REVIEW

Assessment of computational approaches in the prediction of spectrogram and chromatogram behaviours of analytes in pharmaceutical analysis: assessment review

M. Malarvannan[1*], K. Vinod Kumar[2], Y. Padmanabha Reddy[2], Pallaprolu Nikhil[3], Dande Aishwarya[3], V. Ravichandiran[4] and P. Ramalingam[3*]

Abstract

Background Today, artificial intelligence-based computational approach is facilitating multitasking and interdisciplinary analytical research. For example, the data gathered during an analytical research project such as spectral and chromatographic data can be used in predictive experimental research. The spectral and chromatographic information plays crucial role in pharmaceutical research, especially use of instrumental analytical approaches and it consume time, man power, and money. Hence, predictive analysis would be beneficial especially in resource-limited settings.

Main body Computational approaches verify data at an early phase of study in research process. Several in silico techniques for predicting analyte's spectral and chromatographic characteristics have recently been developed. Understanding of these tools may help researchers to accelerate their research with boosted confidence and prevent researchers from being misled by incorrect analytical data. In this communication, the properties of chemical compounds and its relation to chromatographic retention will be discussed, as well as the prediction technique for UV/IR/ Raman/NMR spectrograms. This review looked at the reference data of chemical compounds to compare the predictive ability in silico tools along with the percentage error, limitations, and advantages.

Conclusion The computational prediction of analytical characteristics offers a wide range of applications in academic research, bioanalytical method development, computational chemistry, analytical method development, data analysis approaches, material characterization, and validation process.

Keywords Analytical quality by design, Chemometrics, Computational methods, Infrared spectra, Raman spectra, Spectrograms, UV–visible spectra

*Correspondence:
M. Malarvannan
malarvannan707@gmail.com
P. Ramalingam
drramalingamp@gmail.com
Full list of author information is available at the end of the article

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 2 of 42

Background

The use of computational chemistry in research has been well-acknowledged in recent years and afforded significant research outcomes [1, 2]. There are literature reports on computer code for analysing models, replicating processes, predicting models, and interpreting chemical compounds [3]. Unlike the drug discovery area, the validity of computational techniques in analytical chemistry yet to be explored as a comprehensive tool [4–6]. The computational approach in analytical research is important because simulations of chemical behaviour of an analyte are needed for modelling of analyte response relationship in instrumental methods. Of course, it can be viewed as a visual representation of the connection between the analytical experiment and theoretical prediction [4, 7].

In this era, new chemical entity research is needed in new drug discovery process for treatment, diagnostic, and biomarker research. At this juncture, spectroscopy and chromatography techniques are playing a vital role in the purification, identification, and characterization of the targeted chemical compound [8, 9]. In general, understanding and interpreting the spectrograms and chromatographic retention times of the new compounds is quite difficult for beginners if the researcher is a non-chemist [10]. But, knowledge of spectrogram and chromatography is very essential for researchers and plays a crucial role in the process of developing new drugs. Indeed, the level of expertise and awareness on the accuracy of computation tools could assist the researchers in speeding up the experiments with partial validity of the analytical data [11]. In the current scenario, still there are predatory journals publish data sets that are not reliable if they are not verified [12]. Here, researchers may utilize computational tools to verify the data before citing in their research [4, 7]. The prediction tools of various spectrograms like UV–visible, infrared (IR), Raman, nuclear magnetic resonance (NMR), and mass spectra are now widely accessible to researchers. Similarly, in silico approaches to predict the chromatographic behaviour of an analyte in various chromatographic techniques like HPLC and GC are in existence [13, 14]. The prediction of retention time (t_R) in chromatography is gaining much importance in analytical method development research. Several computational prediction approaches have been reported. Some of them are artificial neural networks (ANNs), response surface methodology (RSM), analytical quality by design (AQbD) [15], design of experiments (DoE), chemometrics, and quantitative structure retention relationship (QSRR) methods [16]. Although the knowledge about artificial intelligence software is limited, several artificial neural network-based programmes are widely available these days. Many researchers spend

a significant amount of time on their experimental work, even though they are shortcomings in computational chemistry. The AQbD and QSRR approaches explore the scientific understanding of critical method variables and method response in chromatography [17, 18]. These methods are still recommended in pharmaceutical method development because it allows regulatory flexibility [19]. In the AQbD approach [20], the tool used in the model development is DoE. In chromatographic research, the quantitative structure retention relationship (QSRR) is a reliable in silico method for predicting molecular systems [21, 22], and it can be used to evaluate complex physicochemical features of analytes in chromatographic analyses and for predicting chromatographic retention parameters [23, 24].

Considering the above discussion, the present assessment review focused on various prediction tools available, and accessible to resource-limited research setups. We have also explored the predictive ability of the different in silico tools with examples pertaining to the reference spectral library. Thus, this review can assist researchers in assessing the tool's reliability from case to case.

Main text

Problems involving the analytical methods

Today, the difficulties in analytical laboratories are the same as they had experienced in the past, although there has been advancement in analytical technology. Analytical laboratories experience difficulties related to the growth and preservation of expertise, maintaining the equipment sensitivities, and introduction of novel methodologies [25]. There are many reports on previous analytical issues with analytes, including method performance [26], a lack of regulatory flexibility [27], complex chemical processes [28], OOT-out of trend [29], and OOS-out of specification [30, 31]. This problem could be mainly raised by three stages such as pre-analytical, post-analytical, and development phase. These can be overcome by utilizing the most modern and advanced computational methods.

Pre-analytical phase

One of the crucial stages in the analysis of the sample is the pre-analytical phase it includes, gathering of literature, sampling, preparation of the sample, transport, and storage. This entire process is the most time-consuming and might occasionally lead to errors [32]. It is widely acknowledged that a degraded sample cannot produce good results. Always, it is important to conduct a literature review before beginning any research on an analyte. There are many databases, books, journals, and websites, but in some instances, information on new analyte may

not be available due to a lack of studies on the analytes, or newly synthesized materials, or a lack of source availability [33]. Next, for new analytical method development, the preparation of a sample is a critical step. A sample processing method is unique for each type of sample, including biological matrix, food products, active compounds, excipients, and pesticides. A given procedure cannot be applied to a different type of analyte without a complete revalidation of the method [34]. Unfortunately, this rule is regularly ignored. Finally, there are several issues with analyte that affect storage and transportation; they are temperature, humidity control, data storage maintenance, and a lack of advancement [35].

Development phase

The selection of the method, procedure, principle, technology, and appropriate recommendations are the main problems that arise throughout the development phase. Unfortunately, it must be acknowledged that no method has yet been developed that satisfies all of these criteria and appropriate for all classes of analytes. This always place restrictions on analytical chemists. It is also crucial to understand whether the analysis's objective is merely screening or accurate quantification. In developing chromatography methods, optimization includes temperature, flow rate, the choice of mobile and stationary phases, separation efficiency, internal standard selection, and validation. Thus, re-optimization are difficult task, if the method fails during method transfer [36]. In the last decade, new chromatographic techniques for the detection of bio-analytes have emerged. One of these techniques is tandem mass spectrometry (LC–MS/MS), which has advantages such as high selectivity and sensitivity but possess disadvantages such as expensive equipment, experienced operators, and more challenging method development [37, 38]. In the development of electrochemistry supported instrument, the general settings for resolution, path of the composite electrochemical response examination, and optimal path of analysis of the multidimensional data are complicated [39].

Post-analytical phase

In this phase, the key challenge is the collection and interpretation of data with analytical techniques, particularly when it comes to clinical research, proteomics, and metabolomics. Additionally, certain sophisticated computations raised problems from the data analysis as well. In general, manual calculations can produce inaccurate findings. From pre- and post-data analyses in chromatography methods, the common troubles are unwanted background signals, baseline drift, unresolved peaks shifting and retention durations, data comparison errors, and improper retention time alignments which are to be addressed [40]. In spectroscopic analyses, specific mathematical transformations that are frequently created for a certain experimental approach are typically used to rectify systematic undesirable signal changes. Baseline shifts (offsets), horizontal shifts, drifts (slope changes), and global intensity effects are some of the systematic signal fluctuations. The significant alteration of signal profiles produced by the derivation transform can mislead the interpretation of final results [41]. Overall, the scheme of application of computational method is shown in Fig. 1.

Prediction of spectrograms

Prediction of C^{13}-NMR and H^1-NMR

NMR is a significant tool for detecting carbon and hydrogen atoms in organic compounds. In the pharmaceutical industry, C^{13}-NMR and H^1-NMR are used to assess drug purity, composition, and chemical shifts of diverse organic molecules. NMR parameters are now calculated by utilizing computational methods in association with chemical structures. AI has created several software tools (e.g. ChemDraw, Chemaxon, etc.) that are now used to predict chemical shifts in H^1-NMR and C^{13}-NMR and offer net intensity, quality, and spectrograms.
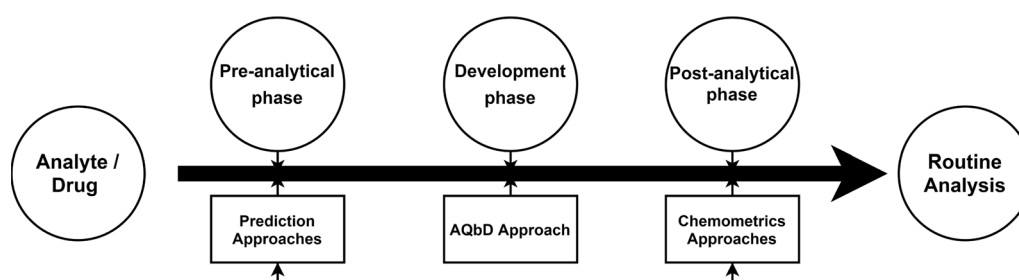

Fig. 1 Scheme of application of computational methods in analytical process

Machine learning approach in NMR prediction

Machine learning (ML) approaches are more beneficial and, in most cases, faster than prediction-based databases like HOSE codes. The database works by finding structural similarities and averaging the experimental data for chemical structures. The similarity between the new and known HOSE codes has little bearing on the accuracy of the prediction. The well-established structure determination approach formerly relied on quantum chemical calculation-based methods such as topical-based DFT calculation. This method is accurate for H1 and C^{13} chemical shift predictions, but considerably more time-consuming and expensive. Today, software tools have been designed to speed up the procedure. The NMR signal characteristics can be visualized more accurately using a machine learning method called "Automatic structure verification (ASV)" based on variables such as temperature, solvent, pH, salt content, concentration, and so on which will affect chemical shifts in laboratory studies. All of these parameters have considered, such a way that NMR can predict the chemical shift for an unknown structure. But, certain other prediction algorithms take some of them into account, still the prediction systems produce variable values. But, the ASV system is capable of properly dealing with overlapping peaks. This is especially important when sections of the compound's relevant peaks, such as significant solvent peaks, are quite close to other signals [42–45]. Few researchers have used this approach, including Jia et al. [46], who have developed a method for extracting data from previously examined ^{13}C and ^{1}H NMR spectra in order to recognize the NMR spectrum. Min Lin and colleagues predicted the chemical shifts based on cutting-edge machine learning [47].

Software handling for NMR Signal prediction

The user can either use a software application to draw the chemical structure of the test molecule or download and paste it into the software. The user will be able to locate the predicted C^{13}-NMR and H^{1}-NMR spectra in 1–5 min after clicking the calculation button. The user can optionally alter the frequency range from 60.0 to 1000 Hz after the prediction. Finally, a pdf document will be generated including the substance's chemical shift, peak intensity, peak quality, molecular location, and coupling constant values [48]. A typical H^{1}-NMR signal for Zidovudine is shown in Fig. 2.

Prediction of UV–Visible Spectra

The UV–Vis absorption spectrum of an organic substance is a key component of its physical makeup. Using predictions of UV–Vis spectra from molecular structural formulas, it is generally quite interesting to design new materials, find potential phototoxic chemicals, and estimate missing spectroscopic data for known molecules [49]. In a recent study, Chan et al. [50] utilized TD-DFT computation approach for rapid ultraviolet–visible spectrum prediction. The method was developed by Urbina et al. [51] using neural network-based computation to predict UV–visible spectrograms.
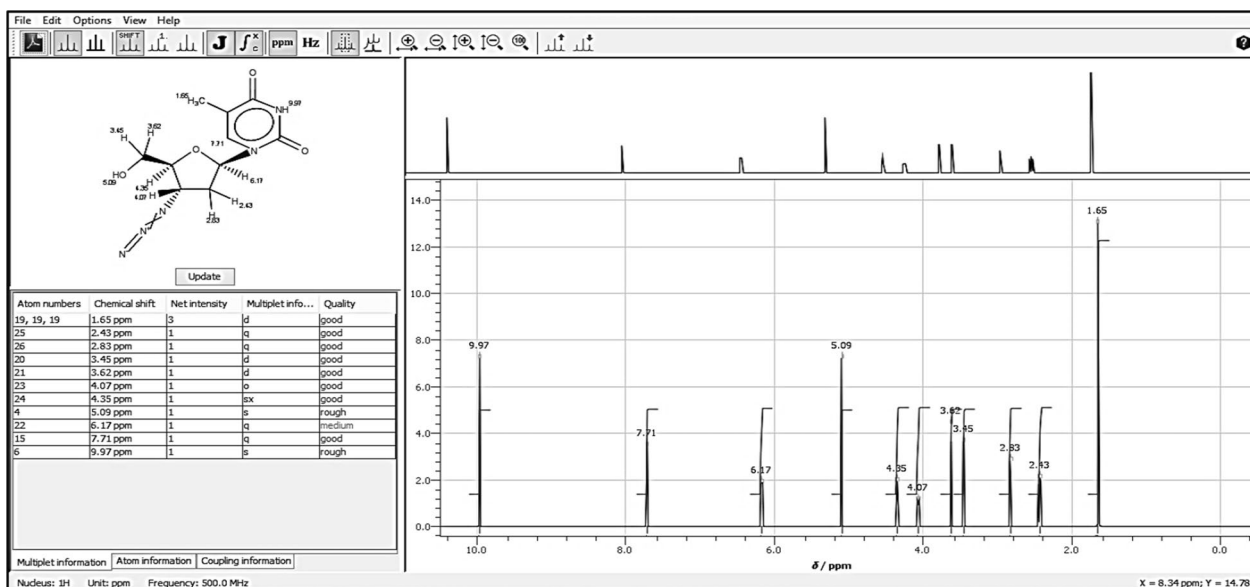

Fig. 2 Typical predicted H^{1}-NMR signals for Zidovudine

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 5 of 42

Time-dependent density functional theory (TD-DFT)

For TD-DFT calculation, the software should be able to analyse the energy of the chemical structure in the excited states, and the probability of transition between energy levels for the chemical molecule. For example, the ORCA programme contains several methods for accurately determining excited state properties. The TD-DFT technique is the most effective of all the approaches. For precise results in this method, an optimized geometry file of the chemical structure is required. To optimize the structure, the user might utilize the "IQmol" software package or another. After that, the user can use Notepad + + to create the input file, with the function code "! B3LYP def2-TZVP", "RIJCOSX" code to speed up the process, "% TDDFT" code to automatically generate the excited state calculation, "NROOTS" flag to determine how many excited states to be added, and "MEXDIM" to determine the maximum dimension of the expansion space. To simulate the analyte employed in the experiment, CPCM may be a solvation model for both the ground and excited states. The number "0" denotes charge, whereas the second number denotes multiplicity. Finally, from the same folder, save this file in "inp" format (tddft.inp). The user may then go to the folder and input a comment "orca tddft.inp > tddft.out" followed by "Enter" to execute the computation on the CMD line (comment prompt). Depending on the molecules involved, it might take some time (10 min–2 h). After the computation is completed, the programme creates an output file in the same folder that contains all of the data [52, 53].

Visualization of UV–visible spectra

The UV–Visible spectrum can be obtained for an unknown analyte instantly using a graphical interface. It does indeed show thin line spectral waves, but some line broadening is required to make the predicted spectra match the experimental one. This is easily accomplished by selecting "Advanced > >" and then, on the "Infrared Spectra Settings" tab, adjusting the "Peak Width" to 10-30 cm^{-1} [54–58]. Figure 3 shows the generated spectrogram of Zidovudine compound.

IR/Raman predictions

For chemical characterization and identification, both infrared (IR) and Raman spectroscopy continued to be essential tools. Recently, McGill et al. [59] developed the IR spectrum prediction procedure using a neural network-based approach. IR and Raman spectra may also be predicted using the ORCA software. It uses "Avogadro" or "IQmol" to compute the frequencies of the molecules. The 3D structure of the analyte is to be analysed and optimized. The ORCA programme can create output on its own. The user must create a new folder and set the optimized geometry structure and input file, similar to the UV–visible computations. The "! B3LYP DEF2-SVP" is the function code, while "OPT FREQ" specifies multiplicity. Finally, save the file in "inp" format in the same location so that the user may navigate the folder and execute "orca foscarnet.inp > foscarnet.out" followed by "Enter" to perform the computation. The output file can be created in the same folder when the operation is finished [7, 54–58]. Figure 4 shows the predicted IR spectrum of foscarnet generated by Avogadro.

