Linear regression analysis of silychristin A, silybin A and silybin B contents in *Silybum marianum*

Sameh F. AbouZid[(1),*], Hayam S. Ahmed[(1)], Abd-El Mageed A. Abd El Mageed[(1)], Abeer S. Moawad[(1)], Asmaa I. Owis[(1)], Shao-Nong Chen[(2,3)], Amandine Nachtergael[(3)], James B. McAlpine[(3)], J. Brent Friesen[(3,4)], Guido F. Pauli[(2,3)]

[(1)]Department of Pharmacognosy, Faculty of Pharmacy, Beni-Suef University, Beni-Suef 62111, Egypt

[(2)]UIC/NIH Center for Botanical Dietary Supplements Research, and [(3)]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood St., M/C 781, Chicago, IL 60612, United States

[(4)]Physical Sciences Department, Rosary College of Arts and Sciences, Dominican University, 7900 West Division Street, River Forest, IL 60305, United States

*Corresponding author

Tel.: +2 082 2317950; Fax: +2 082 2317958

E-mail address: sameh.zaid@pharm.bsu.edu.eg

Abstract

Quantitative correlations between the contents of the flavonolignans silychristin A and silybins A/B provide biosynthetic clues that support a pathway in which one mesomeric form of a taxifolin radical is undergoing an oxidative coupling with a coniferyl alcohol radical. The flavonolignan content and patterns reported in the literature for 53 samples, representing populations of the *Silybum marianum* plant growing in different parts of the world, were subject to a meta-analysis. Linear regression analyses were carried out on these data sets, and a mathematical model was derived that predicts the content of silychristin A relative to the metabolomic pattern of its congeners. The validity of the model was verified by applying it to test samples. This approach could potentially become a tool to enhance the understanding of both the relative composition of the silymarin complex and the biosynthetic pathways that underlie its formation.

Key words: Flavonolignan, linear regression, prediction, quantitative biosynthesis modelling

1. Experimental

1.1. Flavonolignans content in S. marianum *fruits*

The contents of the flavonolignans silychristin A, silydianin, silybin A, silybin B, and isosilybin B in the fruits of *S. marianum* in wild plant populations growing in Egypt has been determined previously (AbouZid et al. 2016). One sample was excluded (W154) because it contained eriodictyol-based flavonolignans. This data was extended by adding data from literature that reported on the contents of these flavonolignans (Shokrpour et al. 2008; Cheilari et al. 2016; Martinelli et al. 2016). In the reports, the contents of the flavonolignans were studied using chromatographic and spectroscopic platforms. Sample names are shown in Table 1 in the same manner as they appeared in the referred studies and reflected collections from different parts of the world (Egypt, Germany Austria, Italy, and Iran). Previous research showed major differences between relative contents of flavonolignans in the different silymarin chemotypes. Nevertheless, because the here considered flavonolignans have shown similar correlation patterns in different chemotypes (Martinelli et al. 2017), the samples were evaluated in bulk without further subdivision between chemotypes. For multiple linear regression, the samples were first divided into two sets: one training set composed of 43 samples and one test set composed of 10 samples. The latter was used for analysis of the predictive performance. The training data set and the test data set were divided at random. For simple linear regression, all 53 samples were used. The contents of the flavonolignans were normalized to mg/g dry weight (DW) of plant material before the statistical analysis was carried out.

1.2. ^{1}H NMR analysis

The NMR spectrometer used was a Bruker model AVANCE III HD (Fällanden, Switzerland) equipped with a BBFO Smart Probe and Bruker 400 AEON Nitrogen-Free Magnet, and operating at a ^{1}H frequency of 400.13 MHz (O1). The following conditions were used for acquisition of the ^{1}H NMR spectra: 30 degrees pulse experiment, excitation pulse 10 µs corresponding to an angle of xyz; acquisition time of 4.1 s; relaxation delay 0.1 s; sweep width 15.1 ppm (8012 Hz); temperature 295.1 K. Sixteen scans were recorded. Free induction decays were Fourier transformed after applying a line broadening factor (LB) of 0.1 Hz and zero filling to 256k. Deuterated

methanol (CD_3OD) was used for sample preparation. ACDLABS 12.0 software was used for post-acquisition processing.

