OXFORD

Journal of AOAC INTERNATIONAL, 104(3), 2021, 571–578

doi: 10.1093/jaoacint/qsaa176
Advance Access Publication Date: 25 January 2021
Article

DRUG FORMULATIONS

Multivariable Manipulation of Spectrophotometric Data with Genetic Algorithm Selection for Novel Quantitative Resolution of Five Antiviral Drugs in Their Pharmaceutical Products

Hebatallah M. Essam[1,*] Amani M. Morsi[2] and Mohammed B. El-Zeiny[3]

[1]Analytical Chemistry Department, Faculty of Pharmacy, Cairo University, Cairo, Egypt, [2]National Organization for Drug Control and Research, Giza, Egypt, [3]Analytical Chemistry Department, Faculty of Pharmacy, Modern University for Technology and Information (MTI), Cairo, Egypt

*Corresponding author's e-mail: heba_essam80@hotmail.com

Abstract

Background: In many real-world situations there are many components in a mixture that produce an enormous amount of information.
Objective: The main task is to build up balanced models that convert these data into meaningful information to deal with. Hence, different chemometric models were applied for the analysis of data obtained from a mixture containing sofosbuvir, ledipasvir, velpatasvir, daclatasvir, and valacyclovir that were recently used internationally for their antiviral activity.
Methods: Partial Least Squares, Spectral Residual Augmented Classical Least Squares, and Concentration Residual Augmented Classical Least Squares designs were applied with and without variable selection procedure [Genetic Algorithm (GA)]. The methods were used for the quantitative analysis of the drugs in laboratory prepared mixtures and real market sample through handling the UV spectral data.
Results: Robust models were obtained by applying GA. The proposed methods were found to be rapid, simple, and required no preliminary separation steps.
Conclusion: These models can be used on a routine basis in quality control laboratories or factories giving competitor results to those obtained by the reported methods.
Highlights: The proposed models offer a powerful analytical alternative for laboratories that consider economic strategies in their requirements.

Viral infection is a burdensome public health issue worldwide. It may hit different body parts causing a lot of diseases with variable severity and life-threatening impact. Hepatitis C is a liver disease caused by hepatitis C virus (HCV). Hepatitis C may be acute or chronic, ranging in severity from mild illness lasting a few weeks to lifelong illness that can ultimately result in cirrhosis, hepatocellular carcinoma, and the need for liver transplantation. Sustained interest in all science aspects has been noticed to study, diagnose, treat, and fight the increased morbidity from HCV infection (1–3). Recently, newly introduced anti-HCV combinations that shoot into the goal and eradicate HCV-RNA with minimal side effects in the early stages of disease replace traditional interferons regimens for treatment.

Sofosbuvir (SOFO) is a nucleotide prodrug that inhibits HCV NS5B (non-structural protein 5B) polymerase approved for the treatment of HCV in combination with other antiviral molecules

Received: 2 November 2020; **Revised:** 5 December 2020; **Accepted:** 17 December 2020

(4). Velpatasvir (VELP) is a new pangenotypic HCV NS5A inhibitor that inhibits genotype 1–6 HCV RNA replicons. Both ledipasvir (LEDI) and daclatasvir (DACL) are first class, directly acting antiviral agents that bind to and inhibit the function of the HCV protein NS5A, which is involved in both viral RNA replication and virus particle assembly (5).

Chicken pox or shingles and herpes are viral diseases that infect the majority of humans, especially those with weakened immune systems. Valacyclovir (VALA) is a prodrug of the antiviral drug acyclovir that exhibits activity against herpes simplex virus types, 1(HSV–1) and 2(HSV–2) and vericellazoster virus (6). The mechanism of action of acyclovir involves the highly selective inhibition of virus DNA replication.

Current in-vitro studies revealed that the combinations of SOFO with VELP, DACL, or LEDI have additive antiviral interactions and lack of cross-resistance. Administration of these co-formulated drugs in a broad range of patients with HCV maintains sustained virologic response (7–9).

The therapeutic importance of these compounds justifies research to establish analytical methods for their determination in bulk drugs and pharmaceutical formulations. In literature, SOFO has been quantified in a presence of its degradation products (10), or simultaneously in combination with VELP (11,12), DACL (13), LEDI (14–15), or ribavirin (16) using variable analytical techniques. On the other hand, VALA was determined in single studies or in the presence of its degradation products and metabolites using spectrophotometric methods (17) and chromatographic methods (18) in different matrices.

Chemometrics as an analytical tool is not a single tool but a range of methods including basic statistics, signal processing, factorial design, calibration, curve fitting, factor analysis, detection, pattern recognition, and neural network with an aim to achieve maximum accuracy, precision, and robustness. This tool has the capacity for analyzing and modeling a wide variety of data types for an even more diverse set of applications. Hence, when the higher-order data is produced, the implementation of chemometrics becomes more powerful and more urgent (19).