Plotting a spectrum

Using Avogadro as a graphical user interface, the IR spectrum may be generated rapidly. To view the visual spectra in a new window, the user can open the saved output file and click "Show Spectra". Although it displays narrow spectral lines, some line widening is necessary to bring the predicted spectra as close to the observed one as possible. This can be readily performed by selecting
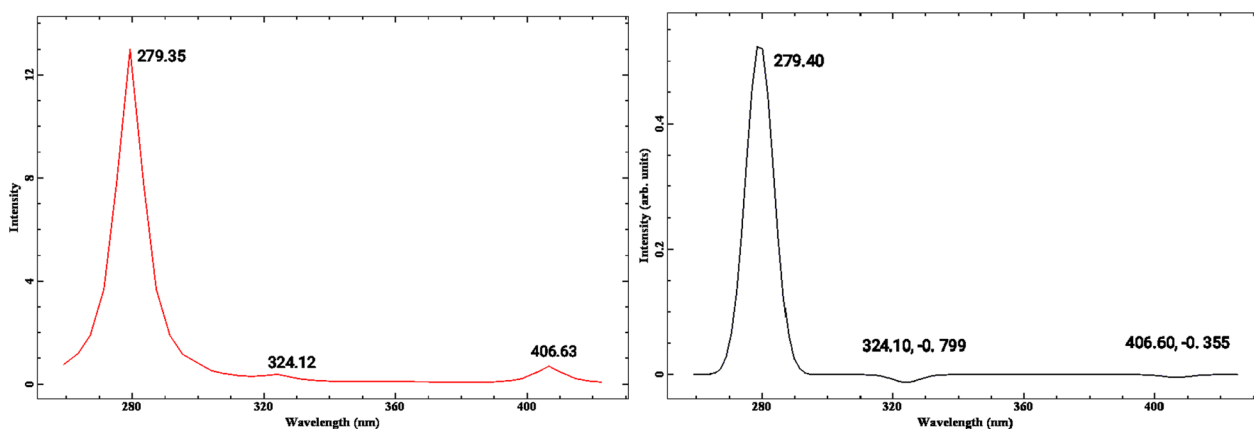

Fig. 3 Predicted CD and UV spectrum of zidovudine generated by Avogadro
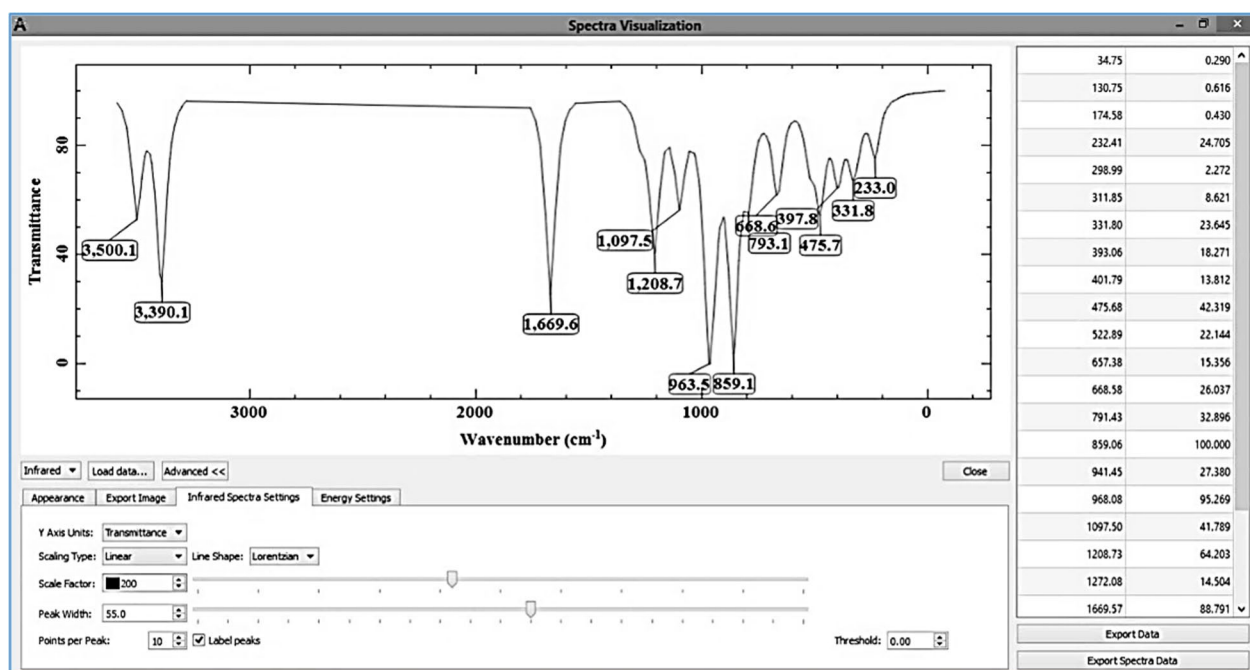
Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 6 of 42

Fig. 4 Predicted IR spectrum of foscarnet generated by Avogadro

"Advanced > >" and then changing the "Peak Width" to 30–130 cm^{-1} on the "Infrared Spectra Settings" tab [54–56].

Mass spectroscopy predictions

The molecular weight of an analytes in pharmaceutical studies is determined by mass spectrometry (MS). In an electron ionization mass spectrometry (EI-MS), an electron beam positively ionizes and fragments the molecules [60]. According to the mass-to-charge (m/z) ratio, the mass spectrum is a distribution of the frequency or intensity of each type of ion [61]. The prediction models calculate the chance of each bond breaking under ionization and the frequency of each ion fragment by using quantum mechanics calculations [62] or machine learning [63]. For large molecules, model's prediction can consume few minutes, depending on the molecule's size. This due to the fact that these techniques must either utilize sophisticated computations to determine molecular orbital energies with high accuracy or stochastically mimic the fragmentation of the molecule. A neural network termed neural electron ionization mass spectrometry (NEIMS) predicts the electron ionization mass spectrum for a particular small molecule and is studied by Jennifer N. Wei and colleagues. Additionally, they found that the forward-only model fails to adequately capture the fragmentation events, but the bidirectional prediction mode does [64] because it directly predicts spectra rather than bond breaking probabilities. As a result, this model is significantly faster than previously reported methods.

Wang et al. utilized the recently developed quantum chemical programme QCEIMS (Quantum Chemical Electron Ionization Mass Spectrometry). QCEIMS can theoretically calculate the spectra for any given chemical structure. However, in order to make quick predictions, approximations and parameter estimations are required, which are important for the precision of QCEIMS predictions. For the MD trajectories, fragment ions are calculated by QCEIMS using Born–Oppenheimer molecular dynamics (MD) within picosecond reaction durations with femtosecond intervals. With this approach, they discovered that tweaking QCEIMS's parameters were not a practical way to enhance simulation outcomes [65, 66]. One of the best tools for in silico mass-spectrum-to-compound identification is CFM-ID, which Wang et al. used to predict more accurate ESI–MS/MS spectra. They added a new method for modelling ring cleavage that models the process as a series of straightforward chemical bond dissociations, and they expanded their handwritten rule-based predictor to cover more chemical classes of analytes [67]. They also listed parameters from molecular topological parameters.

Fluorescence spectroscopy predictions

Fluorescence spectroscopy measures a target analyte fluorescence upon being excited by a laser beam (often UV absorption) [68]. The prediction of analyte's fluorescence

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 7 of 42

features, including the type of fluorescence, emission, and excitation wavelengths [69], can be employed to examine included solvent effects. It has been used to predict the spectra for a variety of fluorescent compounds [70]. The majority of the predicted spectra have molecular masses of 228 or below. In such case, DFT technique can be used for larger molecular weight and chemical emission spectra calculation with solvent effects.

The characterization of electronic excited states depends on the accuracy of simulation spectrum of molecular absorption /or emission and precise techniques like the equation of motion coupled cluster singles and doubles (EOM-CCSD) [71, 72]. In order to increase the emission spectrum qualities, Caricato et al. [73] combined the EOM-CCSD and polarizable continuum (PCM) models and reported that the predicted values of vertical emission energies are in good accord with the available experimental data. Later, DFT was used by Powell et al. [74] to demonstrate the capability of predicted spectra in generating libraries of fluorescence spectra in a digital format. Ye et al. concluded that the statistical requirements for the numerically predicted wavelength were satisfied by the Lasso-RF (Random Forest descriptor) model. Four conjugated bonding-related characteristics were found by the model to contribute primarily to the predicted emission wavelength [75]. Furthermore, Shams-Nateri et al. [76] investigated the link between absorption and emission spectra using the PCA chemometric approach, and they found that the accuracy of emission spectra prediction was improved with the addition of more principal components.

Electrochemistry predictions

Because of the growing interest in electrochemistry as a potential drug core structure and for the development of organic photovoltaic materials, it has recently experienced a huge comeback and provided valuable prediction, filtering, and active learning. This includes a promising optimization of the electrochemical properties of the analytes, investigation of intrinsic electron deficiency, and rendering of the connection between electronic characteristics and substituent effects [77]. Using electrochemistry predictions of compounds using quantum mechanical calculations provides a quick and accurate method for the research. For instance, DFT is regarded as the "workhorse" of recent theoretical investigations in electrochemistry and physics [78].

Electrochemical systems are studied using the popular electrochemical impedance spectroscopy (EIS) characterization approach. The significance of this method is still constrained by several issues. EIS is also extensively employed in the development of sensors [79, 80], in health care [81], drug release [82], testing, and biology

[83] because EIS makes it possible to characterize such systems and helps in identifying crucial variables like conductivities [84], resistances [85], and capacitances [86]. The computational Gaussian processes (GPs) used in this method faced significant challenges including noise, impeded spectrum regression, polarization resistance, and probed frequencies that were not always ideal. An infinite or finite collection of random variables is referred to as a GP, if the joint distribution of any finite subset displays multivariate Gaussian behaviour. Then, GPs may regress and predict it using a prior distribution and a set of assumptions on the characteristics of the observed unknown function [87]. Regression and prediction uncertainty can be measured using GPs and also have so far been used to filter data, predict parameters in diverse situations [88], and enhance experiments in the active learning domain. Liu and Ciucci et al. [89, 90] used GPs to de-convolve the distribution of relaxation duration, a novel approach for EIS analysis. Then, using a finite GP approximation, Maradesa et al. extended this framework to constrain the DFT to be non-negative. Additionally, Py et al. [91, 92] created and validated the method that Ciucci used to assess the quality of EIS spectra using GPs that complied with the Hilbert transform.

Kiss et al. [93] predicted the substituent effects in electrochemical properties of the analyte and comprehended the influence of substituents on the character of the electronic transition and transition density matrices (TDMs). This procedure makes it possible to access the distribution of electrons and holes in the excited state and determine their delocalization. This makes it possible to reveal electronic excitation processes like charge transfer [94]. The imbalance in the TDMs is caused by the presence of electron-donating and electron-withdrawing groups interacting with the hole. The location of the hole is altered when an electron-donating moiety uses mesomeric effects to donate electron density to the hole. Instead of being just inductively impacted, at this instance, the TDM can be described as mesomerically effected. On the other hand, the inductively dominated TDM lacks any localization due to the absence of any major TDM elements on the analyte. The polarity difference has a significant impact on the mesmeric contribution to the TDM. This made it easier to spot the impacts of charge transfer and substitution.

The next field of research addressed the exciton binding energies, which show the Coulomb attraction between the exciton quasiparticles (electron and hole). It is a measurement of the exciton's separability in free charges, and it has a direct impact on how an effective current is produced in optoelectronics [95]. More details on the impacts on the characteristics of the electronic structure are revealed by analysing the HOMO and

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 8 of 42

LUMO energies (EHOMO and ELUMO) [96]. In order to optimize the electrochemical characteristics of an analyte, Min et al. [97] developed and verified a machine learning (ML) approach for electrochemistry. Both output (such as initial capacity and cycle life) and few input (synthesis parameters, ICP-MS data, and X-ray diffraction (XRD) results) variables were used to build several experimental datasets for analyte [98]. When distributing these variables across the entire dataset while building the ML model, a number of primary variables were chosen to serve as suggestions for the optimal experimental parameters.

Prediction of chromatographic retention behaviour
Quantitative structure retention relationship (QSRR)
QSRR is a computational approach for linking chemical structural variables to chromatographic column retention behaviour. Here, Y-variables are frequently employed as dependent variables for predictive or explanatory purposes, whereas X-variables are utilized as independent variables. As a result, Y-variables in QSRR

have connected to solute chromatographic retention, whereas X-variables encode solute molecular structure. QSRR was first used to characterize columns by quantitatively comparing their separation qualities or to supply knowledge for predicting retention mechanisms in various chromatographic settings [22]. A typical QSRR study includes building a retention database of compounds with known chemical structures, computing molecular descriptors for each structure, choosing descriptors, creating a QSRR model, and validation. Figure 5 illustrates a QSRR methodology and work flow.

The most popular methods for expressing chemical structures are molecule 1D descriptors, 2D descriptors, and 3D descriptors. While representing a connection table or a molecular graph, the chemical structure of the solutes of interest is used to compute 2D descriptors, whereas 1D descriptors provide simple chemical information about a solute, such as molecular weight or the number of oxygen atoms in the structure. A molecular descriptor that describes both the general surfaces/or
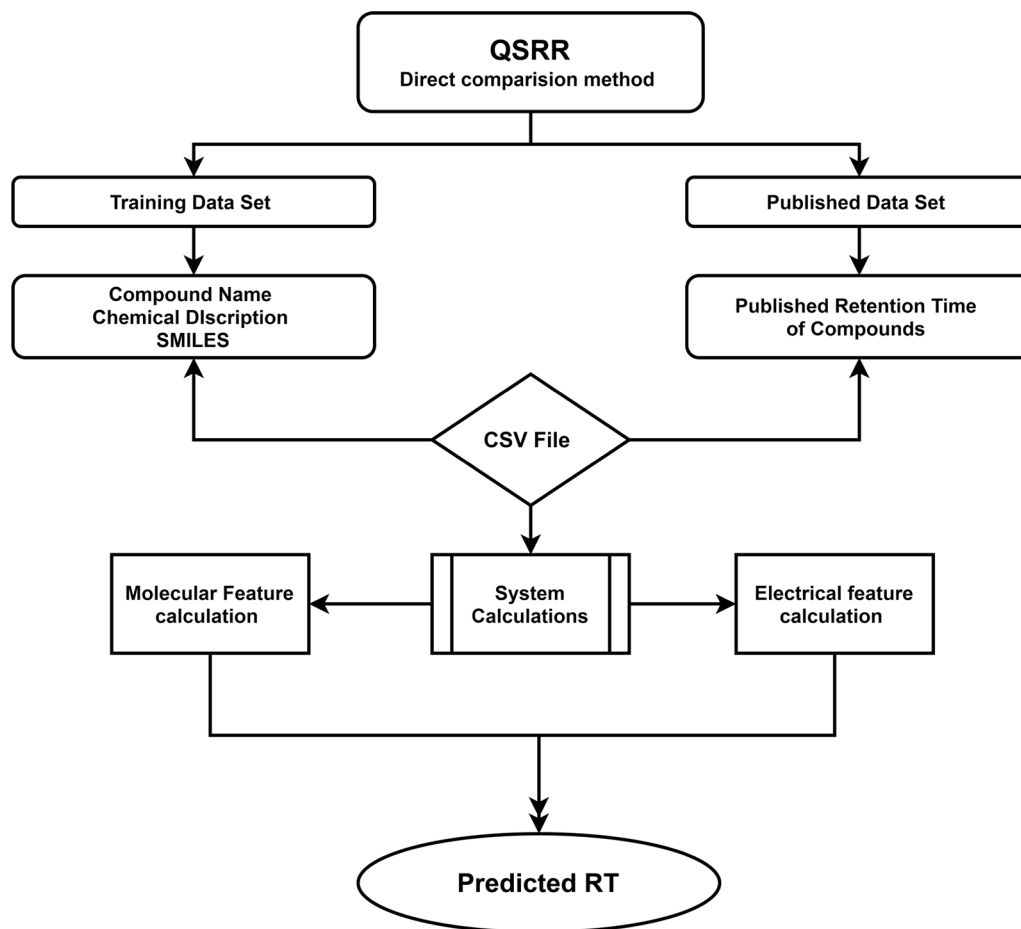

Fig. 5 Scheme of QSRR methodology in chromatography

volumes of molecules and 3D arrangement of structural attributes is known as a 3D molecular descriptor [23].

Depicting the molecular structure of QSRR is one of the key concerns in QSRR modelling. Molecular descriptors that describe chemical structures are typically categorized as physicochemical descriptors and descriptors of the quantum chemical, topological, etc. [99]. The fact is that physicochemical descriptors have a positive correlation with solute retention on chromatographic columns. On a molecular level, quantum chemical descriptors shed light on the process of chromatographic retention, although the link to solute retention is frequently poor, and the calculation is laborious. With today's computational technologies, topological descriptors are easily constructed, but they are unrelated to retention phenomena [24]. There are two methods of the QSRR approach, viz., the direct mapping method and the direct comparison method.

Prediction of retention time by the direct mapping method

It is a simple method for predicting compound retention time on a chromatographic column. It is a web-based solution that allows users to predict retention by submitting their data and receiving expected retention values. Predict is a database available, and this experiment has four steps as follows.

The user can create a CVS file that includes the compound name, real retention time from the PubChem CID or InChIs databases, and stereo-chemical parameters. The user must be able to upload retention data and get new retention predictions easily using a web interface. On the website, the user is initially asked to create a new chromatographic system. Each system will contain two types of columns: (1) a name and (2) a column type (for example: RP, HILIC). (3) column description (for example, Waters and Symmetry C18 columns), (4) eluent system (for instance, 95:5 methanol/water), (5) The eluent's pH (for example, acidic or alkaline), and (6) Eluent additives (for example, 0.1 per cent trifluoroacetic acid). The user will next submit a CSV file containing retention times for chemicals derived from their studies or google scholar in the following phase. Finally, the user may obtain the estimated retention time by clicking "get a prediction" [58].

Prediction of retention time by direct comparison method

QSRR Automator, a python-based software, can be used to predict retention using the direct comparison method. Mordred, a software package that uses the rdkit package, can be used to determine molecular descriptors. Machine learning operations may be performed with the sci-kit learn package. The following is a description of the QSRR Automator Workflow. The training data, which contains the name of each chemical, the structure in the form of a simplified molecular input line entry system (SMILES) text string, and the retention duration, may be created by the user. The programme creates a template and simplifies the input file on its own. After that, the user may submit their training data (chemical descriptions, SMILES, compound name, and actual retention time). The structural and electrical descriptions to be utilized should be used. Functional groups, hybridizations, the number of carbon atoms, and the ring system are all structural properties. Aromaticity and numerous electronegativity calculations are two electrical properties. All of these calculations are simple; unlike more complex fingerprint feature combinations, they can all be done using the Mordred software package, which calculates over 1500 features [100]. The recent data on the QSRR based method were listed in the Table 1.

