Visualization for the "black-box" extraction process of Gegen Qinlian Decoction based on near-infrared spectroscopy

Sijun Wu [a,b], Guoming Zhou [a,b], Xiaoyang Zhang [a,b], Guangpu Fang [a], Jiahe Qian [a], Kun Zhang [c], Wenlong Li [a,b,d*]

[a] College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China;

[b] State key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China;

[c] Sichuan Puhua TCM Technology Co., Ltd, Sichuan, 611135, China;

[d] Haihe Laboratory of Modern Chinese Medicine, Tianjin, 301617, China

[*]Corresponding author (Wenlong Li). *E-mail address:* wshlwl@tjutcm.edu.cn

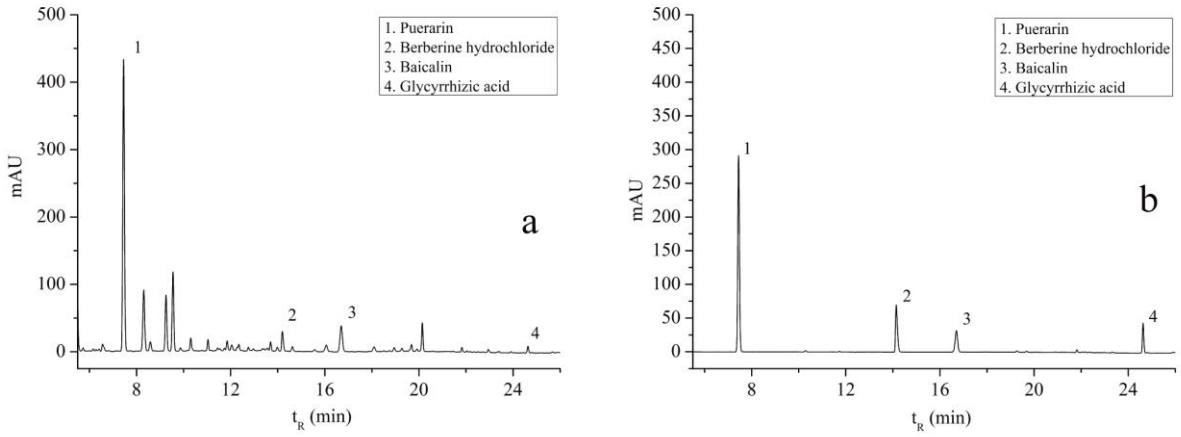

Fig. S1. Representative chromatograms of extracted solution and reference solution. (a) Extracted solution. (b) Reference solution.

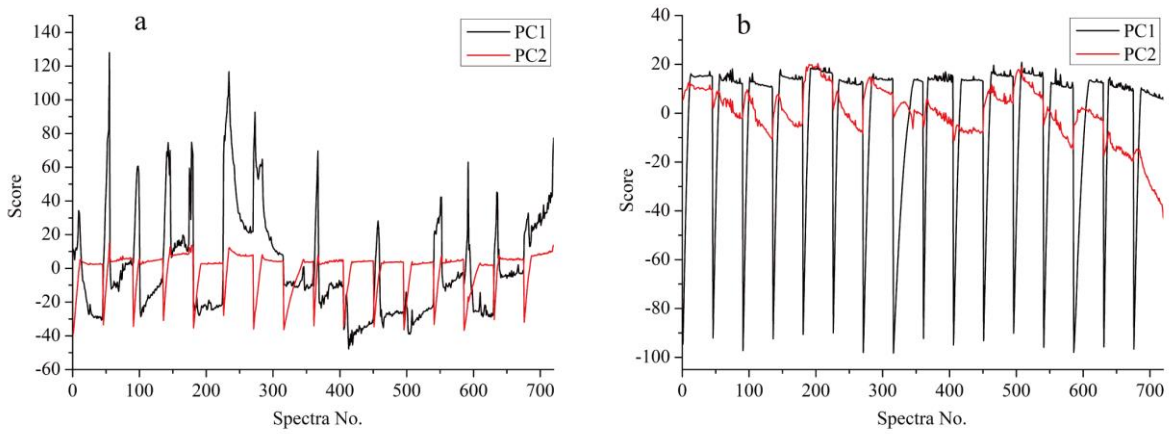

Fig. S2. Scores of the first two principal components of near-infrared spectral data. (a) Without preprocessing. (b) Preprocessing with SNV.