Also, with the advance of computational techniques, chemometrics has become a leading tool for faster analysis of results/data and shorter product development time. It is generally applied for one or more of three primary purposes to explore patterns of association in data, track properties of materials on a continuous basis, and build and use multivariate classification models. There are various algorithms and analogous ways for processing and evaluating the data and they can be implemented to various fields for the analysis of data of a particular manufacturing process, quality control test, or an instrumental output data to provide a rapid quantitative analysis of pharmaceutical products as characterized by the simple, nondestructive, and highly sensitive nature of the method (20).

Fixed dose combinations of SOFO/VELP, SOFO/DACL, and SOFO/LEDI were consecutively approved by the US Food and Drug Administration (FDA) for the treatment of infected patients with HCV (2014–2016), hence the development of single, simple, accurate, and economic method for the analysis of these mixtures in addition to VALA in their pharmaceutical matrices is a demanded mission. Consequently, this work aims to demonstrate the resolving power of multivariate models that can manipulate complicated spectrophotometric information of these compounds giving precise and accurate estimation of their concentrations without preliminary separation steps or sophisticated instruments. These multivariate calibration methods are concentration residual augmented classical least squares (CRACLS) (21,22), spectral residual augmented classical least squares (SRACLS) (23, 24), and partial least squares (PLS) (25–29).

Moreover, GA is used as a variable selection method to enhance the model's performance. The proposed methods are definitely a smart alternative for the simultaneous analysis of the selected antiviral compounds in quality control laboratories.

Experimental

Materials and methods

(a) *SOFO, LEDI, VELP, DACL, and VALA.*—Kindly supplied from the National Organization for Drug Control and Research (NODCAR, Cairo, Egypt). The purity of the standards was verified to be 99.51 ± 1.19, 99.36 ± 0.89, 100.06 ± 0.74, 99.87 ± 0.95, and 100.32 ± 0.99%, respectively, according to the reported methods (4, 5, 14, 30).

(b) *Harvoni® film-coated tablets.*—Manufactured by Gilead Sciences Inc., Foster City, California, USA. It was labeled to contain 90 mg LEDI and 400 mg SOFO per tablet.

(c) *Epclusa® commercial tablets.*—Manufactured by Gilead Sciences Inc., Foster City, CA, USA. Each tablet is claimed to contain 400 mg SOFO and 100 mg VELP.

(d) *Daklinza® 60mg film-coated tablets.*—(Bristol-Myers Squibb Pharma EEIG, UK) were labeled to contain 65.92 mg daclatasvir dihydrochloride, equivalent to 60 mg DACL per tablet.

(e) *Valysernex® film-coated caplets.*—manufactured by EVA PHARMA, El-Sadat, Zawya Abou Muslim, Giza, Egypt, labeled to contain 500 mg VALA per caplet.

(f) *Methanol.*—HPLC grade was purchased from E. Merck, Darmstadt, Germany.

Instrument

Spectrophotometric measurements were performed on Shimadzu UV-Visible spectrophotometer duel beam (Kyoto/Japan), model UV-1650 PC with two matched 1-cm quartz cells. UV-Probe personal spectroscopy software version 2.21 (Shimadzu) was used for operating the instrument. The spectral band width was 1 nm and scanning speed was 2800 nm/min with 0.1 nm interval.

Software

CRACLS, SRACLS, and PLS were performed using our written codes in Matlab 8.2.0.701 (R2013b). GA was performed using PLS toolbox version 2.1, while student's t-test and F-test were performed using Microsoft® Excel.

Procedures

Standard solutions

Standard working solutions for SOFO, VELP, DACL, LEDI, and VALA (100 μg/mL) were prepared by separately dissolving the appropriate amount of each compound in methanol. Then completing to mark in 100 mL volumetric flasks.

Spectral characteristics of anti-viral drugs

Standard solution of each drug of 10 μg/mL was prepared to record the zero-order (D_0) absorption spectra with respect to a blank of methanol over a range of 200–400 nm at 1 nm interval. Due to severe overlapped spectra of the studied drugs, the proposed chemometric methods have been used to analyze this mixture. Wavelengths less than 230 nm were rejected due to the noisy content.

Chemometric design for the proposed models

Prepare 25 mixtures with different ratios of the 5 components according to 5 levels and 5 factors of the experimental design in which the concentration of each component has to be orthogonal to other components in the mixtures. In order to apply the design, different volumes were transferred from the standard working solution of each component into series of 10 mL volumetric flasks. Then, volumes were completed with methanol. The final concentrations were ranging from 6 to 14 μg/mL with 10 μg/mL as a central level of the design for all drugs. The two levels above and below the central level for each drug were chosen to ensure the linearity of the obtained spectral data concerning the ratio of each compound in its pharmaceutical formulation. The scanned absorption spectra of the prepared mixtures were stored in the computer and the data points of each spectrum were transferred to Matlab® for subsequent data selection and processing. Seventeen mixtures were used for building the calibration model (training set), while eight mixtures were chosen as an external validation set to test the prediction of the built multivariate calibration models.