Chemometrics in chromatography

The chemometric approach is widely used in separation science to predict the analysed peak asymmetry, peak overlapping, and peak optimizations. Co-elution of multiple analytes in chromatography significantly complicates quantification of the target analyte due to interference caused by incorrect method optimization. At this juncture, chemometric methods such as principal component analysis (PCA) are widely used in separation science and have now been extended to LC-HRMS analysis for proteomics and metabolomics. In addition, artificial neural networks (ANN), factorial design (FD), partial least squares (PLS), and cluster analysis (CA) are also in place [113, 114]

Chemometrics in one- and two-dimensional chromatography

In the development of two-dimensional (2D) chromatography, the entire first-dimension (1D) effluent is divided into many fractions, each of which is subjected to 2D separation. Two-dimensional chromatography is created by combining the results of 1D liquid chromatography separations (LC×LC). The placements of the spots provide qualitative information, while the intensities of the spots provide quantitative information. However, extracting information from extremely complex molecules like protein digests, metabolic extracts, and oil mixes can be problematic. Even with modern high-resolution chromatography, extracting the entire information of a complex matrix remains a challenging task. Many researchers are constantly working to improve the efficiency of chemometric data processing strategies.

Table 1 Data on the analytical methods reported based on QSRR

S. No.	Research	QSRR Tools/software	Data set	Experimental method	Statistical model for data set	Statistical results	References
1	QSRR for improving structural identification in non-targeted metabolomics	ChromSword—to build the QSRR model, ACD/ChromGenius—descriptors calculation	34 groups of compounds (248 compounds in total)	HPLC/MS Agilent 1100 HPLC QTOF-2 mass spectrometer	PLS (Partial least squares)	R^2=0.8512, %RMSE=8.45, SD=46, N=191	[101]
2	QSRR of ionic liquid cations for characterization of stationary phase for HPLC	Hyperchem 7.5 package with Chemplus—for molecular modelling	8 ionic liquid cations	HPLC-Shimadzu-LC-10Avp pump with manual injector	MLR (Multiple linear regression), PCA (principal component analysis)	Log k versus logP, R^2= >0.9 for 95% confidence intervals. (For 6 different columns) PCs- 97.3% total data variance	[102]
3	QSRR for HPLC Rt prediction of small molecules: endogenous metabolite	Gaussian 03—3D structure calculation, Dragon software version 5.5—for descriptor calculation	9 classes of banned drugs, metabolites, and salts (totally of 146 analytes)	LC–MS method	MLR (Multiple linear regression), PLS (Partial least squares)	R=0.97, R^2=0.95, RMSE=3.45, F=778	[103]
4	QSRR of peptides behaviour in RPLC	MATLAB 7.0.1, Dragon software 5.0, Hyperchem 7.5 package	Training set 50 and test set 19, totally of 69 peptides were used	Four different columns with Merck-Hitachi LaChrom system	GA (Genetic algorithm) to MLR (Multiple linear regression)	R^2= 0.62–0.86, RMSE=1.99–2.99	[104]
5	QSRR analysis of antihypertensives and diuretics	Chempro, Chemaxon, Marvin 4.0.5, Gaussian 98	Totally 17 compounds	SpectraPhoresis 500-Capillary electrophoretic system equipped with UV detector	Optimized ligands and ligand—BCD complexes	R^2=0.914, %RMSE=0.092, F ratio=45.897	[105]
6	QSRR modelling of morphine and its derivatives	Hyperchem 8.0, VLifeMDS 4.4—descriptors calculation, Molegro data modeller	Training set 42 and test set 15, totally 57 morphine and its derivatives	GC–LC method	ANN (Artificial neural network)	R^2=0.9559, %RMSE=0.399	[106]
7	QSRR modelling of coffee flavour compounds	Hyperchem, CODESSA—descriptor calculation	Training set 36 and test set16, totally 52 compounds	NA	SVM (Support vector machines)	Satisfactory relationship between the molecular descriptor and retention time of coffee	[107]
8	QSRR for chromatographic behaviour prediction of antiviral drugs	Hyperchem 8.0, ACD/ChemSketch, QSRR Automator, ADMETlab	Training set 14 and test set 5, totally 19 antiviral drugs	Agilent LC-1200 system. (RP-HPLC Method)	MLR (Multiple Linear Regression), SVR (Support vector for regression)	R^2=0.881, RMSE=0.640	[108]
9	QSRR for naturally occurring phenolic compounds	DRAGAN version 3, MOPAC2009—descriptor calculation	Training set 30 and test set 9, totally 39 phenolic compounds	RPLC-MS method	ANN (Artificial neural network), MLR (Multiple linear regression)	R^2 >0.80, N=30	[109]
10	QSRR for prediction of pesticide retention time	ACD software, ChemDes—descriptor calculation	Total 843 pesticides	UHPLC method	MLR (Multiple linear regression) and SVM (Support vector machines)	R^2=0.85, RMSE=0.69	[110]

Table 1 (continued)

S. No.	Research	QSRR Tools/software	Data set	Experimental method	Statistical model for data set	Statistical results	References
11	QSRR modelling of selected antipsychotics and their impurities	Chemaxon, MOPAC/AMI of chem3D ultra-7.0.0	The training set 147, test set 31, and validation set 32, totally 210 compounds	Thermo scientific Dionex ultra-3000 (RP-HPLC method)	ANN (Artificial neural network)	Training set $R^2 = 0.9962$ RMSE = 0.2954 Test set $R^2 = 0.9829$ RMSE = 0.3633 Validation set $R^2 = 0.9927$ RMSE = 0.4864	[111]
12	QSRR for mycotoxins and fungal metabolites	ISIS Draw-2.5, HyperChem, CODESSA2.63—descriptor calculation	Total 83 compounds	LC–MS/MS method	RBFNN (Radial basis function neural network)	Training set $R^2 = 0.8933$ RMSE = 0.9406 Test set $R^2 = 0.8709$ RMSE = 1.2892	[112]

F Variance ratio, *R* Correlation coefficient, R^2 The square correlation coefficient, *RMSE* The root-mean-squared error

In chromatography, chemometric is an appreciable tool for pre- and post-data analysis to resolve undesired background signals, baseline drift, unresolved peaks, and shifting retention times. Chemometric-based data interpretation, information extraction, and pre-data processing can significantly increase the analytical performance of an existing technique. The various chemometric approaches used in chromatography are penalized partial least squares (PPLS) approaches, multivariate curve resolution and orthogonal subspace projection for background correction, local minimum value approach, baseline estimation, and denoising using scarcity, retention-time-alignment strategies, peak clustering, and principal component analysis (PCA). These methods highlighted the chemometric techniques as the most progressing in silico approach in 1D and 2D chromatography and spectroscopy [115].

Chemometrics in unsupervised and supervised techniques
For understanding the dissimilarity or variance in the data matrix, PCA, independent compound analysis (ICA), and cluster analysis (CA) are used. As a result, the "calibration sets" may be defined as loading vectors and utilized to project unknown data. If data does not cluster against any objective criterion, then supervised procedures such as multivariate calibration methods are applied. Although a regression model may be built utilizing a large number of PCA variables, this approach is referred to as principal component regression (PCR). The data matrix's PCR analysis is mainly based on variance. The partial least squares (PLS) method, also known as a projection to latent structures, is commonly used in the linear supervised method. It finds the route through the data matrix that maximizes the covariance between the matrix and the predicted variable and then creates a regression model [116].

Software tools in chemometrics and their workflow
Chemometric software (for example, BWIQ) is available for on- and off-line quantitative and qualitative spectral measurements to identify principal components. The software classifies the sample as corresponding to the group with the shortest calculated "Mahalanobis distance (a measure of the distance between point-P and distribution D)". The workflow is described in following section.

The complete spectrum will be presented on the screen once you start the software, click "file", open the data, and import it into the software. We may designate spectral files in BWIQ in a variety of ways, including calibration, validation, and ignored files. The "usage" column's dropdown button was used to manually designate the spectrum. The algorithm parameters have been chosen and are accessible in the algorithm properties tab. We may use the sampling method and adjust the calibration file to the o validation file ratio in the property panel, for example, 60:40 (calibration: validation). After that, eliminate any change in the unrelated to chemical variations data sets but rather to scattering, instrumental fluctuations, spectral noise, or background differences in the pre-processing processes. Because the model can analyse the full spectrum, it will be more sensitive to contaminants or changes in the samples that add signals in other spectral areas. However, excluding non-informative or noisy data areas from analysis is an advantage. Then, we have the option of using a chemometric method such as PCA-Mahalanobis distance (MD). In principle component space, the scores plot illustrates the sample clusters. The result shows clusters matching the different classes of principal components. Additional graphs, such as loading and variance, are also available [117].

Different types of chemometrics approaches
Penalized partial least squares approach (PPLS)
This method was initially developed by Whittaker in 1922 to address signal smoothing issues [118]. The goal of PLS is to approximate observed data by resolving conflicts between original data fidelity and the imprecision of fitting data more easily by resolving the model's fit to the data [119]. Assume that Eq. (1) is used to calculate the fidelity and roughness combined in a balanced way:

$$Q = F + \lambda R = \sum_{i=1}^{n} (v_i - z_i)^2 + \sum_{i=2}^{n} (z_i - z_{i-1})^2$$
$$= \left\| v - z^2 \right\| + \lambda \left\| Dz^2 \right\| \quad (1)$$

where z is the fitting vector and v is a vector representing the analyte spectrum, both of which have a length of "n" elements. Fitted z should maintain both the roughness of the fitted vector and fidelity to v. The sum of squares of differences between the vector and element of z and its neighbours can be used to describe F, which stands for fidelity to the analyte spectrum "v", and R, which stands for the roughness of the fitting vector z. A user-adjustable parameter called "λ" finds a balance between fidelity and roughness. Greater λ favours a fitted vector that is smoother.

A weight vector w was added for fidelity in order to use the PLS to estimate the background. Its element w_i may be thought of as a weight that represents the dependability of point I as a component of background. The partial derivatives of Q are equalled to zero $\left(\partial Q / \partial z = 0 \right)$, in order to solve the minimization issue of Eq. (1). The matrix form of the resulting linear system is then used to determine the fit (Eq. 2).

$$(w + \lambda D'D)z = v \tag{2}$$

To use this PLS approach for baseline correction, which is used by Zhang et al. and Cobas, one must first identify the locations of the chromatogram's peaks. In order to determine whether a data point in the chromatogram relates to background or a peak, respectively, a binary mask or weighted matrix can be generated once these peak points are known [120, 121].

$$(w + \lambda D'D)z = wv \tag{3}$$

Additionally, Eilers et al. [122] created the asymmetrical least squares (asLS), which introduces an asymmetry parameter in an effort to address this problem. The weights assigned to positive and negative deviations from the baseline can now be less and bigger, respectively. However, this also takes into account of issues with the baseline that were raised for the introduction of adaptive iteratively reweighted penalized least squares (airPLS) [123], which enables some baseline regions to be fined more than others. By iteratively resolving a weighted penalized least squares problem, airPLS develops a weight vector.

Once the difference between the signal and the fitted vector $|d^t|$ is less than one thousand of the original signal, it is assumed that an accurate weight vector has been established. The PLS approach satisfies the following termination criteria.

$$|d^t| < 0.001 \times |v| \tag{4}$$

In some situations, both approaches overestimate the baseline when a matrix is present. Baek et al. created the asymmetrically reweighted penalizes least squares (arPLS) method as a solution [124]. MairPLS is another technique built on the similar concepts. While comparing to the prior technique, Long Chen et al. [125] collaborative PLS for Raman spectra background correction result was better.

Multivariate curve resolution-alternating least squares (MCR-ALS)

From MCR-ALS, estimations of the chemically significant profiles of the relevant chemical species may be created from mixed experimental data using a bilinear decomposition [126]. Building many MCR-ALS models while investigating suitable quality-of-fit and interpretability of resolved chemical information is commonly required by strategies to determine the optimal number of components in the MCR-ALS model [127]. The data set include complex, heterogeneous samples of unknown composition, spatially resolved chemical images and associated resolved analyte spectra of the individual, pure chemical components. MCR-ALS specifically breaks down an experimental data matrix (DM) [128]

$$DM = CS^T + E \tag{5}$$

where in Eq. (5), the resolved spectrum matrix is ST, the residual error matrix is E, and the concentration profile matrix is C. Three-dimensional experimental data produced by spectroscopic techniques contain spectral (λ or v) and spatial (x and y) information. The 2D experimental data matrix, DM, which contains integrated spatial (both x and y together) and spectral (λ or v) information, is generated from the three-dimensional experimental data before MCR-ALS. This approach applied for the baseline correction and quantitative purpose also for correction of local minimum of the least square errors obtained by various other methods such as singular value decomposition (SVD) or PCA [129].

Principal component analysis (PCA)

The principal component analysis is a popular unsupervised learning technique for reducing the dimensionality of data. The PCA was invented in 1901 by Pearson [130]. In chromatography, PCA is frequently used to examine the outcomes of complicated samples [131] where uncorrelated variables are linearly fit across the data set. The major variation of data is represented by the first component, which also describes the second-most frequent variance in the data, and so on. This chemometrics tool can be particularly helpful when it comes to interpreting highly dimensional data.

The PCA method may be used for interference factor removal, interference factor extraction, and data compression. The following equation illustrates the outcomes of using the singular value decomposition (SVD) method to carry out PCA analysis and get orthogonal principal components (PCs) [132].

$$D = U\Sigma V^T \tag{6}$$

where in Eq. (6), the three matrices U, Σ, and V^T denote scores, singular values, and loadings with sizes of $m \times m$, $m \times n$, and $n \times n$, respectively. D stands for the raw data with a size of $m \times n$ for decomposition [133].

In chromatography, Soares et al. [134] applied the PCA in combination with COW; its interesting use is to compare columns. Prior performing PCA, the chromatograms are first aligned with a COW technique to increase the probability (p-) values. It is possible to determine if there are significant differences between chromatograms by computing the Mahalanobis distances and converting them to p-values. Although this method decreases noise and raises the signal-to-noise ratio (S/N), there is a

possibility that numerous components may become convoluted and that chemical information is lost. According to a report, the ideal bin size depends on the sample [135]. This method may be used to classify samples in complicated or multidimensional data set.

Parallel factor analysis (PARAFAC)

PARAFAC reduces the dimensionality of the data collection, but factor analysis is as similar to PCA. Factor analysis present the data as trilinear and contains three modes, namely spectra, chromatograms, and concentrations [136], whereas PCA is essentially a dimension reduction approach. As a result, it discovers not only a subspace, but also the vector orientations [137]. PARAFAC2, which was developed by Khakimov et al. [138], can similarly handle slight changes in retention time. The three-way array X of dimensions I, J, and K can be described by the PARAFAC decomposition.

$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{7}$$

In Eq. (7), F stands for the number of factors, while X_{ijk}, a_{if}, b_{jf}, c_{kf}, and e_{ijk} are, respectively, elements of X, A, B, C, and E. The loading matrices A, B, and C have dimensions of $I \times F$, $J \times F$, and $K \times F$, respectively. The three-way array of dimensions $I \times J \times K$ is denoted E [139].

The uniqueness of PARAFAC model is that it establishes not only the subspace, but also the location of the axes defining it. Additionally, the PARAFAC model offers a second-order benefit of allowing for the analysis of chemical components even in the presence of unidentified interferences [140]. Tatjana et al. and Na Peng et al. [141] both applied the PARAFAC to the fluorescence analysis, and they discovered that the model of fluorescence had the capacity to quantify and analyse fluorophores quality in analytes and classify the various types of fluorophores. Another study recommended the combination of PARAFAC with fluorescence regional integration for better characterizing analyte and understanding their functionality [142].

Partial least squares (PLS)-based methods

PLS-DA, also known as discriminant partial least squares (D-PLS), is a method for analysing partial least squares. The technique was first developed by Barker and Rayens [143]. Dimension reduction and the construction of a predictive model are the two major components of PLS-DA modelling. It gives a linear delimiter using partial least squares (PLS) regression with the response variables being binary class membership indices (e.g. 0 and 1) for each class. The PLS-2 algorithm, which enables the prediction of a matrix of response variables in multiple components, is used when there are more than two classes involved.

PLS-DA—The components must be orthogonal to one another in the ordinary variant. The non-singular eigenvectors of the covariance matrix C can be used to formulate it [144].

$$C = \frac{1}{(n-1)^2} X^T C_n y y^T C_n X \tag{8}$$

where in Eq. (8), y is the class label vector, C_n is the $n \times n$ centring matrix, and X is the loading matrix. The loading vectors $a_1, \ldots a_d$, which denote the relevance of each feature in that component, are computed iteratively. Its objective for iteration h is as follows:

$$\max \mathrm{cov}(X_h a_h, y_h b_h) \tag{9}$$

where $X1 = X$, y_h and X_h are the residual (error) matrices following transformation with the prior h-1 components, and b_h is the loading for the label vector y_h.

PLS-DA has been used mostly in biomarker and drug discovery research using LC–MS/MS and NMR study of advanced-stage melanoma in blood [145]. Using LC–MS data, Lambrecht et al. [146] employed PLS-DA to classify black rice according to its place of origin. PLS was used by Eleni et al. [147] to predict the diffusion of substances in artificial membranes.

Additionally, orthogonal partial least squares discriminant analysis (OPLS-DA) is designed to distinguish between the discriminating and non-discriminatory dimensions [148]. Using a set of metabolites identified by LC–MS/MS, Zhang et al. [149] applied OPLS-DA to confirm the legitimacy of fruit juices. Shurui et al. [150] used a similar strategy when they used OPLS-DA to HRMS study for non-target metabolomics.