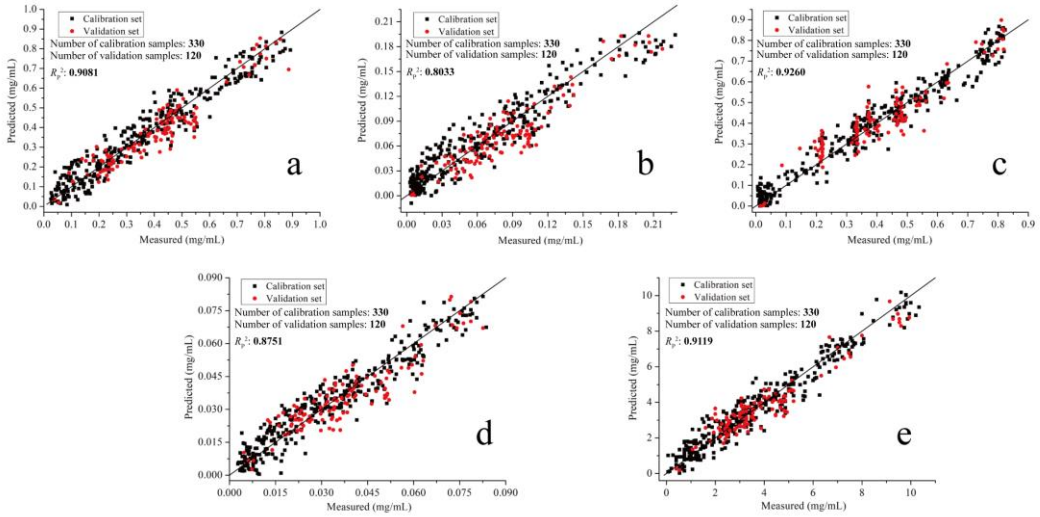

Fig. S3. Correlation diagrams of measured values vs. predictions of CARS-PLS regression models for the components and soluble solids. (a) Puerarin. (b) Berberine hydrochloride. (c) Baicalin. (d) Glycyrrhizic acid. (e) Soluble solids.

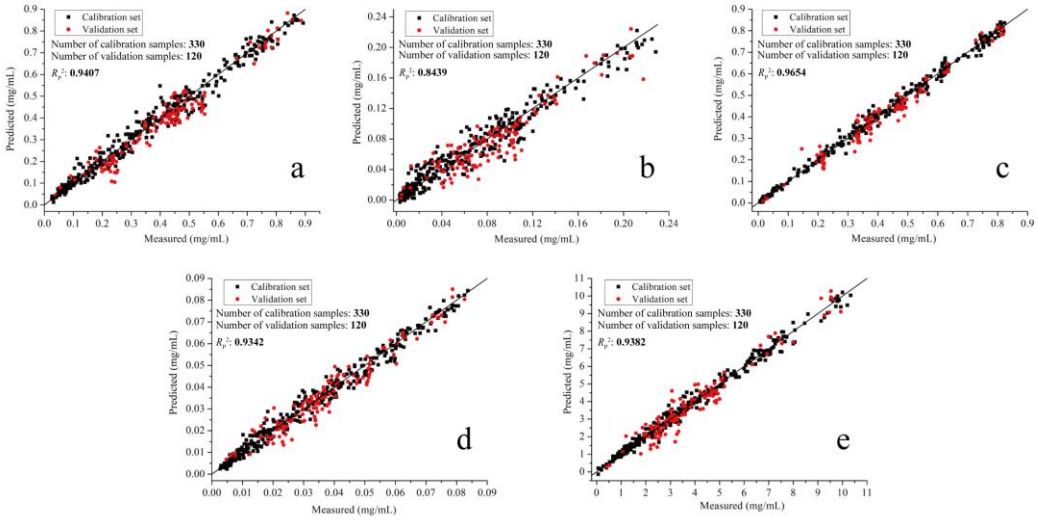

Fig. S4. Correlation diagrams of measured values vs. predictions of GP regression models for the components and soluble solids. (a) Puerarin. (b) Berberine hydrochloride. (c) Baicalin. (d) Glycyrrhizic acid. (e) Soluble solids.