Application of the proposed models for the analysis of Epclusa, Daklinza, Harvoni, and Valysernex commercial preparations

For determination of the studied drugs in their pharmaceutical formulations, 20 tablets of Epclusa (SOFO/VELP) were weighed and then powdered. Then liquid extraction method was applied to remove the interference of industrial additives by placing accurate amount equivalent to 1 tablet in a 250 mL beaker. Fifty millileters of methanol was added, sonicated for 10 min and filtered through 0.5 μm Whatman filter paper into a 100 mL volumetric flask. The residue was washed with methanol three times and washing aliquots were quantitatively collected athen nd completed to the volume with methanol.

Apply the above-mentioned procedure for the preparation of Daklinza (DACL), Harvoni (SOFO/LEDI), and Valysernex (VALA).

Further dilutions were done for each formulation to reach linearity ranges using methanol (10 μg/mL VALA, 12 μg/mL DACL, 9 μg/mL LEDI, 10 μg/mL VELP, and 12 μg/mL SOFO). The procedure described under each method was applied to analyze the drugs in the prepared dosage form solution.

The absorption spectra for the prepared dosage forms dilutions were recorded in the range of 230–400 nm, then processed by the constructed calibration models to estimate the concentration of cited components.

Table 1. The concentrations of sofosbuvir, ledipasvir, velpatasvir, daclatasvir, and valacyclovir in μg/mL in the used experimental design

Mix. No.	SOFO	LEDI	VELP	DACL	VALA
1[a]	10	10	10	10	10
2	10	6	6	14	8
3	6	6	14	8	14
4	6	14	8	14	10
5	14	8	14	10	8
6	8	14	10	8	8
7	14	10	8	8	12
8	10	8	8	12	14
9	8	8	12	14	12
10	8	12	14	12	10
11	12	14	12	10	14
12	14	12	10	14	14
13	12	8	14	14	6
14	10	14	14	6	12
15	14	14	6	12	6
16	14	6	12	6	10
17	6	12	6	10	12
18	12	6	10	12	12
19	6	10	12	12	8
20	10	12	12	8	6
21	12	12	8	6	8
22	12	8	6	8	10
23	8	6	8	10	6
24	6	8	10	6	6
25	8	10	6	6	14

[a] The shaded rows represent the calibration set.

Results and Discussion

The global interest with eco-friendly life coincides with the scientists' concerns of developing green, substantial analytical methods for accurate and precise estimation of life-saving products. The use of spectrophotometric analytical technique in quality control laboratories is a fundamental approach.

Table 2. Parameters of the Genetic Algorithm for the built models

Parameter	Value
Population size	80
Maximum generations	100
Mutation rate	0.005
The number of variables in a window (window width)	4
Percent of population the same at convergence	30
% Wavelengths used at initiation	20
Crossover type	Single
Maximum number of latent variables	6
Cross validation	Random
Number of subsets to divide Data into for cross validation	5
Number of iterations for cross validation at each generation	2

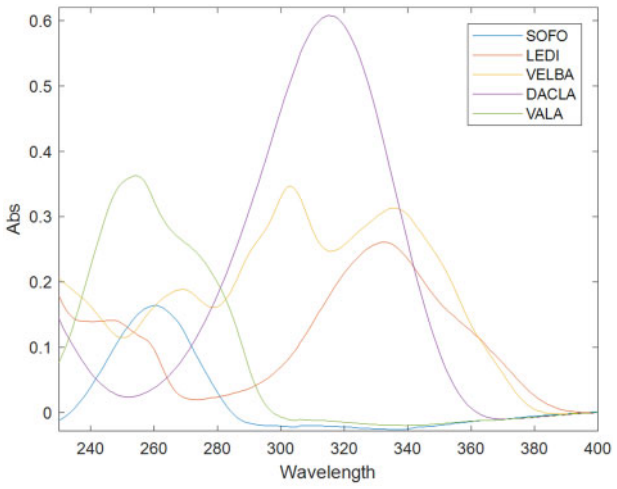

Figure 1. Zero-order spectra of 10 μg/mL from the investigated drugs.

However, the point of ultimate importance in analysis is method selectivity. Lower selectivity is considered a main disadvantage of spectrophotometric methods. To enhance the selectivity, numerical and graphical techniques were implemented for treatment of spectral data which makes the use of spectrophotometric techniques paired with mathematical algorithms and wavelet transform is of the utmost necessity. Thus, this work aims to bring a green, new, easy, and selective chemometric methodology and yet very economical for the determination of analytes in samples (31).

In this study, the spectra of the five active compounds to be quantitatively determined show severe overlap that seriously hindered the resolution of this complex mixture. Therefore, variable multivariate calibration methods were performed to predict the concentration of the five compounds in both calibration and validation set samples, as well as in their market samples; *see* Table 1.

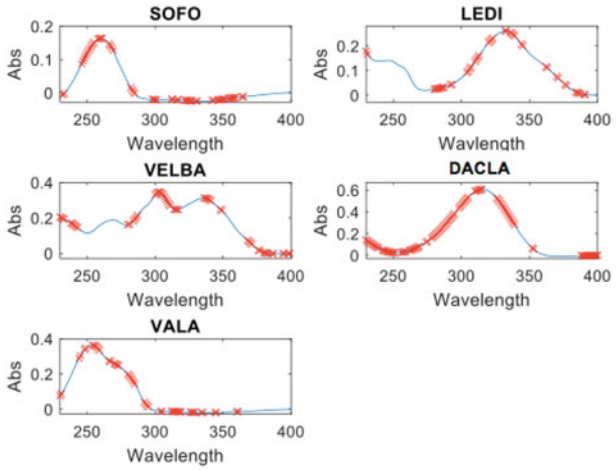

Figure 2. The chosen wavelengths by genetic algorithm (x) for 10 μg/mL of the investigated drugs.