Support vector machines (SVM)

A set of pattern-recognition techniques called support vector machines (SVM) was developed to effectively handle nonlinear data distributions. It is one of the chemometrics' machine learning methods. The fundamental component of SVM is the projection of data points into a space with added dimensions, which serves as a means of identifying linear functions capable of modelling the data [151]. Such modelling functions can be projected back into the space of the original predictors, and producing functions are higher in complexity but lower in dimension (often nonlinear). The use of SVM in discriminant classification is conventional. Nevertheless, several authors offered class-modelling-relevant adjustments. It is important to note the support vector domain description (SVDD) method by Songfeng Zheng [152] used hyperspheres to describe

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 15 of 42

the class spaces, as one of the most popular strategies. Numerous researchers have used this strategy in a variety of analytical studies, including laser-induced breakdown spectroscopy [153], ATR-FT-IR spectroscopy [154], tandem mass spectrometry (MS/MS) [155], and HPLC [156].

Artificial neural networks (ANNs)

ANNs are multilayer networks of linked mathematical operators (neurons). The feed-forward neural network is the most common ANN. Here, each neuron performs as a weighted sum of the input data or outputs of the preceding layer as modified by an activation function (typically linear or logistic function). The proposed algorithms learns from a dataset for predicting event outcomes [157].

In the last decade, artificial neural networks (ANNs) have been developed to determine retention index or time for 1D-GC, 1D-LC, 2D-LC and 2D-GC separations [158, 159]. ANNs are computer programs that "learn" to carry out tasks by taking into account multiple cases. As long as enough input is given, an ANN can detect traits and patterns in data. Then, predictions are made in novel conditions using these traits and patterns. ANNs have been employed in variety of analytical research studies like LC–MS/MS determination [160], GC–MS [161], and HPLC [162]. Moreover, the list of chemometric methods used in analytical techniques were listed in Table 2.

Table 2 List of chemometric methods used in analytical techniques

Analytical technique	Chemometrics method	References
UV–visible spectroscopy	PCR—principal component regression	[163, 164]
	TLRC—trilinear regression calibration	[164]
	MLRC—multi-linear regression calibration	[164]
	PLSR—partial least square regression	[165]
	ILS—inverse least square	[166]
LC-NMR	PCA—principal component analysis	[167]
	OPA—orthogonal projection analysis	[168]
	EFA—evolving factor analysis	[168]
	MCR-ALS—multivariate curve resolution by alternating the least squares approach	[168] [168]
	SFA—sub-window factor analysis	[168]
LC–MS	MCR-ALS—multivariate curve resolution by alternating the least squares approach	[169, 170]
	PARAFAC—parallel factor analysis	[171]
GC–MS	PCA—principal component analysis	[172]
	GRAM—generalized rank annihilation	[173]
FT-IR	PLS-DA—partial least square discriminate analysis	[174]
	PCA—principal component analysis	[175]
HPLC	PCA—principal component analysis	[176]
	GRAM—generalized rank annihilation	[177]
	PARAFAC—parallel factor analysis	[177]
	HCA—hierarchical cluster analysis	[178]
	SIMCA—soft independent modelling of class analogy	[179]
	RAFA—rank annihilation factor analysis	[180]
UPLC	PCR—principal component regression	[181]
	HCA—hierarchical cluster analysis	
Raman spectroscopy	PCA—principal component analysis	[182]
	SIMCA—soft independent modelling of class analogy	
NIR (near-infrared)	PLS—partial least square	[183]
	PCA—principal component analysis	[183]
	SIMCA—soft independent modelling of class analogy	[184]
Spectrofluorometry	PLS—partial least square PARAFAC—parallel factor analysis	[185]

LC–NMR Liquid chromatography nuclear magnetic resonance, *LC–MS* Liquid chromatography–mass spectrometry, *GC–MS* Gas chromatography–mass spectrometry, *FT-IR* Fourier transform infrared, *HPLC* High-performance liquid chromatography, *UPLC* Ultra-performance liquid chromatography

Table 3 Data on the analytical methods reported based on AQbD

S. No	Drug	DoE tool	MODR	tR	Column	Mobile phase	% RSD (precision range)	% Recovery (accuracy range)	References
1	Venetoclax	Full factorial design (2 factors)	FR: 0.3–0.4 CT: 40–55 °C pH: 6–8 ACN%:±3.8%	7.39	C18 (150 mm, ×2.1 mm, 1.7 µ)	5:95% (ammonium bicarbonate in acetic acid and ACN), pH:7.0	0.71–1.11	99.09–100.68%	[190]
2	Valsartan	Box–Behnken design (3 factors)	FR:1.0 CT: 30–40 °C pH: 3.0–3.5	6.70	ODS C18 column (4.6×250 mm, 5 µ)	65:35% (ACN and 0.01 M acetic acid), pH: 3.2	0.068–0.943	99.7–100.9%	[191]
3	Canagliflozin	Factorial design (3 factors)	FR: 1.5 CT: 30–35 °C pH: 3.0–3.5	3.72	C18, (250 mm×4.6 mm, 5 µ)	50:50% (ACN and 0.1% phosphoric acid); pH: 3.0	0.413–0.440	99.08–99.52%	[192]
4	Etofenamate	Central composite design (3 factors)	FR: 0.8–1.2 pH: 5–7 AQ: 5–25	5.30	C18 (250×4.6 mm, 5 µ)	85:15% v/v (methanol and 0.2% TEA in water) pH: 3.0	0.63–0.87	99.6–101.3%	[193]
5	Olanzapine	Rechtschaffen design (4 factors)	FR: 1.5 CT: 40–45 °C pH: 3.0 ACN:±5.3	12.0	Diol (100×4.6 mm, 5 µ)	10:90% v/v (ammonium formate 20 mmol L⁻¹ in 10% ACN and Pure ACN) pH: 3.0	Olz: 1.78 Impurity: 5.9–20.6%	Olz: 100.2% Impurity 1 to 7:107.3–117.9%	[194]
6	Acyclovir and Hydrocortisone	Factorial design (3 factors)	FR: 0.6–1.4 pH: 3.0–7.0 AQ%:30–70	ACY: 3.49 HYD: 9.18	C18, (250×4.6 mm, 5 µ)	50:50% (methanol and water); pH: 6.0	0.63–1.01	98.79–101.41%	[195]
7	Etizolam and propranolol hydrochloride	Central composite design (3 factors)	FR: 0.9–1.1 pH: 3–4 AQ%:30–35	PRO: 3.6 ETZ: 8.2	C18, (250×4.6 mm, 5 µ)	30:70%v/v (0.15% TEA and methanol) pH: 3.5	PRO: 1.02–1.3 ETZ: 0.94–1.4	PRO: 98.8–101.3% ETZ: 98.9–100.5%	[196]
8	Ceftriaxone Sodium	Two-factor and three-level factorial	FR: 1.0 CT:25 °C pH: 5.5–7.5 ACN%: 65–75 AQ%: 25–35	4.15	ODS C18 (250×4.6 mm, 5.0 µ)	70:30% v/v (ACN and water (0.01% TEA) pH:6.5	Interday: 0.55–0.95 Intraday: 0.70–0.94	99.70–101.19%	[15]
9	Telmisartan Hydrochlorothiazide	Three-level factorial design	FR: 0.8–1.2 CT: 30–50 °C pH: 2.5–3.5 ACN%: 65–75 AQ%: 25–35	TMS: 31.18 HCZ: 9.11	ODS, (150×4.6 mm, 3.5 µ)	85:15% v/v (20 mM PDOP and water, ACN in the ratio of 100:900) pH: 3.5	< 2%	105.9–109.4%	[197]
10	Eberconazole nitrate	Full factorial design (3 factors)	FR:1.0–1.2 TBAH: 5–10 mM pH: 2.6–3.2 The organic phase (%v/v): 20–30	7.05	C18 (250×4.6 mm, 5 µ)	25:75%, v/v (methanol–PDOP, 10 mM and TBAH10 mM) pH: 2.8	Interday: 1.66–1.72 Intraday: 1.13–1.16	100.02–100.05%	[198]

Table 3 (continued)

S. No	Drug	DoE tool	MODR	tR	Column	Mobile phase	% RSD (precision range)	% Recovery (accuracy range)	References
11	Cyanidin-3-O-glu-coside	Box–Behnken design	FR: 0.8–1.0 CT: 35–45 °C λ: 510–530 nm Mobile phase A:15–19%; B:81–85%	14.36	C18 (250×4.6 mm, 5 μ)	15:85% v/v (Metha-nol, acetonitrile in the ratio 3:1 and 5% formic acid)	Interday: 0.02–0.03 Intraday: 0.01–0.02	97–102%	[199]
12	Quercetin Dihydrate	Fractional factorial design	FR: 0.7–1 Inj. vol: 10–20 CT: 25—40 Buffer: 30–50 mM	3.14	C18 (250×4.6 mm, 5 μ)	35:65% v/v (ACN and Ammonium acetate buffer (0.1% v/v acetic acid)) pH: 3.5	Interday: 0.030–0.093 Intraday: 0.039–0.068	99.8–100.4%	[200]
13	Atorvastatin	Box–Behnken statisti-cal design	FR: 0.35–1.0 ACN%: 50–90 λ: 235–240 nm	2.43	C18 (250×4.6 mm, 2.2 μ)	50:50% v/v (ACN and Water) pH: 3.5	Interday: 1.25–2.71 Intraday: 1.79–2.70	98.66–99.84%	[201]

FR Flow rate, *CT* Column temperature, *AQ* Aqueous, *ACN* Acetonitrile, *RSD* Relative standard deviation, *TBAH* Tetrabutylammonium hydroxide, *PDOP* Potassium dihydrogen orthophosphate, *TEA* Triethylamine
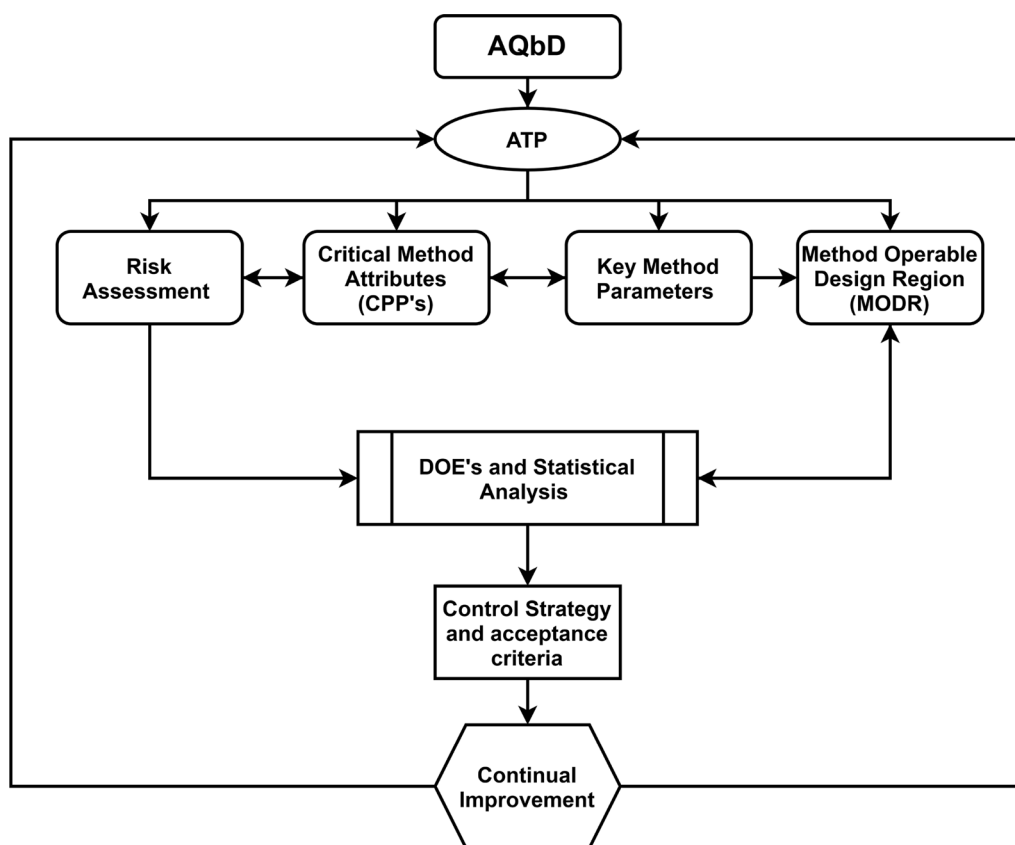
Fig. 6 A typical AQbD approach in analytical method development

Analytical quality by design (AQbD)

Analytical quality by design (AQbD) is an approach for developing robust analytics that is appropriate for regulatory flexibility in pharmaceutical submissions to the FDA. AQbD is widely used in the development of various analytical methods such as UV–visible, FT-IR, Raman, NIR, fluorimetric, HPLC, UHPLC, LC–MS, GC–MS, HPTLC, and SFC. In the pharmaceutical industry, the AQbD tool is integrated with PAT as a real-time process analyser to monitor any given process or material, which generates massive and complex data sets. There is a growing interest in the implementation of AQbD in new analytical method development procedures for wider applications including assays, stability studies, and bioanalytical studies, in analytical method development. While comparing to one-factor-at-a-time (OFAT) approach, AQbD-based analytical methods have demonstrated a high degree of robustness and method performance. Notably, using these techniques reduces the likelihood of human error, and the AQbD approach will not predict any chromatogram but instead explore scientific understanding in method implementation sequences, beginning with the quality of predictions that relate to risk assessment in method choice, then between method parameter and

expected method results, and finally a region for a highly robust and cost-effective approach [186]. The design of experiment (DoE) is a part of AQbD methodology and represents the interaction among the input factors that ultimately affect the technique response and outcomes. Therefore, a typical AQbD methodology starts with an analytical target profile (ATP) and risk and critical evaluation, then uses DoE to optimize the method variables, creates a method operable design region (MODR), and implements a control plan [187–189]. There are works available and comprised in Table 3 and the scheme of methodology illustrated in Fig. 6.

Assessments of prediction ability of prediction software

Assessment of the predictive ability of NMR prediction by Chemaxon

An attempt was made to verify the expected chemical shift values for the chosen test compounds shown in Fig. 7. The original experimental chemical shift values were compared to the predicted chemical shift values of ten chemically divergent structural compounds in this experiment. A per cent error (%) for each chemical shift value was obtained, as well as regression analysis. The per cent error ranged from -26.52 to 35.98%. The
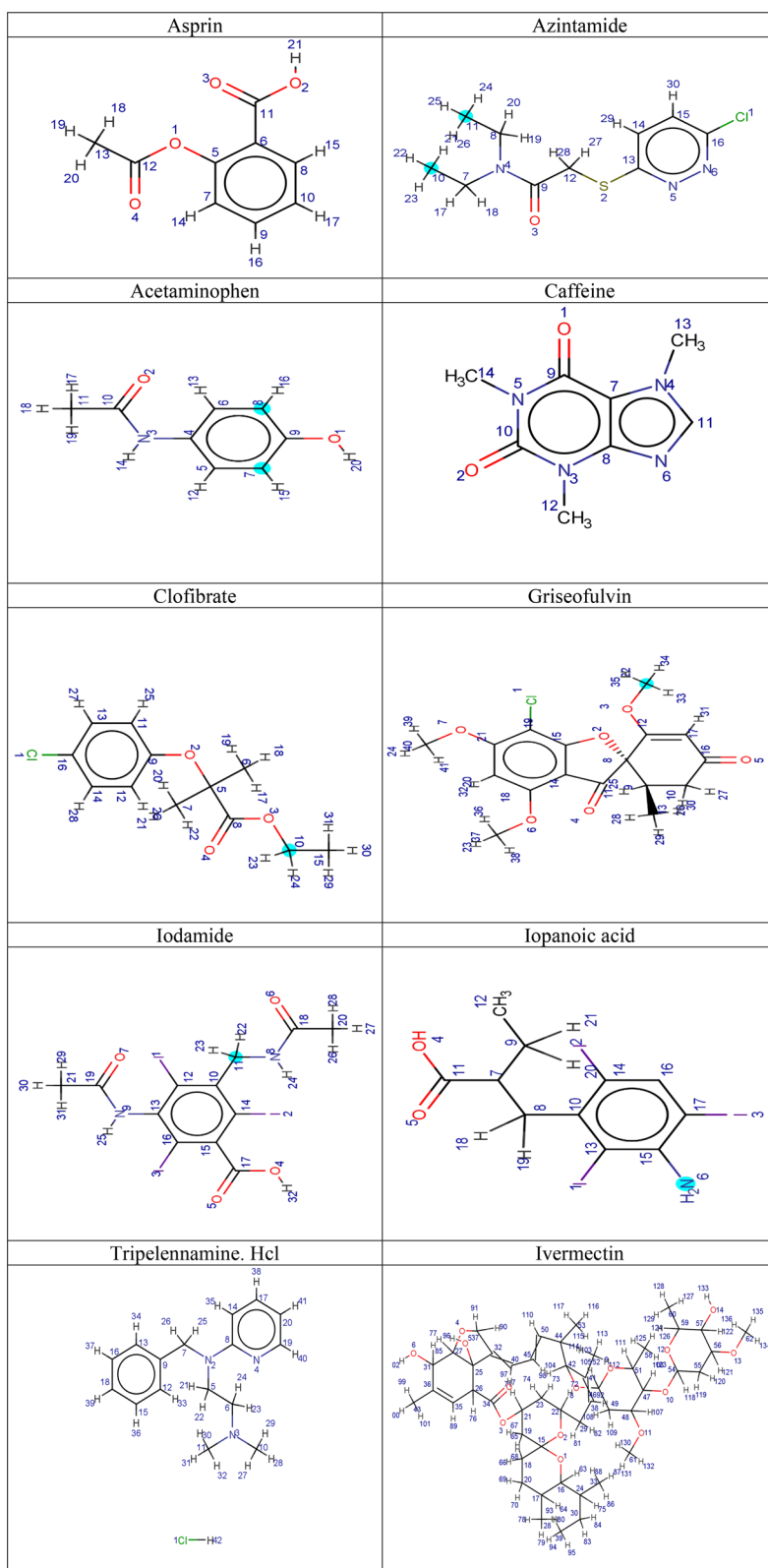
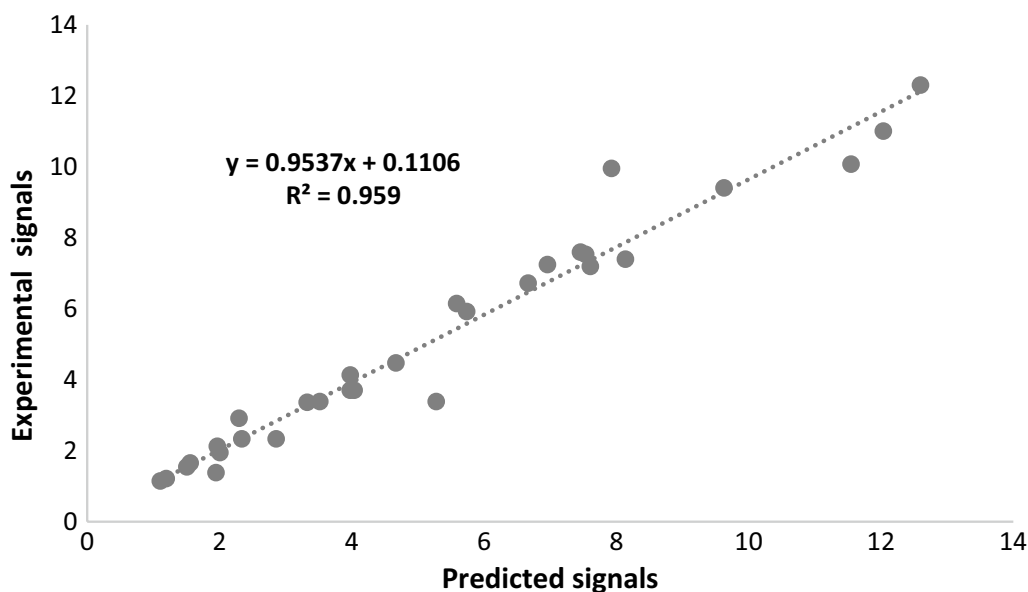**Fig. 7** Chemical structures used for NMR signal prediction assessment