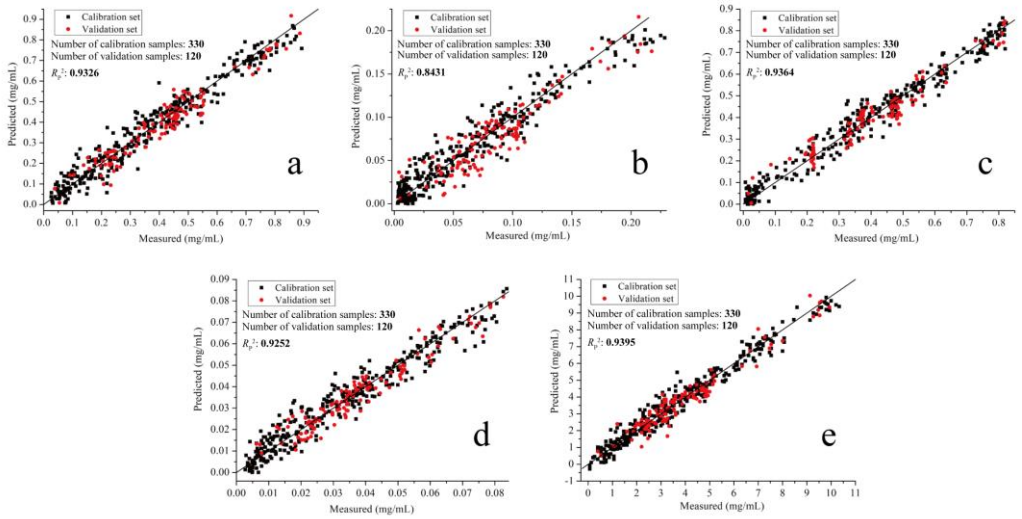

Fig. S5. Correlation diagrams of measured values vs. predictions of ELM models. (a) Puerarin. (b) Berberine hydrochloride. (c) Baicalin. (d) Glycyrrhizic acid. (e) Soluble solids.

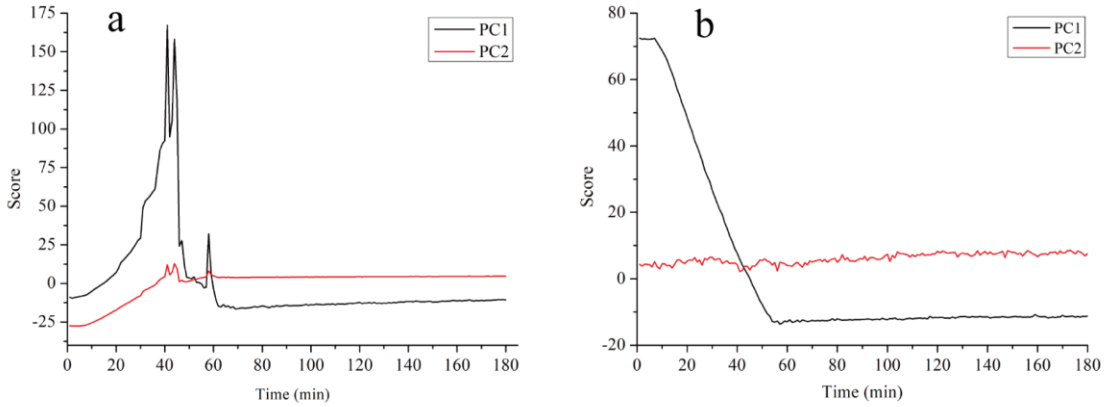

Fig. S6. Scores of the first two principal components of NIR spectral data. (a) Without preprocessing. (b) Preprocessing with 1st der.

Table S1 Optimization of spectral preprocessing methods for four types of components and soluble solids

Preprocessing methods	Puerarin		Berberine hydrochloride		Baicalin		Glycyrrhizic acid		Soluble solids	
	PCs	RMSECV (mg/mL)	PCs	RMSECV (mg/mL)	PCs	RMSECV (mg/mL)	PCs	RMSECV (mg/mL)	PCs	RMSECV (mg/mL)
Raw spectra	9	0.0704	11	0.0166	9	0.0657	5	0.0095	11	0.5451
MSC	8	0.0639	9	0.0197	10	0.0565	8	0.0063	9	0.6183
SNV	9	0.0555	9	0.0185	9	0.0530	9	0.0063	9	0.4314
1st derivative	5	0.0920	5	0.0286	6	0.0996	5	0.0100	5	1.0240
SG smoothing	9	0.0737	11	0.0217	7	0.0779	9	0.0076	9	0.7927
SNV-SG smoothing	8	0.0775	8	0.0241	9	0.0768	8	0.0087	9	0.8275
1st derivative-SG smoothing	5	0.1040	3	0.0316	4	0.1158	3	0.1160	5	1.2037
SNV-1st derivative	6	0.0871	5	0.0272	5	0.0937	5	0.0095	5	0.9667

Table S2 Number of selected variables using the three kinds of methods

Ingredients	Methods	Number of variables
Puerarin	None	1007
	UVE	94
	SPA	34
	CARS	87
Berberine hydrochloride	None	1007
	UVE	78
	SPA	19
	CARS	60
Baicalin	None	974
	UVE	82
	SPA	33
	CARS	105
Glycyrrhizic acid	None	1007
	UVE	91
	SPA	20
	CARS	98
Soluble solids	None	1007
	UVE	82
	SPA	27
	CARS	127