1.3. S. marianum *sample preparation for ^1H NMR analysis*

S. marianum plants were collected beside the Cairo-Alexandria dessert road, Egypt in March, 2014. The plant was botanically authenticated by Dr. Abdel Halim Mohamed, Flora and Phytotaxonomy Department, Agricultural Research Center, Cairo Egypt. Ripe fruits were manually separated from the heads, and kept at -20°C until use. The seed shell was separated from the kernel according to a previously reported method (AbouZid et al. 2016b). Briefly, the fruits (1 g) were soaked in distilled water (10 ml) at room temperature; after 24 h, the water was removed by decantation, and the fruits were subjected to rolling using a wooden roller. During this procedure, the seed shell was separated intact from the kernel. The seed shell was then air-dried, ground and extracted in methanol (HPLC grade, x3). The combined extracts were evaporated to dryness, dried in a desiccator, and subjected to spectroscopic analysis.

1.4. Statistical analysis

Simple and multiple simple linear regression analysis were carried out using XLSTAT2017. The results from the multiple linear regression were analyzed for variance using Fisher's F test in order to either accept or reject the null hypothesis; whether or not the explanatory variables bring significant information to the model.

References

AbouZid S, Chen SN, Pauli GF. 2016. Silymarin content in *Silybum marianum* populations growing in Egypt. Ind Crop Prod 83: 729-737.

AbouZid, S., Chen, S.N., McAlpine, J.B., Friesen, J.B., Pauli, G.F., 2016. *Silybum marianum* pericarp yields enhanced silymarin products. Fitoterapia 112, 136-143.

Cheilari A, Sturm S, Intelmann D, Seger C, Stuppner H. 2016. Head-to-head comparison of ultra-high-performance liquid chromatography with diode array detection versus quantitative nuclear magnetic resonance for the quantitative analysis of the silymarin complex in *Silybum marianum* fruit extracts. J Agric Food Chem 64: 1618-1626.

Martinelli T, Potenza E, Moschella A, Zaccheria F, Benedettelli S, Andrzejewska J, 2016. Phenotypic evaluation of a milk thistle germplasm collection: fruit morphology and chemical composition. Crop Sci 56: 3160-3172.

Shokrpour M, Mohammadi SA, Moghaddam M, Ziai SA, Javanshir A. 2008. Variation in flavonolignan concentration of milk thistle (*Silybum marianum*) fruits grown in Iran. J Herbs Spices Med Plants 13: 55-69.

Table S1

Correlation coefficients (R) of silybin A, silybin B, and silychristin A contents in *Silybum marianum* samples.

Silychristin A	Silybin B	Silybin A	
0.994	0.988	**1**	Silybin A
0.984	**1**		Silybin B
1			Silychristin A

Table S2

Analysis of the variance for the multiple linear regression model used to predict the silychristin A content in *Silybum marianum*.

Pr > F	F	Mean squares	Sum of squares	DF	Source
< **0.0001**	1687.814	211.868	423.736	2	Model
		0.126	5.021	40	Error
			428.757	42	Corrected Total

Computed against model Y=Mean(Y)

Type I Sum of Squares analysis (Silychristin A):

Pr > F	F	Mean squares	Sum of squares	DF	Source
< **0.0001**	3374.852	423.639	423.639	1	Silybin A
0.383	0.776	0.097	0.097	1	Silybin B

Type III Sum of Squares analysis (Silychristin A):

Pr > F	F	Mean squares	Sum of squares	DF	Source
< **0.0001**	67.604	8.486	8.486	1	Silybin A
0.383	0.776	0.097	0.097	1	Silybin B

Table S3

Experimental values of silychristin A content (mg/g dry weight plant material), values predicted by model equation for the test samples and the respective errors.

Reference	Sample	Predicted silychristin A content	Experimental silychristin A content	Errors Percent (predicted – experimental /experimental * 100)
AbouZid et al. 2016	3P64W	5.25	5.11	2.74
	3P64Br	5.21	5.19	0.39
	3P64B	5.20	5.08	2.36
	3W	4.17	3.86	8.03
	P154	4.36	4.13	5.57
Cheilari et al. 2016	PL01	4.02	3.82	5.24
Martinelli et al. 2016	G22	11.49	11.88	-3.28
	G23	13.35	14.08	-5.47
	G24	2.03	2.57	-21.01
	G25	1.97	2.41	-18.26

Table S4

Prediction of silybin A and B content (mg/g DW plant material) using simple linear regression analysis. Data are based on the sample set reported by Cheilari et al. 2016.