Fig. 8 Regression for H^1-NMR predicted *versus* experimental signals
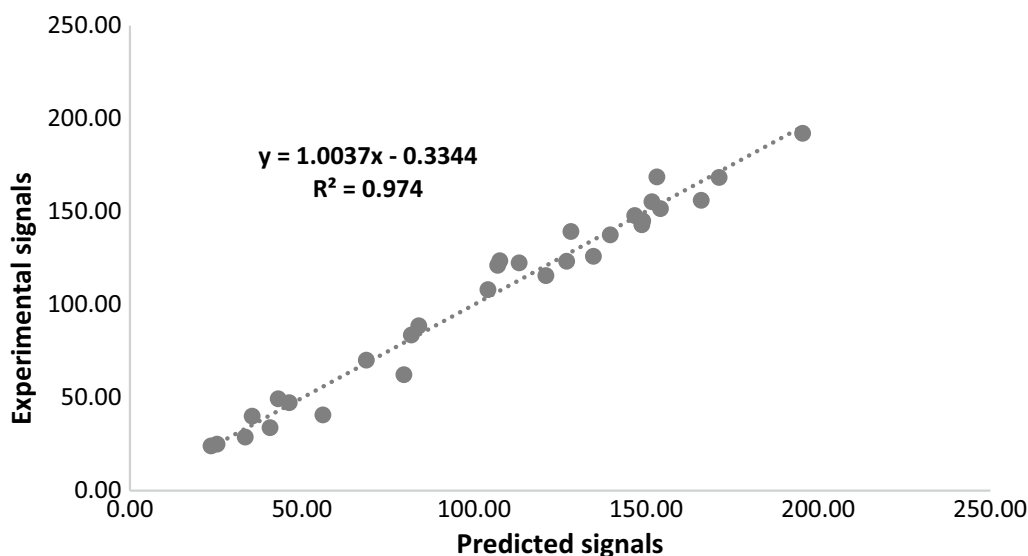

Fig. 9 Regression for C^{13}-NMR predicted *versus* experimental signals

correlation's graphs in Figs. 8 and 9 show R^2 value of 0.959 (H^1-NMR) and 0.974 (C^{13} NMR). This indicates the accuracy of NMR signal prediction. According to prediction results, in H^1-NMR, aliphatic proton error ranged from -26.52 to 35.98%, whereas aromatic proton error ranged from -25.47 to 9.21%. The aliphatic carbon error ranged from -14.41 to 27.54% in the C^{13} NMR, whereas the aromatic carbon error ranged from -14.95 to 6.49 per cent. Finally, we conclude the aliphatic error was greater when compared with the aromatic error all those data was presented in the Table 4.

Assessment of the predictive ability of ORCA
For UV–Visible prediction
Here, originally obtained wavelength maximum (λmax) values were compared to the predicted wavelength values of fifteen structurally divergent structural compounds. A per cent error (%) for each wavelength value was obtained, as well as regression analysis. The error rate was found to be between -2.27 and 18.69%. The correlation's graph in Fig. 10 *shows* R^2 value of 0.926. This demonstrates the accuracy of UV–visible prediction. The results demonstrate that when methanol is used as a prediction

Table 4 Comparative data for H^1-NMR and C^{13} NMR signals of predicted *versus* experimental signals

Compound	1H NMR signals							13C NMR signals							References
	MHz	A.no.	M.t	Std.	Pred.	Dev.	% error	MHz	A.no.	M.t	Std.	Pred.	Dev.	% error	
Ivermectin	300.0	82	m	5.74	5.92	−0.18	−3.14	250.0	50	s	139.60	137.48	2.12	1.52	[202]
		75	d	3.98	3.7	0.28	7.04		47	s	81.80	83.60	−1.80	−2.20	
		52	s	1.51	1.54	−0.03	−1.99		37	s	68.70	70.08	−1.38	−2.01	
Azintamide	200.0	15	d	7.46	7.59	−0.13	−1.74	200.0	13	s	166.01	156.07	9.94	5.99	[203]
		7	q	3.52	3.38	0.14	3.98		14	s	128.16	139.27	−11.11	−8.67	
		10	t	1.11	1.14	−0.03	−2.70		12	s	40.76	33.80	6.96	17.08	
Caffeine	60.0	11	s	7.54	7.53	0.01	0.13	60.0	9	s	151.69	155.31	−3.62	−2.39	[204]
		13	s	3.98	4.13	−0.15	−3.77		11	s	148.73	142.95	5.78	3.89	
		12	s	3.33	3.36	−0.03	−0.90		12	s	33.54	28.74	4.80	14.31	
Aspirin	100.0	2	s	12.04	11.00	1.04	8.64	200.0	12	s	171.20	168.35	2.85	1.66	[205]
		10	sx	7.61	7.19	0.42	5.52		10	s	134.70	125.96	8.74	6.49	
		13	s	2.34	2.33	0.01	0.43		7	s	126.90	123.27	3.63	2.86	
Griseofulvin	100.0	20	s	5.59	6.14	−0.55	−9.84	60.0	11	s	195.45	192.10	3.35	1.71	[206]
		22	s	4.04	3.70	0.34	8.42		19	s	104.04	108.02	−3.98	−3.83	
		26	m	2.30	2.91	−0.61	−26.52		10	s	35.52	40.00	−4.48	−12.61	
Iodamide	100.0	9	s	7.93	9.95	−2.02	−25.47	60.0	15	s	149.00	145.03	3.97	2.66	[207]
		11	s	4.67	4.47	0.2	4.28		10	s	107.50	123.57	−16.07	−14.95	
		20	s	2.01	1.94	0.07	3.48		11	s	56.10	40.65	15.45	27.54	
Acetaminophen	500.0	6	s	9.63	9.40	0.23	2.39	60.0	6	s	153.15	168.66	−15.51	−10.13	[208, 209]
		8	d	6.67	6.72	−0.05	−0.75		8	s	120.91	115.60	5.31	4.39	
		10	s	1.97	2.12	−0.15	−7.61		10	s	23.61	23.97	−0.36	−1.52	
Clofibrate	60.0	14	d	6.96	7.24	−0.28	−4.02	100.0	9	s	154.18	151.59	2.59	1.68	[210]
		7	s	1.56	1.65	−0.09	−5.77		10	s	79.62	62.29	17.33	21.77	
		15	t	1.20	1.21	−0.01	−0.83		6	s	25.34	24.90	0.44	1.74	
Tripelennamine Hcl	100.0	1	s	11.55	10.07	1.48	12.81	60.0	14	s	113.10	122.42	−9.32	−8.24	[211]
		17	sx	8.14	7.39	0.75	9.21		20	s	106.90	121.05	−14.15	−13.24	
		11	s	2.86	2.33	0.53	18.53		5	s	43.10	49.31	−6.21	−14.41	
Iopanoic acid	100.0	4	s	12.60	12.30	0.3	2.38	100.0	15	s	146.70	147.88	−1.18	−0.80	[212]
		6	s	5.28	3.38	1.9	35.98		14	s	83.95	88.58	−4.63	−5.52	
		20	o	1.95	1.38	0.57	29.23		8	s	46.30	47.23	−0.93	−2.01	

MHz: Megahertz, *A.no* Atom number, *M.t* Multiplet information, *Std* Standard, *Pred* Predicted, *Dev* Deviation, *% Error*: Percentage error

solvent, the error ranges from 0.0 to 18.69 per cent, whereas water has a range of −2.27 to 11.73%. Finally, we conclude that more error is observed when using methanol as a solvent for prediction compared with water. The resulting data were presented in the Table 5.

For Raman and infrared
The predicted Raman shift and infrared absorption frequency for the selected test substances was verified with experimental values. Here, the predicted frequency values of ten chemically divergent structural compounds are verified with original experimental frequency values. The % error for each frequency value and regression analysis was calculated. The % error was observed between −30.04 and 29.26%. The R^2 value (Figs. 11 and 12) for the correlation was 0.946 for both the Raman shift and the infrared absorption frequency. This indicates the reliability of Raman shift and infrared absorption frequency prediction. The results reveals that aliphatic single bond compound error ranged from −20.02 to 29.26%, double and triple bonded compound error ranged from −4.20 to 13.14%, and hydroxyl function group compound error ranged from −21.01 to 3.55%, and aromatic ring compound error ranged from −11.43 to 18.07%. As a result, we find that the aliphatic single bond compound error was greater than other errors, moreover the comparative data was presented in the Table 6 clearly.
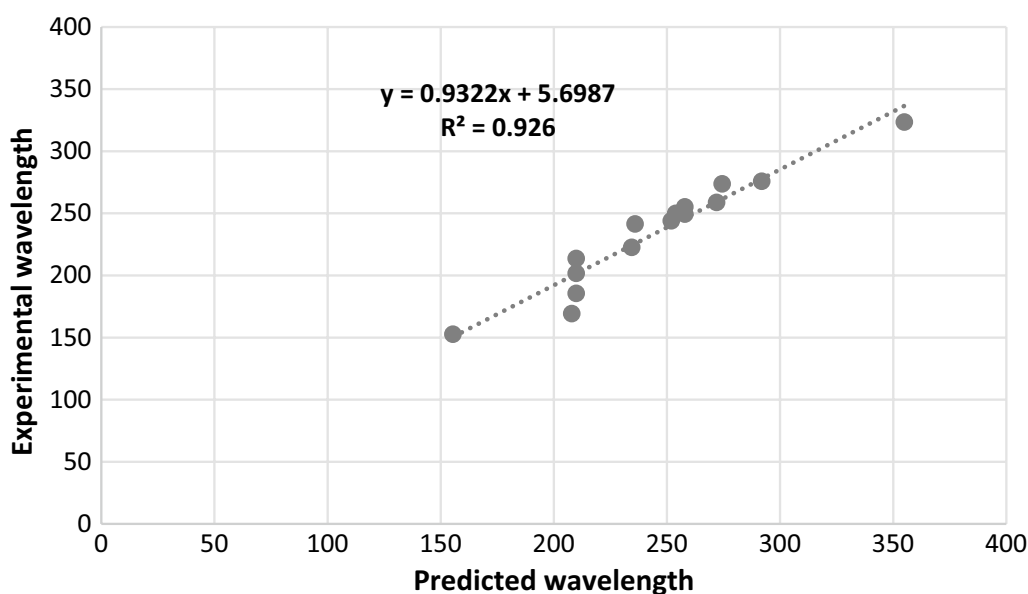
Fig. 10 Regression plot for UV–visible of predicted versus experimental frequency

Table 5 Comparative data of UV–visible spectral data: predicted versus experimental

Compound	UV–visible absorption (wavelength)					References
	Solvent	Exp. (nm)	Pred. (nm)	Dev.	%Error	
Caffeine	Methanol	272	258.72	13.28	4.88	[204]
Malic acid	Water	210	213.51	− 3.51	− 1.67	[213, 214]
Primidone	Methanol	258	249.18	8.82	3.42	[214]
Acetaminophen	Water	254	249.85	4.15	1.63	[215]
Benzocaine	Methanol	292	275.65	16.35	5.60	[216]
Pyrrole	Water	210	185.36	24.64	11.73	[217]
Anthracene	water	355	323.37	31.63	8.91	[218]
Metformin	Water	236	241.35	− 5.35	− 2.27	[219]
Chlorobutanol	Water	210	201.6	8.4	4.00	[220]
Acetic acid	Methanol	208	169.12	38.88	18.69	[221]
Aniline	Methanol	234.5	222.52	11.98	5.11	[221]
Coumarin	Methanol	274.5	273.66	0.84	0.31	[221]
Allopurinol	Methanol	252	243.91	8.09	3.21	[221]
Methyl paraben	Methanol	258	255.06	2.94	1.14	[221]
Formaldehyde	Methanol	155.5	152.49	3.01	1.94	[221]

Furthermore, an attempt has been made to verify the predicted infrared absorption frequency for lamivudine and zidovudine with all functional frequencies. In this study, the predicted frequency values of lamivudine and zidovudine structural functional group frequencies were compared to the original experimental frequency values. The % error for each frequency value and regression analysis was calculated. The % error was observed between − 24.26 and 18.89%. The R^2 value for the correlation graph in Fig. 13 was shown to be 0.970 for both lamivudine and zidovudine absorption frequencies. This also demonstrates the reliability of frequency prediction using a single compound prediction with all functional groups was presented in the Table 7.
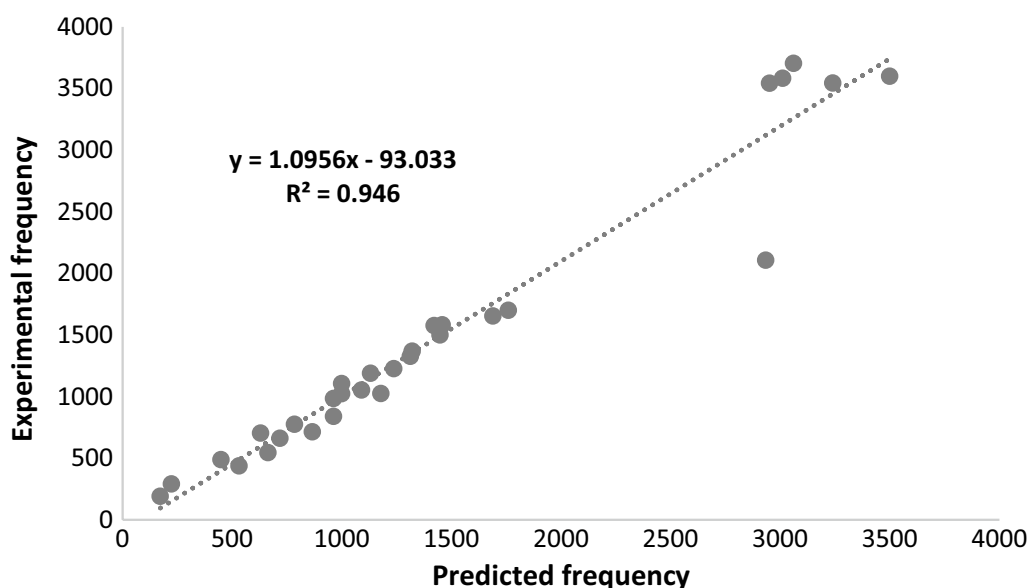
Fig. 11 Regression plot for Raman shift of predicted versus experimental frequency
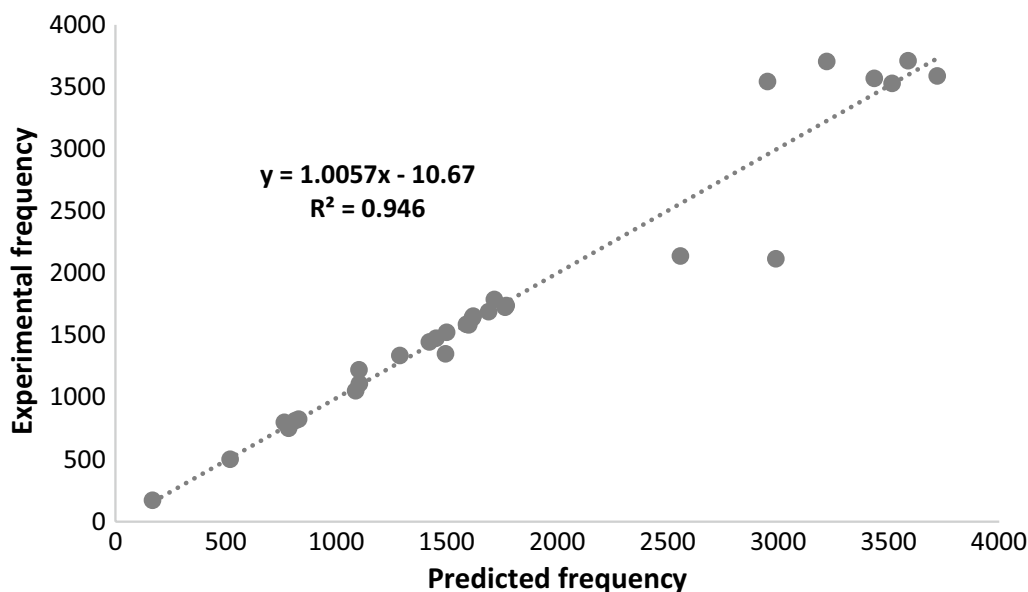

Fig. 12 Regression plot for infrared absorption of predicted versus experimental frequency

Assessment of predictive ability QSRR Automator with reference data

The QSSR retention predictions for antiviral drugs were conducted. For that reference information was gathered from various research publications, and different antiviral drugs with C_{18} column elution were selected. The predicted retention time for the test set of drugs was compared to the published retention time data. The % error for each retention time and regression coefficient were calculated. The % error was observed in the range of -20 to 20%. This can be observed clearly in the histogram plot in Fig. 14 and 15, and the R^2 value for the correlation was 0.947. This indicates the reliability of retention time prediction. Topological and 2D descriptors like MW, AATS, MATS, GATS, Axp, n6aHRing, NsNH2, and SLogP are the most often used contributing descriptors, and they depend on the analyte chemical structure. Table 8 was presented the results of QSRR predicted and experimental retention time of Antiviral class drugs.