GA as variable selection method

Constructing an analytical model with the least noise and cost is critical in a competitive environment. However, shortening the duration of any analytical activity usually requires the engagement or adoption of additional techniques. GA could be used to establish the fitness of measured data by evaluating the most informative variables and removes irrelevant ones to generate the most optimal outcome resulting in a less complex model built with less variables. It can codify information of each variable as a gene along a finite length string and use objective functions as mutation, crossover, and selection procedures to determine the promising regions. The main issue in applying GA is the optimization of its parameters. One hundred independent short runs were done to reduce the probability of occurrence of over-fitting (32). The optimal parameter settings for GA model are shown in Table 2. After application of GA model, the absorbance matrix was minimized to about 33, 26, 33, 56, and 26% of its original size for SOFO, VALP, DACL, LEDI, and VALA, respectively, as revealed in Figure 2.

Table 3. Assay validation sheet of sofosbuvir, ledipasvir, velpatasvir, daclatasvir, and valacyclovir by the proposed models

Model	SOFO	LEDI	VELP	DACL	VALA	Model	SOFO	LEDI	VELP	DACL	VALA
GA-PLS-1						**PLS-1**					
RMSEC[a]	0.05305	0.34355	0.0287	0.05332	0.0481	**RMSEC**[a]	0.0594	0.3170	0.0232	0.06550	0.0635
RMSEP[b]	0.09585	0.41247	0.0543	0.07330	0.0681	**RMSEP**[b]	0.0961	0.5064	0.0782	0.12523	0.1604
RRMSEP[c]	1.03630	4.12470	0.6040	0.71519	0.0633	**RRMSEP**[c]	1.0398	5.0649	0.8698	1.22182	1.4923
BCRMSE[d]	−0.0020	0.19126	−0.0004	0.00160	−0.0075	**BCRMSE**[d]	0.0060	0.24920.9941	−0.0053	−0.00299	0.0242
r[e]	0.9998	0.9930	0.9999	0.9999	0.9999	**r**[e]	0.9997	0.4164	0.9999	0.9997	0.9998
intercept[e]	0.0255	0.4273	−0.0066	0.0222	0.0006	**intercept**[e]	0.0418	0.9605	0.0111	0.0596	−0.009
slope[e]	0.9977	0.9593	1.0006	0.9979	0.9999	**slope**[e]	0.9962		0.9999	0.9944	1.0008
GA-CRACLS						**CRACLS**					
RMSEC[a]	0.09954	0.37111	0.0524	0.0540	0.0658	**RMSEC**[a]	0.0746	0.3454	0.0324	0.06590	0.0896
RMSEP[b]	0.09907	0.32215	0.0498	0.07043	0.0756	**RMSEP**[b]	0.0888	0.3766	0.0571	0.09758	0.0966
RRMSEP[c]	1.07102	3.2214	0.5537	0.68720	0.7039	**RRMSEP**[c]	0.9607	3.7660	0.6352	0.95205	0.8991
BCRMSE[d]	0.00995	0.10839	0.0014	0.00327	0.0002	**BCRMSE**[d]	0.0084	0.1615	−0.0092	0.00094	0.0104
r[e]	0.9993	0.9914	0.9999	0.9999	0.9998	**r**[e]	0.9996	0.9926	0.9999	0.9997	0.9995
intercept[e]	0.0068	0.2483	−0.035	0.017	0.0096	**intercept**[e]	0.0067	0.2991	0.0238	0.0532	0.0145
slope[e]	0.9994	0.977	1.0031	0.9984	0.9991	**slope**[e]	0.9994	0.9723	0.9979	0.995	0.9986
GA-SRACLS						**SRACLS**					
RMSEC[a]	0.08383	0.35011	0.0346	0.05469	0.0896	**RMSEC**[a]	0.0650	0.3225	0.0243	0.07281	0.0693
RMSEP[b]	0.08259	0.36754	0.0435	0.07202	0.0685	**RMSEP**[b]	0.0990	0.4448	0.0746	0.11341	0.1923
RRMSEP[c]	0.89289	3.67539	0.4844	0.70265	0.6374	**RRMSEP**[c]	1.0706	4.4489	0.8295	1.10650	1.7890
BCRMSE[d]	−0.0036	0.15376	−0.0015	0.00200	−0.0039	**BCRMSE**[d]	0.0068	0.2078	−0.0056	−0.00091	0.0362
r[e]	0.9995	0.9924	0.9999	0.9998	0.9995	**r**[e]	0.9997	0.9936	0.9999	0.9996	0.9997
intercept[e]	0.0598	0.2681	−0.0139	0.0206	−0.0276	**intercept**[e]	0.0416	0.293	0.0143	0.0609	−0.027
slope[e]	0.9946	0.9752	1.0012	0.9981	1.0026	**slope**[e]	0.9963	0.9729	0.9987	0.9943	1.002

[a] Root Mean Square Error of Calibration.
[b] Root Mean Square Error of Prediction.
[c] Relative Root Mean Square Error of Prediction.
[d] Bias Corrected Root Mean Square Error of Prediction.
[e] Data of the straight line plotted between predicted concentrations versus actual concentrations of the calibration set.