	Silychristin A	Predicted Silybin A	Predicted Silybin B	Predicted Silybin A+B	Experimental Silybin A+B (qNMR)	Error (Predicted - Experimental)
PL01	**3.72**	**3.07**	**3.90**	**6.98**	**8.61**	**-1.63**
PL02	**1.06**	**0.46**	**1.17**	**1.63**	**2.22**	**-0.58**
PL03	**1.14**	**0.54**	**1.26**	**1.80**	**2.41**	**-0.61**
PL04	**1.59**	**0.98**	**1.71**	**2.69**	**3.07**	**-0.38**
PL05	**1.44**	**0.83**	**1.56**	**2.40**	**3.11**	**-0.71**
PL06	**1.28**	**0.68**	**1.40**	**2.08**	**2.11**	**-0.04**
PL07	**1.51**	**0.90**	**1.64**	**2.54**	**3.01**	**-0.48**
PL08	**1.04**	**0.44**	**1.15**	**1.59**	**2.48**	**-0.88**
PL09	**1.48**	**0.87**	**1.60**	**2.48**	**2.82**	**-0.34**
PL10	**1.37**	**0.76**	**1.49**	**2.26**	**2.81**	**-0.55**

Figure captions

Figure S1. The standardized regression coefficients for the correlations between the contents of silychristin A and those of silybin A and B.

Figure S2. Comparison of the experimental and predicted values of Silychristin A content (mg/g DW plant material) in *Silybum marianum* fruits. The experimental values are taken from Table 4. The predicted Silychristin A values were obtained by entering the Table 4 data for Silybins A & B in the equation: Silychristin A content = 0.56169 + 0.909663 * Silybin A content + 0.092599 * Silybin B content.

Figure S3. ^{1}H NMR spectrum of silymarin (0-8 ppm) showing the signals of silydianin at 2.90 ppm, silychristin A at 5.55 ppm, silybin A at 7.10 ppm, silybin B at 7.08 ppm, isosilybin A at 7.15 ppm, and isosilybin at B 7.13 ppm.

Figure S4. Simple linear regression analysis between silychristin A content vs silybin A content (A), silychristin A content vs silybin B content (B), as well as silydianin content vs isosilybin B content (C). The data is expressed as mg/g DW plant material. The values are taken from Table 1.

Figure S1

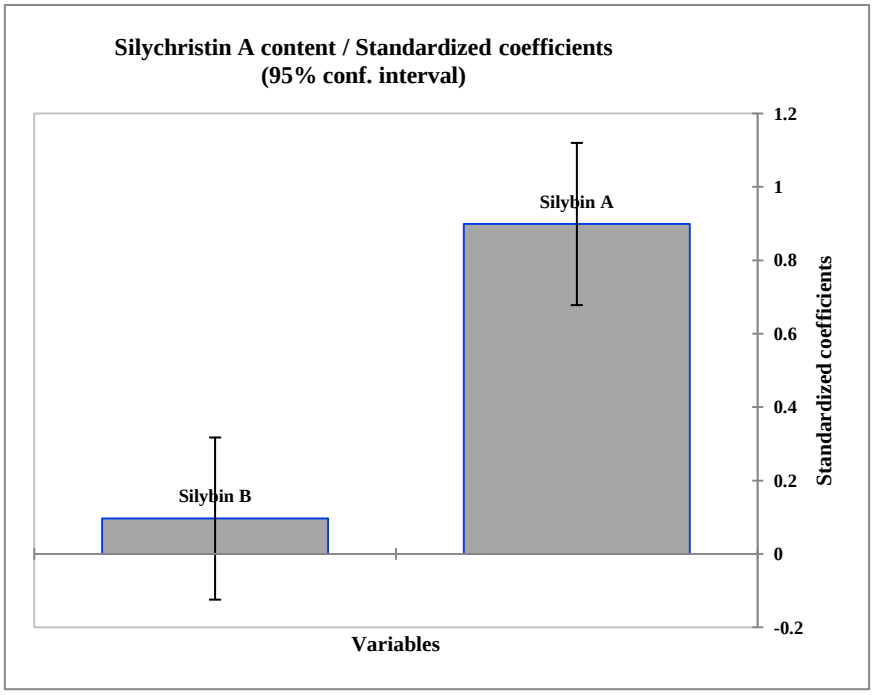

Silychristin A content / Standardized coefficients
(95% conf. interval)
Silybin A
Silybin B
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Standardized coefficients
Variables