Table 6 Comparative data on Raman shift and infrared absorption frequency: predicted versus experimental values

Compound	Raman frq. (cm^{-1})				Infrared frq. (cm^{-1})				References
	Std.	Pred.	Dev.	%Error	Std.	Pred.	Dev.	%Error	
Malic acid	1422	1575	−153	−10.76	1716	1788	−72	−4.20	[215]
	963	984	−21	−2.18	1288	1336	−48	−3.73	
	532	436	96	18.05	1103	1222	−119	−10.79	
Primidone	3012	3582	−570	−18.92	2990	2115	875	29.26	[209, 214]
	1322	1367	−45	−3.40	1620	1653	−33	−2.04	
	963	838	125	12.98	765	799	−34	−4.44	
Acetaminophen	2934	2106	828	28.22	3588	3710	−122	−3.40	[209, 215]
	1449	1499	−50	−3.45	3435	3568	−133	−3.87	
	664	544	120	18.07	1690	1689	1	0.06	
Chloral hydrate	1760	1699	61	3.47	3720	3588	132	3.55	[222]
	223	290	−67	−30.04	1765	1725	40	2.27	
	172	190	−18	−10.47	169	171	−2	−1.18	
Halothane	1313	1326	−13	−0.99	1600	1584	16	1.00	[223]
	1179	1024	155	13.15	814	812	2	0.25	
	718	661	57	7.94	520	502	18	3.46	
Hydralazine hydrochloride	3061	3704	−643	−21.01	3220	3704	−484	−15.03	[215, 224]
	1238	1226	12	0.97	1590	1587	3	0.19	
	449	487	−38	−8.46	785	751	34	4.33	
Lithium carbonate	1459	1582	−123	−8.43	2558	2137	421	16.46	[225]
	1091	1053	38	3.48	1495	1350	145	9.70	
	866	713	153	17.67	1088	1053	35	3.22	
Phenobarbital	1690	1653	37	2.19	1770	1739	31	1.75	[226]
	1000	1104	−104	−10.40	1500	1523	−23	−1.53	
	630	702	−72	−11.43	830	824	6	0.72	
Sorbitol	3240	3543	−303	−9.35	2952	3543	−591	−20.02	[227]
	2952	3542	−590	−19.99	1422	1446	−24	−1.69	
	1132	1188	−56	−4.95	1105	1109	−4	−0.36	
Trimethoprim	3500	3599	−99	−2.83	3516	3528	−12	−0.34	[228]
	1000	1022	−22	−2.20	1616	1637	−21	−1.30	
	785	775	10	1.27	1452	1475	−23	−1.58	

Short-time Fourier transform (STFT) method for assessment

The short-time Fourier transform (STFT) study revealed sufficiency to recognize that all the assessment results of C^{13}, H^1-NMR, UV–visible, IR, Raman are accurate and have also been used for further evaluation purpose. Based on the density power of frequency spectrogram, it is most likely that the yellow or red colour denoted high power, and the blue colour is low power. If the spectrograms had the same frequency power or should have produced results close to the acceptable prediction, we would have concentrated on the greater frequency power of both the predicted and experimental data sets. While both will have different frequency powers, this indicates that the prediction was inaccurate.

The H^1-NMR power frequency of the predicted and experimental results are nearly identical and the highest power index should be in the range of 18.25–17.39, respectively (Fig. 16). This indicates that both outcomes are accurate. With regard to the C^{13}-NMR, the power frequency of both the predicted and the experimental data are depicted in Fig. 17, both of which exhibit the identical power index value of 44.17 and 44.01, respectively. This demonstrated the validity of the data, and the spectrogram revealed very slight frequency differences in the lower power range, which are visible in the blue colour peaks. The highest frequency indexes in the UV–visible Fig. 18 are 44.89 and 44.82, for both predicted and experimental results, respectively, in the same frequency index.
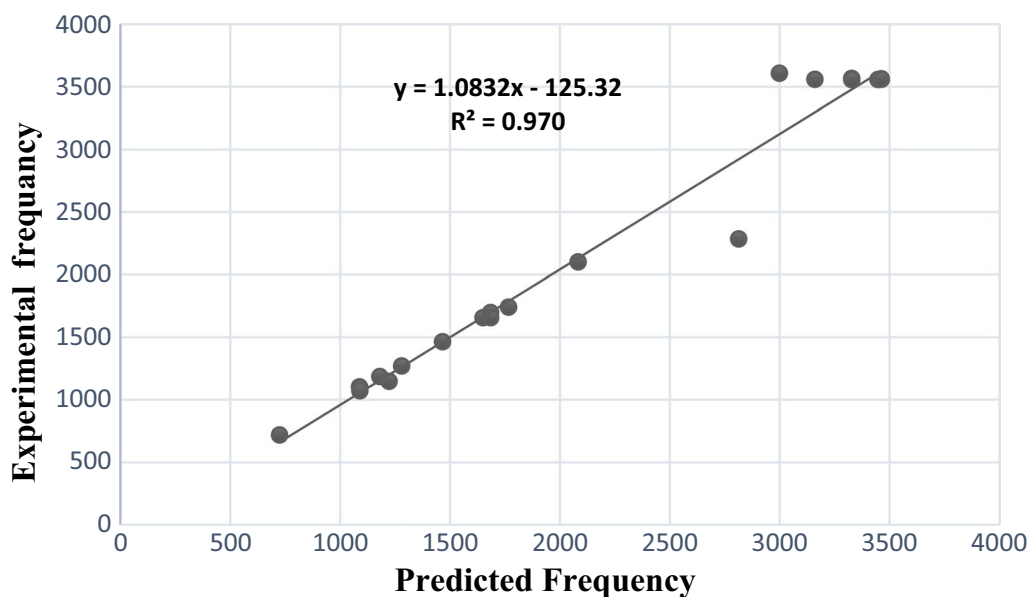
Fig. 13 Regression plot for infrared absorption of predicted versus experimental frequency of lamivudine and zidovudine

Table 7 Comparative infrared data of single compound with functional group frequency: predicted vs experimental values

| Drug	Targeted functional assignments	Frequency range	Published freq (Cm^{-1})	Predicted freq (Cm^{-1})	Intensity	Deviation	%Error	References
Lamivudine	Heterocyclic amine, NH stretch	3490–3430	3328.4	3564.43	S	− 236.03	− 7.09	[229, 230]
	Primary alcohol, OH stretch	3645–3630	2999	3607.1	S	− 608.1	− 20.28	
	C–NH$_2$, Aromatic primary amine, NH stretch	3510–3460	3448	3556.55	S	− 108.55	− 3.15	
	C=O amide	1725–1700	1767	1738.05	S	28.95	1.64	
	C=C stretch	1680–1620	1651	1653.5	S	− 2.5	− 0.15	
	Tertiary amine, CN stretch Aromatic amino	1210–1150	1181.6	1183.21	S	− 1.61	− 0.14	
	C–H stretch (aromatic)	3130–3070	3328.4	3556.55	W	− 228.15	− 6.85	
	Aryl thioethers, C–S stretch	715–670	725.17	714.51	S	10.66	1.47	
	C–O–C, C–O stretch, cyclic ether	1140–1070	1088.8	1099.8	S	− 11	− 1.01	
	C–OH Stretch	1200–1020	1223.1	1144.24	S	78.86	6.45	
	CH3 Stretch, C–H stretch (methyl asym./sym. Stretch)	2970–2950/2880–2860	2854	3546.32	S	− 692.32	− 24.26	
Zidovudine	Heterocyclic amine, NH stretch	3490–3430	3462.9	3557.08	S	− 94.18	− 2.72	[228, 231]
	Primary alcohol, OH stretch	3645–3630	3462.9	3561.75	S	− 98.85	− 2.85	
	C=O amide	1725–1700	1685.8	1695.08	S	− 9.28	− 0.55	
	C–O–C, C–O stretch, cyclic ether	1140–1070	1090.3	1093.72	S	− 3.42	− 0.31	
	Open-chain azo (–N=N–)	1630–1575	2083.3	2098.49	S	− 15.19	− 0.73	
	C=C stretch	1680–1620	1685	1654.46	W	30.54	1.81	
	Tertiary amine, CN stretch Aromatic amino	1210–1150	1280.7	1268.1	S	12.6	0.98	
	C–H stretch (aromatic)	3130–3070	3160.7	3557.08	W	− 396.38	− 12.54	
	Aromatic secondary amine, NH stretch	3450	3462	3561.75	S	− 99.75	− 2.88	
	CH3 Stretch, asym./sym. bend	1470–1430/1380–1370	1466.7	1461.2	M	5.5	0.37	
	CH3 Stretch, C–H stretch (methyl asym./sym. Stretch)	2970–2950/2880–2860	2814.4	2282.89	S	531.51	18.89	
	C–OH Stretch	1200–1020	1090.3	1069.58	S	20.72	1.90	
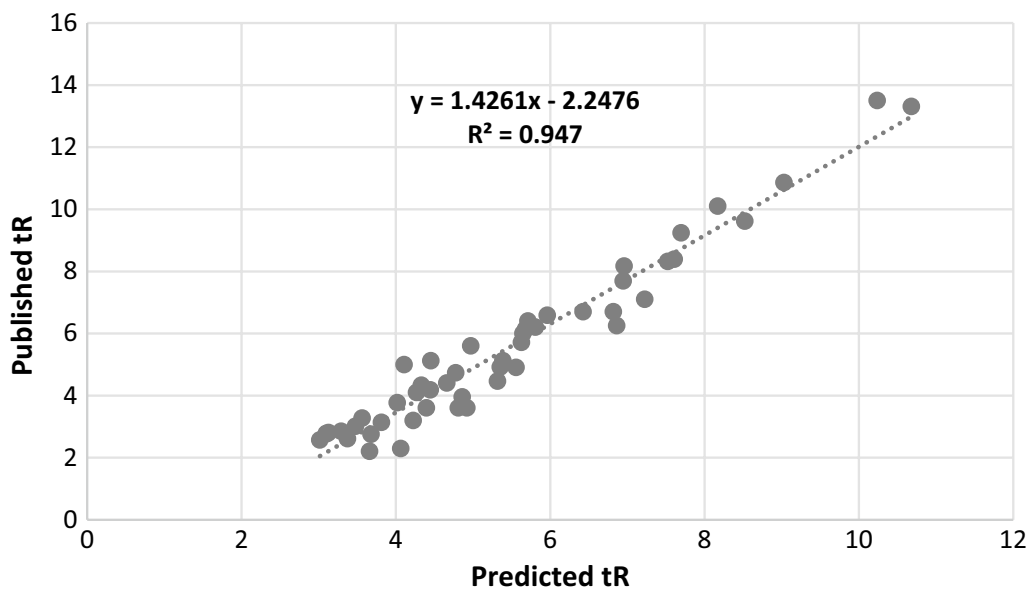
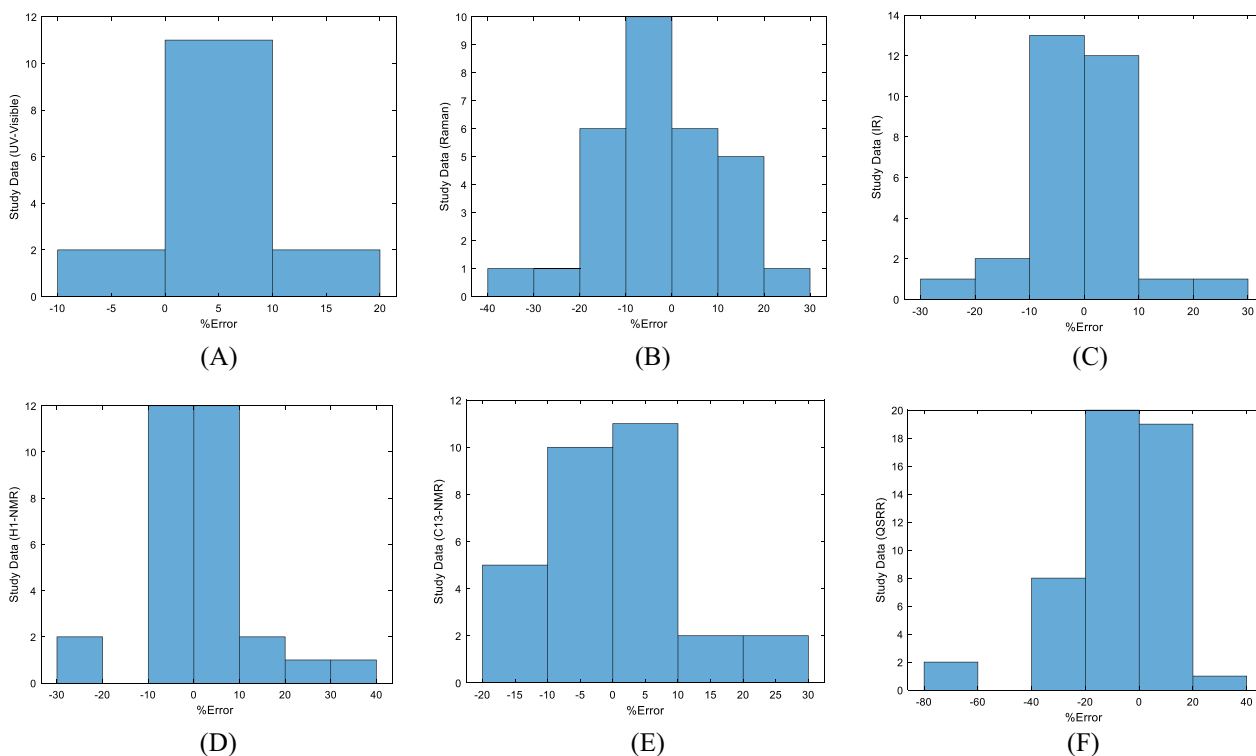**Fig. 14** Regression plot for a retention time of predicted versus published

Fig. 15 Illustrate Histogram plot for all the % errors in prediction study **A** UV–visible—symmetric error, **B** Raman—right skewed error, **C** IR—symmetric error, **D** H1-NMR—right skewed error, **E** C13-NMR—left skewed error, **F** QSRR—left skewed error