Table 4. Statistical comparison for the results obtained by the proposed methods and the reported methods for investigated drugs in bulk powder

Models	Mean	SD	N	Variance	Student's *t* test [a] (2.306)	*F* value [a] (9.014)
			SOFO			
PLS-1	100.36	0.89	4	0.79	1.25	1.08
GA-PLS-1	100.09	1.31	4	1.72	0.72	2.02
CRACLS	99.97	1.19	4	1.42	0.60	1.67
GA-CRACLS	99.97	1.03	4	1.06	0.64	1.25
SRACLS	100.28	0.82	4	0.67	1.17	1.27
GA-SRACLS	100.20	1.12	4	1.25	0.92	1.47
Reported method (4)	99.51	0.92	6	0.85		
			LEDI			
PLS-1	100.27	1.12	4	1.25	1.05	1.14
GA-PLS-1	99.51	1.16	4	1.35	0.17	1.05
CRACLS	101.05	1.55	4	2.40	1.69	1.69
GA-CRACLS	100.45	1.45	4	2.10	1.12	1.48
SRACLS	100.60	1.64	4	2.69	1.20	1.89
GA-SRACLS	99.63	1.01	4	1.02	0.32	1.39
Reported method (14)	99.36	1.19	6	1.42		
			VELP			
PLS-1	100.27	0.66	4	0.44	0.39	1.25
GA-PLS-1	100.49	0.86	4	0.74	0.73	1.35
CRACLS	100.23	0.62	4	0.38	0.33	1.45
GA-CRACLS	100.58	0.62	4	0.38	1.00	1.45
SRACLS	100.30	0.50	4	0.25	0.48	2.20
GA-SRACLS	100.49	0.81	4	0.66	0.74	1.20
Reported method (4)	100.06	0.74	6	0.55		
			DACL			
PLS-1	99.82	0.63	4	0.40	0.08	2.25
GA-PLS-1	99.59	1.04	4	1.08	0.38	1.20
CRACLS	99.89	0.62	4	0.38	0.03	2.37
GA-CRACLS	99.75	0.79	4	0.62	0.18	1.45
SRACLS	99.84	0.62	4	0.38	0.05	2.37
GA-SRACLS	99.66	0.93	4	0.86	0.62	1.05
Reported method (5)	99.87	0.95	6	0.90		
			VALA			
PLS-1	99.65	0.56	4	0.31	1.05	3.16
GA-PLS-1	100.02	1.44	4	2.07	0.34	2.11
CRACLS	99.75	0.68	4	0.46	0.86	2.13
GA-CRACLS	99.94	0.90	4	0.81	0.53	1.21
SRACLS	99.65	0.63	4	0.40	1.02	2.45
GA-SRACLS	99.79	1.13	4	1.28	0.68	1.31
Reported method (28)	100.32	0.99	6	0.98		

[a] The values in the parenthesis are the corresponding theoretical values of *t* and *F* at ($\alpha = 0.05$).

PLS model

This multivariable calibration technique is persistently used in quantitative spectral analysis to obtain very selective information from unselective data. It involves the simultaneous decomposition of both concentration and spectral data matrices into set of pairs called loadings and scores. Each pair is called latent variable (LV). In order to determine the correct number of LVs to be used for modeling the auto-scaled data, cross-validation with random subsets selection was performed to calculate the cumulative PRESS (prediction residual error sum of squares). Haaland and Thomas's criterion (23) was used to calculate the optimum number of LVs to be used in the models. Then, the root mean square error of cross validation (RMSECV) of each LV is statistically compared to that of the LV with the lowest RMSECV. The

Table 5. Statistical comparison for the results obtained by the application of the proposed methods and the reported methods for the analysis of the investigated drugs in their pharmaceutical formulations