Figure S2

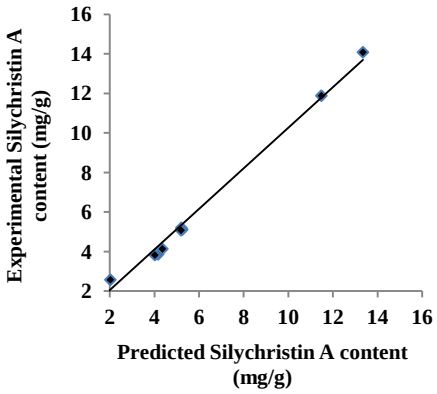

Experimental Silychristin A content (mg/g)
16
14
12
10
8
6
4
2
2
4
6
8
10
12
14
16
Predicted Silychristin A content (mg/g)

Figure S3.

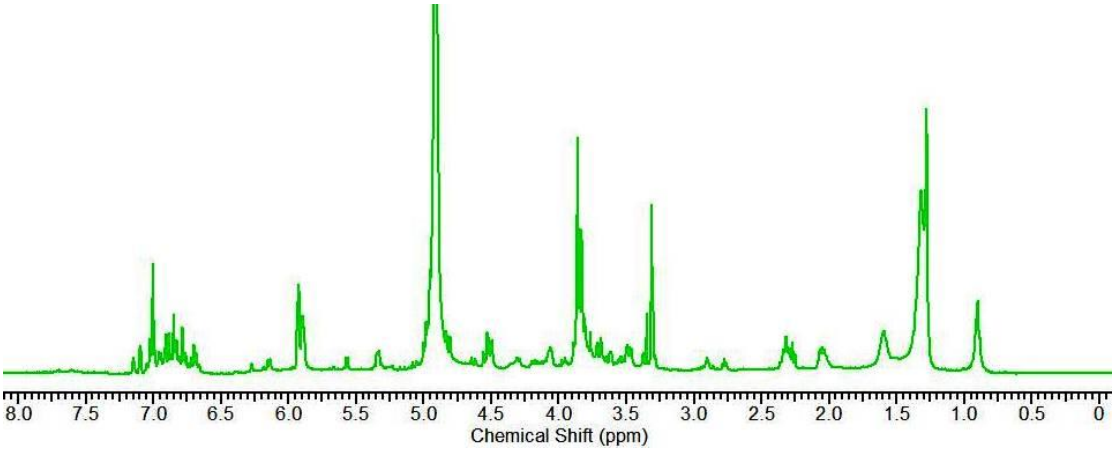
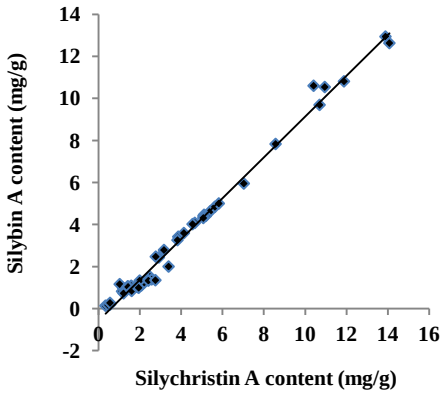
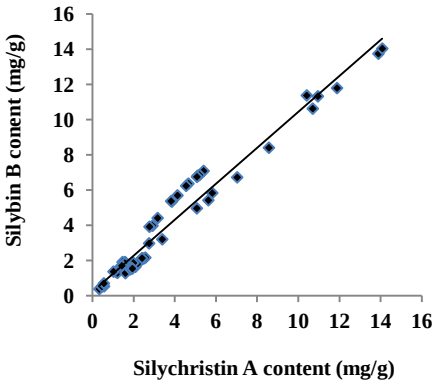


Figure S4.

A



B



C