Table 8 Comparative data of predicted and experimental retention in HPLC using the QSRR approach [232–279]

| S.no.	Drug	Predicted tR (mins)	Published tR (mins)	% Error	Column	Contributing descriptors
1	Stavudine	3.29238	2.85	− 15.52	C18	MW, AATS, MATS, GATS, Axp, SLogP
2	Lamivudine	4.33172	4.33	− 0.04	C18	MW, AATS, MATS, GATS, Axp, PEOE_VSA5,
3	Nevirapine	7.6097	8.39	9.30	C18	MW, AATS, MATS, GATS, Axp, nFARing
4	Boceprevir	4.81008	3.6	− 33.61	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, nFARing, SRW03
5	Penciclovir	5.80932	6.201	6.32	C18	MW, AATS, MATS, GATS, Axp, n6aHRing
6	Abacavir	5.38772	5.12	− 5.23	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, SRW03
7	Acyclovir	7.69738	9.241	16.70	C18	MW, AATS, MATS, GATS, Axp, NsNH2, SLogP
8	Atazanavir	7.52508	8.32	9.55	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, SLogP
9	Daclatasvir	5.68164	6.138	7.43	C18	MW, AATS, MATS, GATS, Axp, SLogP
10	Didanosine	3.47798	3.007	− 15.66	C18	MW, AATS, MATS, GATS, Axp, SLogP
11	Doravirine	6.95882	8.166	14.78	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, SLogP
12	Efavirenz	8.5244	9.612	11.32	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, n10FRing
13	Glecaprevir	3.66206	2.205	− 66.08	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, nFARing
14	Ledipasvir	3.67926	2.751	− 33.74	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, nFARing
15	Lopinavir	6.81954	6.698	− 1.81	C18	MW, AATS, MATS, GATS, Axp, JGI10, SLogP
16	Maraviroc	4.06444	2.29	− 77.49	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, JGI10
17	Rilpivirine hydrochloride	3.5655	3.273	− 8.94	C18	MW, AATS, MATS, GATS, Axp, PEOE_VSA5
18	Ritonavir	5.62858	5.71	1.43	C18	MW, AATS, MATS, GATS, Axp, PEOE_VSA5
19	Saquinavir	5.96362	6.583	9.41	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, n10FRing,
20	Sofosbuvir	5.56006	4.905	− 13.35	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, NdsssP
21	Tenofovir	5.35406	4.916	− 8.91	C18	MW, AATS, MATS, GATS, Axp, NdsssP
22	Tenofovir alafenamide	6.42914	6.697	4.00	C18	MW, AATS, MATS, GATS, Axp, NdsssP, JGI10,
23	Tenofovir disoproxil dimer	4.92236	3.6	− 36.73	C18	MW, AATS, MATS, GATS, Axp, Xc-4d, NdsssP,
24	Velpatasvir	3.81528	3.133	− 21.78	C18	MW, AATS, MATS, GATS, Axp, nBonds, nFARing
25	Zidovudine	3.016	2.566	− 17.54	C18	MW, AATS, MATS, GATS, Axp, SLogP
26	Amantadine	7.22792	7.1	− 1.80	C18	MW, AATS, MATS, GATS, Axp, n10FRing
27	Cidofovir	4.77664	4.734	− 0.90	C18	MW, AATS, MATS, GATS, Axp, SdsssP
28	Foscarnet	4.9733	5.6	11.19	C18	MW, AATS, MATS, GATS, Axp, SdsssP
29	Ganciclovir	4.22576	3.196	− 32.22	C18	MW, AATS, MATS, GATS, Axp, SLogP
30	Indinavir	4.26906	4.1	− 4.12	C18	MW, AATS, MATS, GATS, Axp, nFARing
31	Fosamprenavir	6.86348	6.247	− 9.87	C18	MW, AATS, MATS, GATS, Axp, Xc-4d
32	Nelfinavir	9.03022	10.86	16.85	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, PEOE_VSA5, nFARing
33	Entecavir	4.44892	4.18	− 6.43	C18	MW, AATS, MATS, GATS, Axp, SLogP
34	Nitazoxanide	5.31866	4.46	− 19.25	C18	MW, AATS, MATS, GATS, Axp, Ns, N5ring, SMR
35	Raltegravir	10.68226	13.31	19.74	C18	MW, AATS, MATS, GATS, Axp, Xc-6d, PEOE_VSA5
36	Remdesivir	4.6616	4.4	− 5.95	C18	MW, AATS, MATS, GATS, Axp, Xc-6d, NdsssP
37	Ribavirin	3.37434	2.606	− 29.48	C18	MW, AATS, MATS, GATS, Axp, NaaN, PEOE_VSA5
38	Rimantadine	10.23922	13.5	24.15	C18	MW, AATS, MATS, GATS, Axp, n10FRing,
39	Trifluridine	3.10676	2.77	− 12.16	C18	MW, AATS, MATS, GATS, Axp, Xc-6d,
40	Vidarabine	3.10048	2.779	− 11.57	C18	MW, AATS, MATS, GATS, Axp, SLogP
41	Zanamivir	4.39834	3.6	− 22.18	C18	MW, AATS, MATS, GATS, Axp, SLogP
42	Valganciclovir	8.1715	10.1	19.09	C18	MW, AATS, MATS, GATS, Axp, SLogP
43	Pibrentasvir	4.10828	4.996	17.77	C18	MW, AATS, MATS, GATS, Axp, SLogP
44	Bictegravir	5.64744	5.998	5.84	C18	MW, AATS, MATS, GATS, Axp, SLogP
45	Emtricitabine	3.12968	2.805	− 11.58	C18	MW, AATS, MATS, GATS, Axp, PEOE_VSA5,
46	Voxilaprevir	4.02044	3.768	− 6.70	C18	MW, AATS, MATS, GATS, Axp, nAtom, nP, nf, Xc-6d, NsF
47	Simeprevir	5.71292	6.398	10.71	C18	MW, AATS, MATS, GATS, Axp, Xch-3d, n10FRing
48	Valacyclovir	4.86238	3.952	− 23.04	C18	MW, AATS, MATS, GATS, Axp, NsNH2, n9Ring
49	Tipiracil	4.45506	5.118	12.95	C18	MW, AATS, MATS, GATS, Axp, SLogP
50	Favipiravir	6.94756	7.69	9.65	C18	MW, AATS, MATS, GATS, Axp, n6aHRing
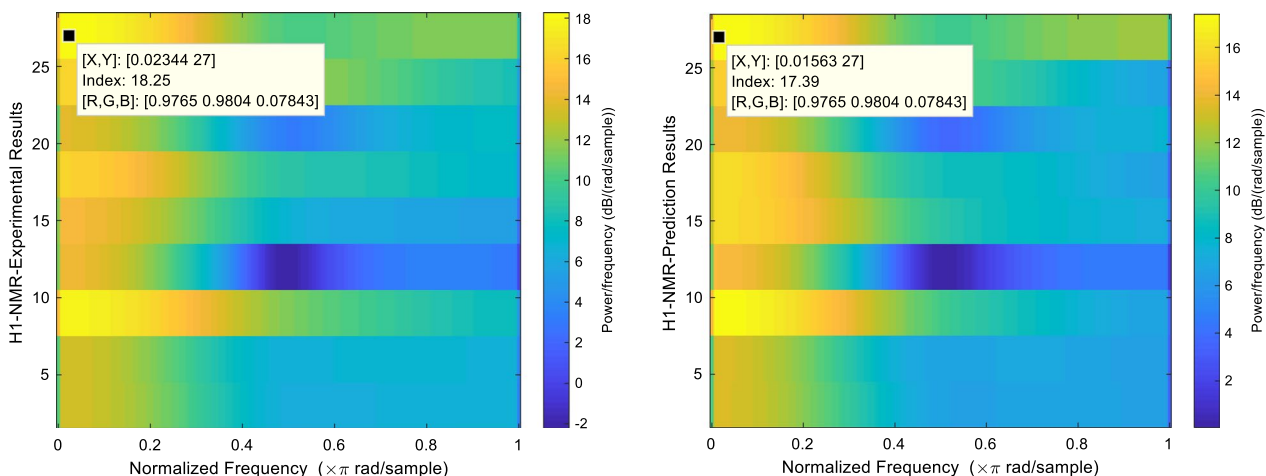
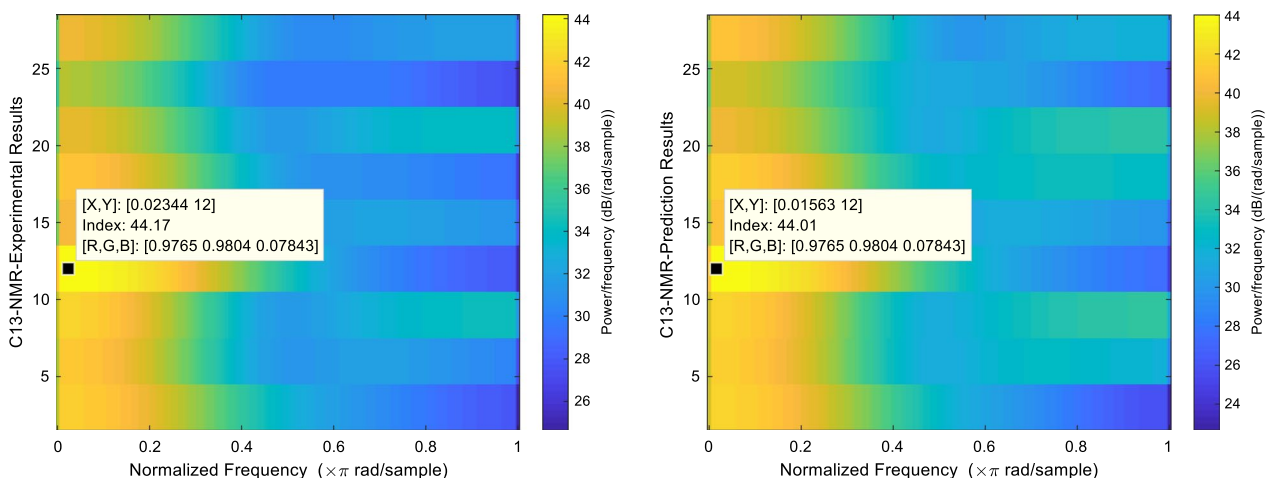**Fig. 16** STFT spectrogram plot of H1-NMR for predicted and published results

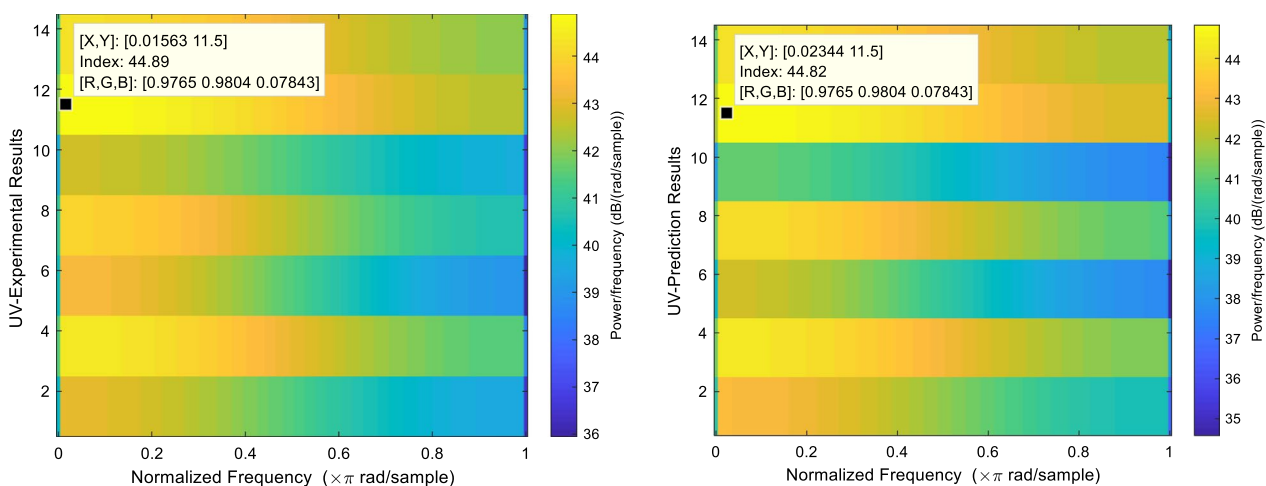**Fig. 17** STFT spectrogram plot of C13-NMR for predicted and published results

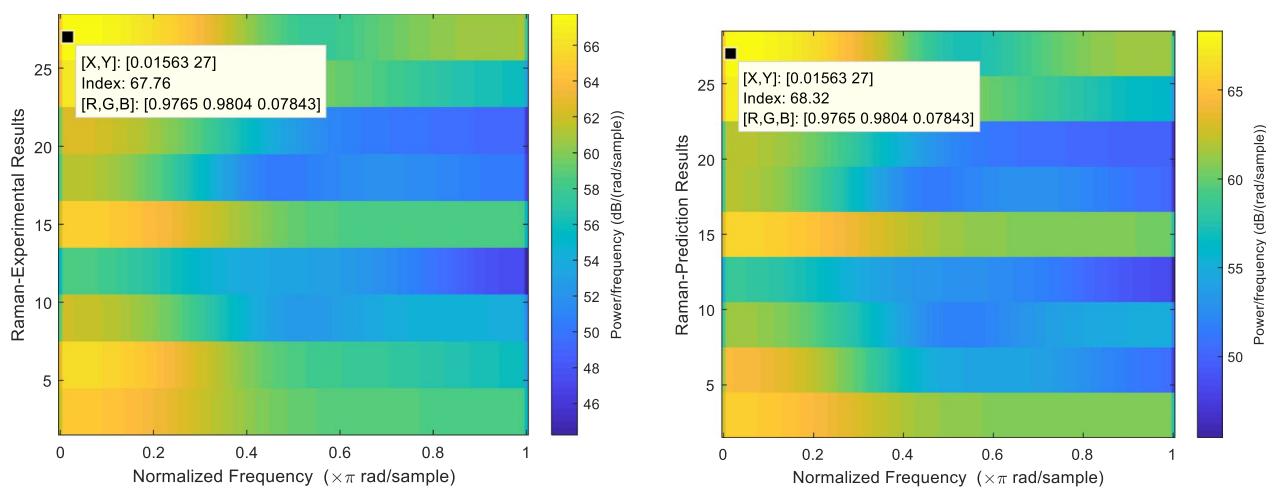**Fig. 18** STFT spectrogram plot of UV–visible for predicted and published results

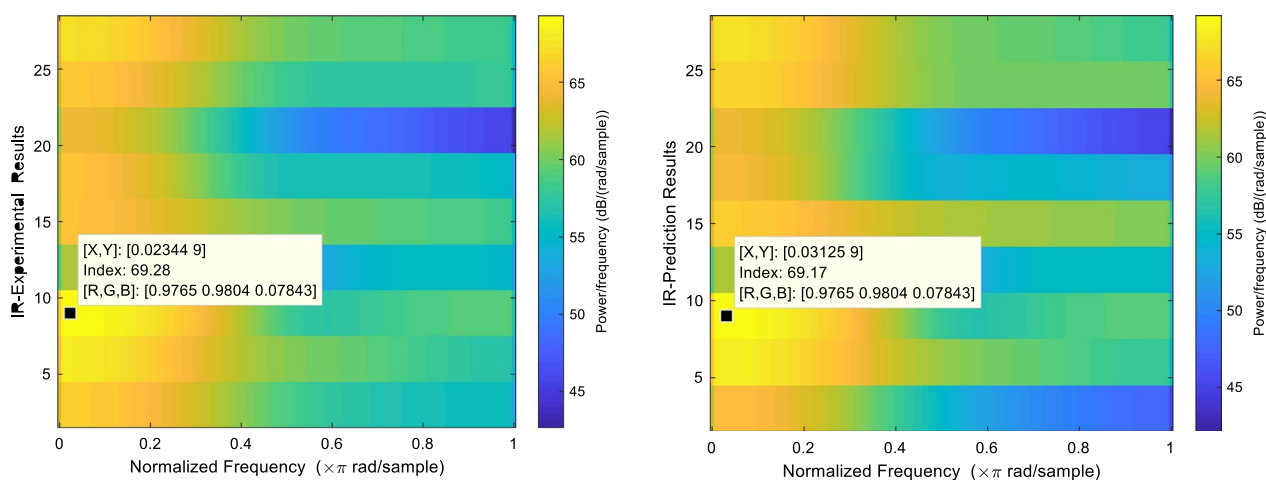**Fig. 19** STFT spectrogram plot of Raman for predicted and published results

Fig. 20 STFT spectrogram plot of IR for predicted and published results

Finally, the Raman and infrared power frequency index can be observed in Figs. 19 and 20 that both the predicted and experimental data are shown using the same frequency index and Fig. 21 shows the STFT spectrogram 3D plot for IR prediction results all these results providing us to confirm that the prediction was accurate.

Discussion

Our assessment afforded the acceptable results, however few software-related constraints, particularly time consumption of 5–20 h for TD-DFT calculations. The prediction error will produce more erroneous findings when the data set is small and a prediction tool required to be unique. Therefore, a large data set is necessary for successful finding; but, in some cases, a large data set can also result in inaccurate prediction, e.g. a complicated structure with multiple classes of variables

takes longer time to process, and the impact on the prediction process ultimately leads to wrong results, which is disappointing for a research study. Therefore, careful planning in the dataset and systematic prediction are required to produce reliable research findings. Then, while collecting the reference data set, we stumbled into issues with some data not being present in the reference library. In that situation, leaving the compound and switching to another approach might be an option. For example, if two distinct spectrum results for the same chemicals are found in certain reference data, such case data optimization need to be performed. There are several online reference data sources available for mass spectroscopy; however, there are fewer for infrared Raman spectra. In Table 9, all these problems and challenges related to spectrogram prediction are listed.
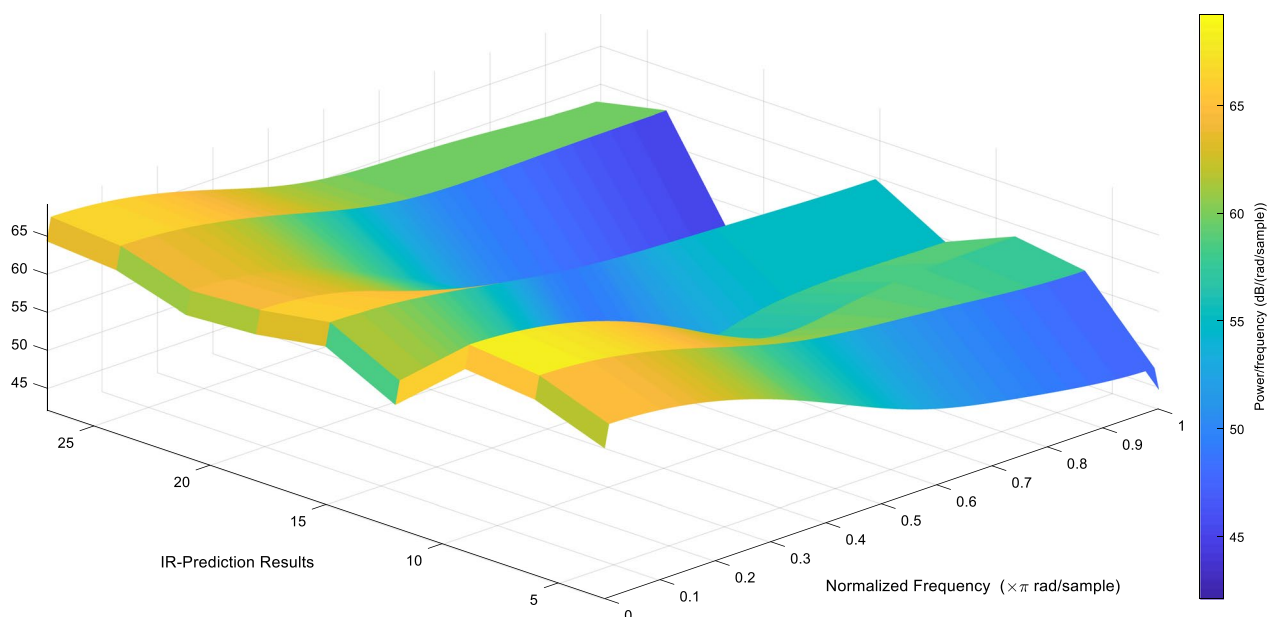
Fig. 21 STFT spectrogram 3D plot for IR prediction results