Dosage form	Models	Mean	SD	N	Variance	Student's *t* test [a] (2.776)	*F* value [a] (19.000)
Harvoni tablet claimed to contain 90 mg LEDI/400 mg SOFO per tablet	SOFO						
	PLS-1	90.20	1.31	3	1.72	0.792	1.69
	GA-PLS-1	91.13	1.22	3	1.49	0.185	1.46
	CRACLS	89.70	1.05	3	1.10	1.500	1.08
	GA-CRACLS	89.76	0.98	3	0.96	1.537	1.06
	SRACLS	91.19	1.18	3	1.39	0.256	1.36
	GA-SRACLS	91.26	0.99	3	0.98	0.366	1.28
	Reported method (14)	90.96	1.01	3	1.02		
	LEDI						
	PLS-1	92.14	0.88	3	0.77	2.256	1.59
	GA-PLS-1	92.56	1.10	3	1.21	1.571	1.02
	CRACLS	96.26	1.20	3	1.44	2.389	1.17
	GA-CRACLS	94.91	1.24	3	1.54	0.948	1.25
	SRACLS	94.25	0.89	3	0.79	0.317	1.56
	GA-SRACLS	94.23	0.82	3	0.67	0.300	1.84
	Reported method (14)	93.99	1.11	3	1.23		
Epclusa tablet claimed to contain 100 mg VELP/400 mg SOFO per tablet	SOFO						
	PLS-1	89.30	0.95	3	0.90	2.316	1.06
	GA-PLS-1	89.49	0.89	3	0.79	2.121	1.08
	CRACLS	88.17	1.03	3	1.06	2.362	1.25
	GA-CRACLS	87.87	1.11	3	1.23	2.639	1.45
	SRACLS	89.66	0.87	3	0.76	0.548	1.12
	GA-SRACLS	90.30	0.83	3	0.69	0.338	1.23
	Reported method (4)	90.06	0.92	3	0.85		
	VELP						
	PLS-1	90.89	0.96	3	0.92	2.527	1.30
	GA-PLS-1	92.53	0.83	3	0.69	0.333	1.03
	CRACLS	94.54	0.88	3	0.77	2.507	1.08
	GA-CRACLS	94.62	0.81	3	0.66	2.735	1.08
	SRACLS	91.55	1.00	3	1.00	1.592	1.41
	GA-SRACLS	92.73	0.95	3	0.90	0.041	1.27
	Reported method (4)	92.76	0.84	3	0.71		
	DACL						
Daclinza tablet claimed to contain 60 mg DACL per tablet	PLS-1	97.81	0.89	3	0.79	1.311	1.05
	GA-PLS-1	99.09	1.10	3	1.21	0.373	1.46
	CRACLS	97.84	1.06	3	1.12	1.160	1.35
	GA-CRACLS	99.17	0.89	3	0.79	0.527	1.05
	SRACLS	97.48	0.88	3	0.77	1.781	1.08
	GA-SRACLS	99.16	0.93	3	0.86	0.507	1.04
	Reported method (5)	98.78	0.91	3	0.83		
Valysernex caplet claimed to contain 500 mg DACL per caplet	VALA						
	PLS-1	99.80	0.77	3	0.59	1.076	1.22
	GA-PLS-1	98.22	0.96	3	0.92	2.257	1.28
	CRACLS	101.72	1.06	3	1.12	2.316	1.56
	GA-CRACLS	101.49	0.91	3	0.83	2.222	1.15
	SRACLS	97.93	0.95	3	0.90	2.649	1.25
	GA-SRACLS	97.99	1.04	3	1.08	2.436	1.50
	Reported method (28)	99.89	0.85	3	0.72		

[a] The values in the parenthesis are the corresponding theoretical values of *t* and *F* at ($\alpha = 0.05$)?.

model of choice is that with lowest number of LVs that shows no significant differences to that with LV of the lowest RMSECV. It was found that 7, 4, 7, 4, and 7 LVs are acceptable numbers for modeling the selected data for PLS models for SOFO, VALP, DACL, LEDI, and VALA, respectively, while 6, 4, 6, 4, and 6 LVs are acceptable numbers for modeling them for GA-PLS models.

CRACLS model

CRACLS is an improved method of the traditional CLS one. It can used to resolve the mixture of SOFO, VALP, DACL, LEDI, and VALA using the experimental data obtained from UV spectra of this mixture. Moreover, it can estimate the pure spectral profiles of all the five active compounds. To gain the most of benefits of CRACLS; the model with the minimum number of iterations showing no significant differences to the model built with the iterations of the minimum RMSEV is used. It was found that 6, 4, 7, 4, and 7 iterations are acceptable numbers for modeling the selected data for CRACLS models for SOFO, VALP, DACL, LEDI, and VALA, respectively, while 6, 4, 6, 4, and 6 iterations are acceptable numbers for modeling them for GA-CRACLS models.

The main advantage of CRACLS was that it has retained the qualitative benefits of classical least squares (CLS) and has maintained the flexibility of PLS modeling.

SRACLS method

SRACLS is another modification for CLS introduced to eliminate its disadvantage. In SRACLS, the coefficient matrix (**K**) is calculated and used to obtain a new estimate for the spectral matrix. Spectral residual matrix is then calculated and subsequently subjected to principal component analysis (PCA). The coefficient matrix (K) is then augmented with one or more loading vectors that give analogous predictability to PLS models. It was found that 6, 3, 6, 3, and 6 PCs are acceptable numbers for modeling the selected data for SRACLS models for SOFO, VALP, DACL, LEDI, and VALA, respectively, while 5, 3, 5, 3, and 5 PCs are acceptable numbers for modeling them for GA-SRACLS models.

