# Supporting Information

#### Towards Normalization Selection of Raman Data in the Context of Protein Glycation: Application of Validity Indices to PCA Processed Spectra

Alsamad Fatima<sup>1</sup>, Gobinet Cyril<sup>1</sup>, Vuiblet Vincent<sup>1</sup>, Jaisson Stephane<sup>2</sup>, Piot Olivier<sup>1, 3\*</sup>

\*Corresponding Author: E-mail: <u>olivier.piot@univ-reims.fr</u>

#### List of figures:

**Figure S1.** Color scaled reconstructed maps of PCA-LDA classification accuracy on calibration set and test set obtained as function of the number of principal component (PC) by type of normalization SNV (standard normal variate), phenylalanine band (Phe), amide I band (AI), MSC (multiplicative signal correction) and EMSC (extended multiplicative signal correction). LDA models were trained on 3 collagen batches in each time (column to the left) and one batch served for test set (column to the right).

**Figure S2.** Color scaled reconstructed maps of XB validity index values obtained as function of the number of principal component (PC) by type of normalization SNV (standard normal variate), phenylalanine band (Phe), amide I band (AI), MSC (multiplicative signal correction) and EMSC (extended multiplicative signal correction). Column of figures to the left corresponds to XB <sub>calibration</sub> values obtained from collagen batches used to train LDA models. Column of figures to the right corresponds to XB <sub>test</sub> values obtained from collagen batch used as test in the LDA model. XB values are shown according to a decimal logarithm scale.

Figure S3. Pure CML Raman reference spectrum acquired at 785 nm.

### List of tables:

**Table S1.** DB <sub>calibration</sub>, XB <sub>calibration</sub>, XB <sub>test</sub> and DB <sub>test</sub> validity indices values and accuracy percentages of PCA-LDA models obtained for each of the first 7 PCs of PCA scores applied on SNV, Phe (phenylalanine), AI (amide I), MSC and EMSC normalized data sets of the 4 batches of untreated and acid-treated collagens. Model 1: training test on batches #2, #3 and #4 and test set on batch #1, model 2: training test on batches #1, #3 and #4 and test set on batch #2, model 3: training test on batches #1, #2 and #4 and test set on batch #3, model 4: training test on batches #1, #2 and #3 and test set on batch #4. Minima of DB and XB are underlined. Highest accuracy percentages are in bold.

#### Figure S1.



Number of F 63





Calibration set composed of batches 1#, 2# and #4



Calibration set composed of batches 1#, 2# and #3



#### Classification accuracy on the test set as function of principal component's number by type of normalization

Test set corresponding to batch #1



Test set corresponding to batch #2



Test set corresponding to batch #3



Test set corresponding to batch #4



Figure	<b>S2</b> .
--------	-------------





Figure S3.

## Table S1.

	SNV	Training set		Test set			
	PC number	DB <sub>calibration</sub>	XB <sub>calibration</sub>	Accuracy %	DB test	XB test	Accuracy %
	PC1	5,4	12,1	60	12,7	80,2	28
	PC2	<u>1,7</u>	<u>1,2</u>	67	60,8	1326,3	51
Ξ	PC3	3,8	6,2	50	2,4	2,8	33
ode	PC4	3,0	3,7	61	1,4	0,7	67
Ĕ	PC5	5,4	15,3	55	1,6	1,1	33
	PC6	6,7	15,6	50	1,5	0,8	48
	PC7	25,1	258,2	47	25,1	258,2	48
	PC1	10,1	40,9	50	2,0	1,4	55
	PC2	11,5	66,8	50	7,3	18,8	50
12	PC3	5,5	11,0	52	5,5	11,0	66
ode	PC4	<u>1,3</u>	0,7	77	1,8	1,1	58
Ĕ	PC5	4,6	7,7	55	2674,4	2839618,6	43
	PC6	2,8	3,6	64	4,4	9,2	59
	PC7	14,1	113,3	50	<u>1,3</u>	<u>0,8</u>	37
	PC1	6,4	14,2	52	7,8	28,9	60
	PC2	2,8	3,4	61	1,3	0,7	27
3	PC3	5,5	11,4	58	1,5	0,8	60
ode	PC4	2,2	3,7	68	177,2	15058,2	43
ž	PC5	<u>2,1</u>	2,0	64	1,0	0,4	73
	PC6	5,9	13,6	53	0,9	<u>0,3</u>	23
	PC7	17,2	159,6	45	208,0	17804,6	50
	PC1	6,8	17,4	50	4,5	9,1	54
	PC2	20,9	202,2	45	1,4	0,8	50
4	PC3	<u>1,6</u>	<u>1,1</u>	68	6,4	15,8	59
ode	PC4	1,7	1,2	68	16,6	102,2	45
Σ	PC5	8,6	50,7	45	1,1	0,5	80
	PC6	2,5	2,6	65	<u>0,6</u>	<u>0,1</u>	25
	PC7	31,8	517,6	45	18,3	120,6	50

