta-ur-Rahman (Ed.). Elsevier, UK, **2001**, 179–234.
7. Man, S.; Gao, W.; Zhang, Y.; Huang, L.; Liu, C. Chemical study and medical application of saponins as anti-cancer agents. *Fitoterapia* **2010**, *81*:703–714.
8. Bedir, E.; Pugh, N.; Calis, I.; Pasco, D.S.; Khan, I. A. Immunostimulatory effects of cycloartane-type triterpene glycosides from *Astragalus* species. *Biol. Pharm. Bull.* **2000**, *23*, 834–837.
9. Yesilada, E.; Bedir, E.; Çalış, İ.; Takaishi, Y.; Ohmoto, Y. Effects of triterpene saponins from *Astragalus* species on in vitro cytokine release. *J. Ethnopharmacol.* **2005**, 96: 71–77.