In order to downgrade the model intricacy and/or empower the model predictability, the PLS, CRACLS, and SRACLS models for the studied drugs were constructed using the selected variables by the GA technique. While the built models with and without GA procedures displayed the same number of LVs, the PLS model paired with GA revealed lower root mean square error of calibration (RMSEC) and lower root mean square error of prediction (RMSEP), as shown in Table 3. While for the CRACLS model, pairing with GA lowers the number of iterations by 14.29% for both DACL and VALA. Furthermore, the use of variable selection procedure of GA with the built SRACLS model reduces the number of required PCS by 16.67% for SOFO, DACL, and VALA.

This improvement in the models predictability might be through finding out the most informative regions in the analyzed spectra.

Method validation and statistical analysis

RMSEP is calculated by summing all squared prediction errors during cross-validation and is an indicator of the reliability and predictive ability of the model. The lower the RMSEP value the higher is the predictive ability of the model. It can be used to find the optimum number of components. The "best" model consists of as few predictor variables as possible and shows the lowest or almost the lowest RMSEP. The accuracy of the predictions or total error can be expressed as the percentage as the relative root MSEP (RRMSEP) (33) which takes into consideration the mean of the real concentrations.

Trueness is verified by the absence of bias, which can be evaluated taking into account errors on both axes of slope and intercept for the real concentrations and the concentrations predicted by the model. So, BCMSEP can be a good indication for model precision or variance in the prediction (34). These parameters and other validation parameters were shown in Table 3.

The results obtained by applying the proposed PLS, CRACLS, and SRACLS methods for determination of SOFO, VELP, DACL, LEDI, and VALA in bulk powder were statistically compared to the reported methods (4, 5, 14, 30). Statistical study showed no significant difference between the proposed and reported methods regarding both accuracy and precision as the calculated t-values and F-values were less than the theoretical ones (35), as shown in Table 4.

Analysis of real market sample

Applicability of the optimized models was tested using real samples of pharmaceutical dosage forms for the simultaneous determination of SOFO, VELP, DACL, LEDI, and VALA. The statistical analysis of the obtained results by the proposed methods and those obtained by the reported methods for the analysis of Harvoni (14), Epclusa (4), Daklinza (5), and Valysernex (30) showed lower calculated values of t-test and F-test than the tabulated ones. That indicates no significant differences between the proposed and the reported methods, as shown in Table 5.

Conclusion

Multivariate calibration techniques have resolution power that can resolve complex mixtures of five components with strong spectral overlapping. The use of full spectral data still has some uninformative variables that might increase the model complexity. Variable selection along with multivariate techniques get the advantage of giving more robust models with removal of uninformative variables resulting in less complex and/or better model predictability without deteriorating its performance.

Six multivariate calibration models (PLS, GA-PLS, SRACLS, GA-SRACLS, CRACLS, and GA-CRACLS) have been introduced as powerful chemometric methods for simultaneous analysis of SOFO, VELP, DACL, LEDI, and VALA in their laboratory prepared mixtures and pharmaceutical dosage forms. Coupling with GA as a variable selection step before these chemometric models was proved to be more advantageous over the models built with full spectral data in producing less complex models with improved predictability. They offered lower RMSEP, less LVs, lower SD, and better regression parameters. Thus, by demonstrating the results in hands, GA models can be recommended as variable selection models as preprocessing step before building the conventional chemometric methods. The proposed methods are very strong competitors for the usually used HPLC method in analysis of complex mixtures in quality control laboratories

Conflict of Interest

The authors declare that there is no any funding resources ar any conflict of interest with other institutions and all the work is original and belong solely to the authors.

References

1. Sofia, M.J. (2019) *HCV: The Journey from Discovery to a Cure*, Vol II, Arbutus Biopharma, Inc, Warminster, PA, USA
2. Smolic, M., Vcev, A., & Wu, G. (2017) Update on Hepatitis C, IntechOpen, Croatia
3. Dominique, S.C. (2019) *Viral Hepatitis: Chronic Hepatitis C*, Springer International Publishing, Switzerland