	Phe	Training set		Test set			
	PC number	<b>DB</b> calibration	XB <sub>calibration</sub>	Accuracy %	DB test	XB test	Accuracy %
	PC1	21,9	193,1	47	1,2	0,6	46
	PC2	4,1	6,0	54	<u>2,3</u>	2,2	65
Ξ	PC3	2,2	1,8	65	5,9	15,4	52
ode	PC4	3,9	6,3	48	2,4	3,1	44
ž	PC5	2,4	2,4	65	2,7	3,2	51
	PC6	34,5	494,3	47	2,8	2,5	55
	PC7	5,0	10,2	54	1,6	0,9	38
	PC1	30,7	366,1	43	158,9	9844,9	47
	PC2	5,8	14,1	50	<u>1,1</u>	0,5	64
12	PC3	3,1	4,1	64	2,9	2,9	56
ode	PC4	4,8	8,0	55	11,2	50,6	61
ž	PC5	<u>1,5</u>	<u>1,1</u>	64	1,5	1,1	52
	PC6	2,5	2,4	68	2,2	2,3	52
	PC7	6,1	14,7	57	4,0	6,5	43
	PC1	112,6	4455,1	43	8,8	31,2	47
	PC2	3,1	3,5	52	7,3	18,4	50
13	PC3	558,5	119089,8	42	1,2	0,6	67
ode	PC4	4,1	8,1	53	1,4	1,0	47
ž	PC5	12,3	56,2	42	4,9	10,3	37
	PC6	2,2	2,4	59	0,8	0,2	83
	PC7	<u>1,7</u>	<u>1,5</u>	69	2,4	2,3	50
	PC1	5,5	10,6	50	2,7	2,6	54
	PC2	3,0	3,7	63	32,5	510,8	50
4	PC3	<u>1,7</u>	<u>1,1</u>	70	2,4	2,0	47
ode	PC4	5,2	10,8	60	2,6	2,6	63
ž	PC5	2,1	2,2	58	5,6	9,8	48
	PC6	11,6	58,2	55	2,0	1,9	73
	PC7	3,2	4,9	60	<u>1,5</u>	1,0	73

	AI	Training set		Test set			
	PC number	DBcalibration	XB <sub>calibration</sub>	Accuracy %	DB test	XB test	Accuracy %
	PC1	10,1	43,7	51	13,1	91,5	35
	PC2	<u>1,6</u>	1,0	69	1,5	0,7	69
Ξ	PC3	25,6	261,8	46	<u>0,7</u>	0,2	35
odel	PC4	7,4	25,1	44	1,1	0,5	39
Ĕ	PC5	2,8	2,9	56	2,5	2,3	71
	PC6	8,9	40,2	51	1,3	0,7	28
	PC7	8,9	29,1	48	3,8	4,4	43
	PC1	57,0	1409,2	45	2,1	1,8	50
	PC2	2,8	3,5	55	2,5	2,1	65
2	PC3	2,0	1,7	68	2,1	1,4	56
ode	PC4	16,9	112,3	45	4,8	8,8	65
Ĕ	PC5	<u>2,0</u>	1,6	68	<u>2,1</u>	1,8	68
	PC6	3,8	5,5	59	33,3	469,7	45
	PC7	15,8	148,4	45	2,5	2,4	64
	PC1	6,8	16,3	47	6,4	20,2	53
	PC2	<u>2,0</u>	1,7	65	75,5	2414,2	50
13	PC3	5,3	11,5	53	0,8	0,3	73
ode	PC4	68,0	1925,6	47	2,0	1,6	57
Σ	PC5	2,2	2,9	64	2,5	2,4	53
	PC6	3,5	5,3	61	1,2	0,6	67
	PC7	7,5	29,2	56	1,0	0,5	47
	PC1	15,8	100,3	50	3,7	6,7	54
	PC2	1,8	1,2	65	16,4	125,3	47
14	PC3	<u>1,8</u>	1,2	65	9,8	30,8	51
ode	PC4	4,5	8,1	45	1,7	1,1	36
ž	PC5	3,4	4,9	55	1,9	1,7	72
	PC6	13,8	85,4	55	7,2	20,5	40
	PC7	3,8	6,8	65	2,3	2,2	54