4. Rezk, M.R., Monir, H.H., & Marzouk, H.M. (2019) *J. Spectrochimica Acta Part A* **213**, 159–166
5. Chakravarthy, V.A., Sailaja, V.,B.B., & Kumar, A.P. (2016) *Asian J. Pharm. Clin. Res*. **9**, 61–66
6. Prasad Reddy, D.V., Gurupadayya, B.M., Manohara, Y.N., & VIjaya Bhaskar, V. (2007) *Asian J. Chem*. **19**, 2797–2800
7. Feld, J.J., Jacobson, I.M., Hézode, C., Asselah, T., Ruane, P.J., Gruener, N., Abergel, A., Mangia, A., Lai, C.-L., Chan, H.L.Y., Mazzotta, F., Moreno, C., Yoshida, E., Shafran, S.D., Towner, W.J., Tran, T.T., McNally, J., Osinusi, A., Svarovskaia, E., Zhu, Y., Brainard, D.M., McHutchison, J.G., Agarwal, K., & Zeuzem, S. (2015) *N. Engl. J. Med*. **373**, 2599–2607
8. Everson, G.T., Towner, W.J., Davis, M.N., Wyles, D.L., Nahass, R.G., Thuluvath, P.J., Etzkorn, K., Hinestrosa, F., Tong, M., Rabinovitz, M., McNally, J., Brainard, D.M., Han, L., Doehle, B., McHutchison, J.G., Morgan, T., Chung, R.T., & Tran, T.T. (2015) *Ann. Intern. Med*. **163**, 818–826
9. Foster, G.R., Afdhal, N., Roberts, S.K., Bräu, N., Gane, E.J., Pianko, S., Lawitz, E., Thompson, A., Shiffman, M.L., Cooper, C., Towner, W.J., Conway, B., Ruane, P., Bourlière, M., Asselah, T., Berg, T., Zeuzem, S., Rosenberg, W., Agarwal, K., Stedman, C.A.M., Mo, H., Dvory-Sobol, H., Han, L., Wang, J., McNally, J., Osinusi, A., Brainard, D.M., McHutchison, J.G., Mazzotta, F., Tran, T.T., Gordon, S.C., Patel, K., Reau, N., Mangia, A., & Sulkowski, M. (2015) *N. Engl. J. Med*. **373**, 2608–2617
10. Nebsen, M., & Elzanfaly, E.S. (2016) *J. Chromatogr. Sci*. **54**, 1631–1640
11. Nalla, S., & Rao, J.S. (2017) *Wjpps* **6**, 1596–1601
12. Patel, J.P., & Chaudhary, A.B. (2018) *World J. Pharm. Pharmaceut. Sci*. **7**, 813–825
13. Abdallah, O.M., Abdel-Megied, A.M., & Gouda, A.S. (2018) *Biome. Chromatogr*. **32**, e4186
14. EL-Shorbagy, H.I., Elsebaei, F., Hammad, S.F., & Elbrashy, A.M. (2018) *Spectrochimica Acta Part A: Molec. Biomolec. Spectrosc*. **205**, 398–409
15. Hassouna, M.E.M., & Mohamed, M.A. (2018) *J. Analyt. Pharmaceut. Res*. **7**, 92–99
16. Wadie, M.A., Mostafa, S.M., El.Adl, S.M., & Elgawish, M.S. (2017) *J. Pharm. Biol. Sci*. **12**, 60–68
17. Kumar, T.A., Gurupadayya, B.M., Rahul Reddy, M.B., & Prudhvi Raju, M.V. (2011) *J. Pharm. Res*. **4**, 24–27
18. Bhavar, G.B., Pekamwar, S.S., Aher, K.B., & Chaudhari, S.R. (2014) *Int. J. Pharm. Sci. Rev. Res*. **25**, 53–58
19. Nariman, A.E., Ali, M.Y., Nagiba, Y.H., Mahmoud, A.T., & Mohamed, A. (2016) *J. AOAC Int*. **99**, 948–956
20. Mahmoud, A.T., & Adel, M.M. (2019) *Spectrosc. Lett*. **52**, 339–345
21. Haaland, D.M., Melgaard, D.K., & Wehlburg, C.M. (2002) *Appl. Spectrosc*. **56**, 615–624
22. Attia, K.A.M., Nassar, M.W.I., El-Zeiny, M.B., & Serag, A. (2017) *Spectrochimica Acta. Part A, Molec. Biomolec. Spectrosc*. **170**, 117–123
23. Haaland, D.M., & Melgaard, D.K. (2000) *Appl. Spectrosc*. **54**, 1303–1312
24. Attia, K.A.M., Nassar, M.W.I., El-Zeiny, M.B., & Serag, A. (2016) *Spectrochimica Acta. Part A, Molec. Biomolec. Spectrosc*. **159**, 1–6
25. Gerlach, R.W., Kowalski, B.R., & Wold, H.O.A. (1979) *Anal. Chim. Acta*. **112**, 417–421
26. Geladi, P., & Kowalski, B.R. (1986) *Anal. Chim. Acta*. **185**, 1–17
27. Stone, M. (1974) *J. Roy. Stat. Soc. B* **36**, 111–133
28. Wold, S. (1978) *Technometrics* **20**, 397–405
29. Attia, K.A.M., Nassar, M.W.I., El-Zeiny, M.B., & Serag, A. (2016) *Spectrochimica Acta. Part A, Molec. Biomolec. Spectrosc*. **156**, 54–62
30. Ramakrishna, V.S., kumar, D.R., Malleswara Rao, N.V.N., & Babu, C.R. (2012) *Int. J. PharmTech Res*. **4**, 1009–1014
31. Dinç, E., Özdemir, A., & Baleanu, D. (2005) *Talanta* **65**, 36–47
32. Brown S.D., Tauler, R., & Walczak, B. (2009) Comprehensive chemometrics chemical and biochemical data analysis, 1st Ed., Elsevier.
33. Sales, F., Callao, M.P., & Rius, F.X. (1999) *Analyst* **124**, 1045–1051
34. Alonso Lomillo, M.A., Domínguez Renedo, O., & Arcos Martínez, M.J., (2001) *Anal Chim. Acta*. **449**, 167–177
35. Smilde, A., Bro, R., & Geladi, P. (2004) Multi-way Analysis in Chemistry and Related Fields, John Wiley & Sons, Ltd