Model 1	C number PC1 PC2 PC3 PC4 PC5 PC6 PC7	DB <sub>calibration</sub> 5,4 <u>1,6</u> 3,9 3,0 6,1 7,0	<b>XB</b> <sub>calibration</sub> 12,1 1,1 6,4 3,7 19,5	Accuracy % 60 65 53 58	DB <sub>test</sub> 14,0 59,8 2,3	<b>XB</b> test 97,1 1285,8 2 6	Accuracy % 33 51
Model 1	PC1 PC2 PC3 PC4 PC5 PC6 PC7	5,4 <u>1,6</u> 3,9 3,0 6,1 7,0	12,1 <u>1,1</u> 6,4 3,7	60 <b>65</b> 53 58	14,0 59,8 2,3	97,1 1285,8 2.6	33 51
Model 1	PC2 PC3 PC4 PC5 PC6 PC7	<u>1,6</u> 3,9 3,0 6,1 7,0	<u>1.1</u> 6,4 3,7	<b>65</b> 53	59,8 2,3	1285,8 2.6	51
Model 1	PC3 PC4 PC5 PC6 PC7	3,9 3,0 6,1 7,0	6,4 3,7	53 58	2,3	26	
Mode	PC4 PC5 PC6 PC7	3,0 6,1 7.0	3,7 10 5	58		2,0	33
Ň	PC5 PC6 PC7	6,1 7.0	10.5	50	1,4	0,7	67
	PC6 PC7	70	19,5	51	1,5	1,0	33
	PC7	1,0	17,3	50	1,3	0,7	48
		25,7	261,4	45	2,4	2,5	48
	PC1	10,1	41,0	50	2,0	1,5	58
	PC2	11,4	65,5	52	7,1	17,8	50
2	PC3	5,6	11,4	52	4,1	7,0	66
ode	PC4	<u>1,3</u>	0,7	77	1,8	1,1	54
ž	PC5	4,7	8,0	50	58,6	1346,0	50
	PC6	2,7	3,4	68	4,1	8,1	59
	PC7	13,0	95,5	50	1,4	0,8	42
	PC1	6,3	13,9	52	8,0	30,4	60
	PC2	2,7	3,4	61	1,3	0,7	30
e	PC3	5,2	10,4	52	1,5	0,8	60
del	PC4	2,1	3,2	74	9,1	37,5	47
ž	PC5	2,3	2,4	59	1,0	0,4	73
	PC6	5,7	12,9	58	0,9	0,3	23
	PC7	27,2	383,1	55	8,4	29,2	50
	PC1	6,7	17,1	55	4,5	9,2	54
	PC2	23,6	258,6	45	1,4	0,8	58
4	PC3	1,6	<u>1,1</u>	70	6,4	15,6	59
del	PC4	1,7	1,1	65	16,2	96,4	45
Mo	PC5	9,2	58,9	50	1,1	0,5	80
	PC6	2,5	2,7	65	0,6	0,1	25
	PC7	39,9	807,7	45	12,1	53,3	50

	EMSC	Training set			Test set		
	PC number	<b>DB</b> <sub>calibration</sub>	<b>XB</b> <sub>calibration</sub>	Accuracy %	<b>DB</b> <sub>test</sub>	XB test	Accuracy %
	PC1	7,8	25,3	55	11,0	59,9	53
	PC2	<u>1,7</u>	1,2	61	<u>1,0</u>	0,4	61
-	PC3	2,6	2,9	56	5,2	9,8	52
leb	PC4	13,6	66,1	50	3,8	5,7	61
Mo	PC5	5,2	8,9	50	5,2	8,9	48
	PC6	7,8	37,0	53	4,9	10,1	44
	PC7	172,6	13100,8	47	1,6	1,2	50
	PC1	264,1	29383,1	41	1,7	1,1	50
	PC2	4,0	7,3	59	3,5	4,3	64
2	PC3	3,0	2,9	64	2,5	2,8	51
leb	PC4	<u>1,5</u>	<u>0,9</u>	68	<u>1,3</u>	0,7	65
Ň	PC5	9,1	38,8	50	8,6	28,1	47
	PC6	10,7	53,1	41	6,6	16,0	54
	PC7	76,8	2922,0	48	2,8	3,3	51
	PC1	6,2	13,2	52	8,2	31,4	53
	PC2	<u>2,2</u>	1,9	61	36,7	589,9	50
e	PC3	2,8	4,6	59	0,7	0,2	77
lab	PC4	3,1	3,9	50	0,9	0,3	77
Ĕ	PC5	17,9	151,7	53	2,3	2,0	57
	PC6	3,6	5,3	57	1,1	0,5	30
	PC7	5,5	21,7	57	1,8	1,4	53
	PC1	12,0	54,1	50	5,9	15,8	54
	PC2	3,6	5,0	60	1,3	0,7	72
4	PC3	<u>1,2</u>	0,6	75	4,0	6,7	38
labo	PC4	3,3	5,0	63	2,3	2,5	50
M	PC5	3,8	6,6	60	<u>1,0</u>	0,5	20
	PC6	2,7	3,1	63	1,9	1,4	29
	PC7	103,2	4992,3	45	4,9	9,7	52