Table 9 Spectrogram behaviour predictions' limitations and benefits

S. No.	Prediction/method	Limitations/disadvantages	Advantage/applications
1	NMR spectra prediction	Based on machine learning Require huge data sets to be trained for effective findings Defining data requirement is difficult Probability of a poor model is high Training set should be unique Predictions are always accompanied by uncertainties [42] The predictor Chemaxon requires the uniqueness of chemical structures Creation of large, complex chemical structures is challenging Unable to predict higher frequency range compounds (60.0–1000 Hz)	Prediction is possible in one or two minutes even if the structure is huge and complicated Estimates the NMR signals of new and exciting compounds (60.0–1000 Hz) User-friendly and suitable for natural product dereplication More uniqueness for a new chemical in comparison to all others [273] Well-known method for facilitating structural elucidation Even for a molecule of polyatomic (H, O, C, S, N, B, F, I, Cl, Br, P, Si, As, Ge, Sn, and Se,), it can predict 13C and 1H NMR spectra [48]
2	UV–visible spectra prediction	The solvation model methanol produced delayed findings as compared to the solvation modal water The prediction of very complicated compounds takes significantly longer, around 5–12 h Some compounds provide the desired outcomes after 24 h The spectrum visualization gave positive findings, although the spectrogram saved was of poor quality	TD-DFT is one of the most common approaches to predicting UV–visible spectra Simple and not computationally too expensive [50] Alternative solvation models, investigators may compute the number of excited states of various substances The error is frequently lower than using the experimental technique The computation is completed considerably faster

Table 9 (continued)

S. No.	Prediction/method	Limitations/disadvantages	Advantage/applications
3	Infrared and Raman spectra prediction	The optimization of the structure is extremely critical Little variations in the optimized structure lead to incorrect absorption frequency predictions Raman intensities are a more involved problem than IR intensities They depend much more on the experimental setup The prediction of very complicated compounds takes significantly longer, around 24–48 h Even so, the compounds do not always provide the desired outcomes Low quality of the spectrogram generated	The prediction was quite straightforward and user-friendly The functional code was not very complex Infrared/Raman predictors may be used to estimate the absorption frequencies of novel and exciting compounds For example, using the functional code "! Opt numFreq" to predict the IR and Raman spectra of a single molecule at the same time [52]
4	Mass spectroscopy predictions	The library matching strategy has a coverage issue Expensive and time-consuming for capturing additional spectra Prediction techniques are expensive to compute The performance of the forward prediction mode is weak	A variety of reference data sources, including the NIST, NIH, and EPA MS databases, are available for prediction Bidirectional prediction mode captures the fragmentation events more accurately Electron ionization can be accurately predicted using quantum chemistry calculations
5	Fluorescence spectroscopy predictions	Smaller data sets produce less accuracy Higher degree conjugated bonding patterns for emission spectra are required to be calculated Supportive chemometric techniques, such as PCA, are necessary	Lowering of the calculation's computational cost It offers an easy method to perform effective geometry optimizations There is good agreement between the computed values of vertical emission energy
6	Electrochemistry predictions	First, as EIS data is inherently noisy, spectra regression and prediction at unknown frequencies are made more difficult Typically, tested frequencies have a predetermined number of points per decade that are log-spaced	Few deviations and a short experimentation time Active learning acquisition with GPs

In QSRR approach, more than 50 compounds are needed for prediction of retention time. During operation, we noticed that whenever a data set was given, it was based on predicting values nearby. This may be a problem with the QSRR Automator programme, but more sophisticated software for retention time prediction is already available, so we can utilize it for alternative purposes. The chemometric theory is entirely mathematically based, understanding AQbD and chemometrics is more critical in nature. If a chemist is not familiar with mathematical, it will be harder to develop a prediction process. Each approach in chemometrics has a unique methodology, thus experts are required for both planning and result evaluation. Additionally, we noted in the literature study that there is less research on the electrochemistry spectroscopic prediction with chemometrics. Generally, the electrochemistry prediction will be employed in the technological field, but only when few drugs are discovered and developed. Since there are so many variables that might influence the results, such as instrument setting, calibration, process, and model selections, some AQbD method failure will certainly occur in the case of method replacements. However, this strategy is most effective at minimizing method transfer, OOS, and OOT failure rates. These are presented in Table 10.

The differences between the physical and chemical data predictions are also illustrated in Table 11. By comparison, the physical data prediction is simpler than the chemical data prediction because the latter requires a larger number of supporting techniques and programmes. In addition, it requires larger number of descriptors, and is more challenging for beginners and students.

Table 10 Chromatography behaviour predictions' limitations and benefits

S. No	Prediction/method	Limitations/disadvantages	Advantage/applications
1	QSRR and software	Prediction must require more than 50 compounds for study Statistical models that extrapolate outside the training set's retention times may produce erroneous findings Even if the training set included a range of chemically varied molecules with structures that differ from the training sets are unlikely to be successfully predicted To accommodate the changes in the more complicated compounds, the training set must be bigger With an equal-sized training set, lipids provide better predictions than metabolites It is insufficient to separate very near co-elute compounds It is difficult to predict retention time when using several columns and conditions	The predictor places the majority of predictions within one or two minutes/ One minute of their real value equals about 5–8% of the retention time Approximately, all predictions fall between 2–5 min about 10–25 per cent depending on the column It will quickly and easily produce and save a large number of models In-house data tests yielded comparable results To estimate retention times for numerous columns and conditions It is adequate to enhance confidence in exact identifications Compound of the same interests that are clearly identified
2	AQbD	When overcome, can result in method failures and, in certain cases method replacements Chromatographic methods where the number of analytes is required for effective separation Since there are so many factors that impact the method's outcomes Applying the AQbD paradigm to analytical methods is justifiable Instrument settings, sample characteristics, procedure parameters, and calibration model selection are examples of these factors [193]	Enhanced method efficiency; fewer trials, resulting in lower method cost; time utilization; levels of compliance; and knowledge of the extremes As the technique demonstrates a link between the method variables and performance The analyst gains confidence in the method's effectiveness Analytical techniques are re-evaluated regularly to resolve any gaps in method performance To avoid failures in method transfer, OOS and OOT, AQbD methodology might be used AQbD allows for regulatory flexibility, but it necessitates to high level of robustness
3	Chemometrics	Some good laboratory-based analysers are not all mathematically minded, so they did not want to overburden their studies with maths Teaching and learning, chemometrics is still having problems integrating itself Chemometrics is partial because the fundamental body of information is overburdened Any new content must supplant previous subjects Basic statistical knowledge, such as univariate calibration, precision, accuracy, and uncertainty, is necessary [275]	Chemometrics technique advancement is continual, rapid, and efficient With the improvements in exploratory tools, they provide rich information about chemical systems [276] Adaptability for analysis of complex chemical process data in the industry [277] Quality control of herbal drugs, food analysis like vegetables [278–280], fruits [281] grains [282, 283], proteins [284] etc. Environmental chemistry studies [285] and assessment of the results The development of high-throughput chromatographic and spectroscopic data calculations
3.1	ANN	Retention time of new congener cannot be predicted No information about the relationship between molecular property and retention behaviour can be obtained [286]	An accurate and reliable Human brain way it is working Optimize the separation without employing analyte properties
3.2	SVM	Introducing bias in the results Class-modelling technique should only consider sensitivity when deciding on the ideal circumstances for process parameters	Exhibit better overall performance These models and the experimental result agree well The distribution of non-target samples is prevented from the target samples
3.3	MLR-ALS	Fail to develop an appropriate QSPR model Less prediction when compare with the SVM model [287]	Measuring the amount of variables Improved selectivity through improved chemical information separation from interference effects and higher signal-to-noise ratios, which improve chemical distribution visualization

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 33 of 42

Table 10 (continued)

S. No	Prediction/method	Limitations/disadvantages	Advantage/applications
3.4	PARAFAC	In some cases, PARAFAC is too restrictive to model the data in a sensible way Problems with scattering and missing values, for instance, have attracted specific focus	Quicker and more reliable computationally than SVD Applications include five-way data analysis and analysis of variance Relatively straightforward mathematically
3.5	PCR and PLSR	Risk of overfitting due to bias-variance trade-off Differences between bias and variance can have a significant impact on predictions [288]	Multi-resolution of multi-component mixtures Accuracy ranging from 98 to 103% [166, 288–290]
3.6	PLS	The method has more challenging Minor inconvenience compared to the risk of change correlation [291] Overstate the baseline when additive noise is present	A six-component combination was resolved using PLS [292] It is used to build multivariate calibration models [293]
3.7	PCA	Sample load at the small scale was insufficient to define the second principal compound Without a mathematical model the analysis is inaccurate Direct PCA is not suitable for the raw data	Predict chromatogram shapes at different scale of operation [294] Single principal compound is sufficient to explain most
3.8	TLRC and MLRC	Cannot be employed more than three compound mixtures Inaccurate calculations are caused by errors in wavelength set selection	Commonly employed in ternary mixture multi-resolution [164] A combination of strongly overlapping spectra, the accuracy is in the region of 99–101% More reliable with simple mathematical calculation [295]

Table 11 Differences in physical and chemical data predictions

Physical data prediction	Chemical data prediction
Falls under the predictions of mass, capacity, density, response time, and size	Falls under prediction of the electronegativity, ionic potential, bonding energy, electrochemical characteristics, fragmentation, and structural attributes
There is no more uncomplicated process	The process is little complicated
Involvement of fewer descriptors in the prediction process	Involvement of a large number of descriptors in the prediction process
1D, 2D, and topological descriptors are mostly utilized for the prediction [271]	All topological, electronic, 1D, 2D, and 3D descriptors were utilized for the prediction [272]
The database size is typically smaller	The database is larger and more complex
Canonical SMILES are includes in the data set, mostly, 2D chemical structure is sufficient for the prediction	Data sets used for prediction includes canonical SIMILES, ChEMBL, and 2D and 3D chemical structures
Predictions scores typically produce immediate results	Prediction outcomes should not be direct; instead, they correlate with other outcomes to produce final outcomes
There are fewer accessible methods and approaches; most are QSPR-based	Numerous techniques and strategies are employed, such as QSAR and QSRR
Mathematical calculations only involved such as regression and correlations	There are included quantum chemical calculations like TD-DFT and DFT
The final outcomes are unaffected by the less accurate physical data predictions in some cases	Less accurate chemical data prediction will have an impact on the outcomes

Conclusions

Finally, with acceptable accuracy and the least feasible variation, the present review of computational approaches in spectrum prediction was concluded. Overall, students and researchers are considerably utilizing the in silico tools in computational chemistry and indicate the reliability of such tools in research. The development and application of computational approaches in analytical research and development are our key objectives. As we observed, computational analytical behaviour prediction offers a wide range of applications in academic research, bioanalytical method development, computational chemistry, analytical method development, data analysis approaches, material characterization, and validation. Still, the prediction error of these tools need to be minimized for better accuracy, thus it will be explored much more in exploratory research in future.

Abbreviations

UV	Ultraviolet
IR	Infrared
NMR	Nuclear magnetic resonance
GC	Gas chromatography
ANNs	Artificial neural networks
RSM	Response surface methodology
AQbD	Analytical quality by design
DoE	Design of experiments
QSRR	Quantitative structure retention relationship
OOT	Out of trend
OOS	Out of specification
AI	Artificial intelligence
ASV	Automatic structure verification
TD-DFT	Time-dependent density function theory
EI-MS	Electron ionization mass spectrometry
NEIMS	Neural electron ionization mass spectrometry
QCEIMS	Quantum chemical electron ionization mass spectrometry
MD	Molecular dynamics
ESI–MS/MS	Electrospray ionization mass spectroscopy
EOM-CCSD	Equation of motion coupled cluster singles and doubles
PCM	Polarizable continuum model
QSAR	Qualitative structural activity relationship
Lasso-RF	Random forest descriptor
EIS	Electrochemical impedance spectroscopy
GPs	Computational Gaussian processes
TDMs	Transition density matrices
ICP-MS	Inductively coupled plasma mass spectroscopy
XRD	X-ray diffraction
CSV	Comma-separated values
SMILES	Simplified molecular input line entry system
RMSE	Root-mean-squared error
FD	Factorial design
PLS	Partial least squares
CA	Cluster analysis
PCR	Principal component regression
TLRC	Trilinear regression calibration
MLRC	Multi-linear regression calibration
PLSR	Partial least square regression
ILS	Inverse least square
PCA	Principal component analysis
OPA	Orthogonal projection analysis
EFA	Evolving factor analysis
MCR-ALS	Multivariate curve resolution by alternating the least squares approach
SFA	Sub-window factor analysis
PARAFAC	Parallel factor analysis
GRAM	Generalized rank annihilation
PLS-DA	Partial least square discriminate analysis
HCA	Hierarchical cluster analysis
SIMCA	Soft independent modelling of class analogy
RAFA	Rank annihilation factor analysis
LC–NMR	Liquid chromatography nuclear magnetic resonance
LC–MS	Liquid chromatography–mass spectrometry
GC–MS	Gas chromatography–mass spectrometry
FT-IR	Fourier transform infrared
HPLC	High-performance liquid chromatography
UPLC	Ultra-performance liquid chromatography
HPTLC	High-performance thin-layer chromatography
SFC	Supercritical fluid chromatography
PAT	Process analytical technique
OFAT	One-factor-at-a-time
FR	Flow rate
CT	Column temperature
AQ	Aqueous
ACN	Acetonitrile
RSD	Relative standard deviation
TBAH	Tetrabutylammonium hydroxide
PDOP	Potassium dihydrogen orthophosphate
TEA	Triethylamine
STFT	Short-time Fourier transform
QSPR	Qualitative structural property relationship

Declarations

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors whose names are listed in this manuscript that they have NO affiliations with or involvement in any organization or entity with any financial or nonfinancial interest in the subject matter or materials discussed in this manuscript.

Author details
[1]Department of Pharmaceutical Analysis, National Institute of Pharmaceutical Education and Research (NIPER)-Kolkata, Kolkata, West Bengal 700054, India. [2]Department of Pharmaceutical Analysis, Raghavendra Institute of Pharmaceutical Education and Research (RIPER)-Autonomous, Anantapur, Andhra Pradesh 515721, India. [3]Department of Pharmaceutical Analysis, National Institute of Pharmaceutical Education and Research (NIPER)-Hajipur, Hajipur, Bihar 844402, India. [4]Department of Natural Products, National Institute of Pharmaceutical Education and Research (NIPER)-Kolkata, Kolkata, West Bengal 700054, India.

References

1. Nova A, Maseras F (2013) Enantioselective synthesis. In: Comprehensive inorganic chemistry ii (second edition): from elements to applications, pp. 807–831
2. Genheden S, Reymer A, Saenz-Méndez P, Eriksson LA (2017) Computational chemistry and molecular modelling basics
3. Polanski J, Gasteiger J (2016) Computer representation of chemical compounds. J Puzyn T Eds. https://doi.org/10.1007/978-94-007-6169-8_50-1
4. Gerlich M, Neumann S (2013) Metfusion: integration of compound identification strategies. J Mass Spectrom 48(3):291–298
5. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S (2010) In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf 11(1):1–12
6. Peironcely JE, Rojas-Chertó M, Tas A, Vreeken R, Reijmers T, Coulier L, Hankemeier T (2013) Automated pipeline for de novo metabolite identification using mass-spectrometry-based metabolomics. Anal Chem 85(7):3576–3583

7. Snyder HD, Kucukkal TG (2021) Computational chemistry activities with avogadro and orca. J Chem Educ 98(4):1335–1341. https://doi.org/10.1021/acs.jchemed.0c00959

8. Kotha RR, Natarajan S, Wang D, Luthria DL (2019) Compositional analysis of non-polar and polar metabolites in 14 soybeans using spectroscopy and chromatography tools. Foods 8(11):557

9. Kaleta M, Oklestkova J, Novák O, Strnad M (2021) Analytical methods for the determination of neuroactive steroids. Biomolecules 11(4):553

10. Vandierendonck A (2017) A comparison of methods to combine speed and accuracy measures of performance: a rejoinder on the binning procedure. Behav Res Methods 49(2):653–673

11. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK (2021) Artificial intelligence in drug discovery and development. Drug Discov Today 26(1):80

12. Udayakumar V, Periandy S, Ramalingam S (2011) Experimental (ft-ir and ft-raman) and theoretical (hf and dft) investigation, ir intensity, raman activity and frequency estimation analyses on 1-bromo-4-chloroben-zene. Spectrochim Acta A Mol Biomol Spectrosc 79(5):920–927. https://doi.org/10.1016/j.saa.2011.03.049

13. Guideline, I. H. T. (2017). Technical and regulatory considerations for pharmaceutical product lifecycle management q12. Paper presented at the International Conference on Harmonisation of Technical Require-ments for Registration of Pharmaceuticals for Human Use.

14. FDA. (2011). Food drug administration. Pharmaceutical quality system (ich 10) conference. Accessed Jul 2021 from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/q10-pharmaceutical-quality-system

15. Patel KY, Dedania ZR, Dedania RR, Patel U (2021) Qbd approach to hplc method development and validation of ceftriaxone sodium. Future J Pharm Sci 7(1):141. https://doi.org/10.1186/s43094-021-00286-4

16. Peraman R, Bhadraya K, Padmanabha Reddy Y (2015) Analytical quality by design: a tool for regulatory flexibility and robust analytics. Int J Anal Chem. https://doi.org/10.1155/2015/868727

17. Agatonovic-Kustrin S, Zecevic M, Zivanovic L, Tucker I (1998) Applica-tion of artificial neural networks in HPLC method development. J Pharm Biomed Anal 17(1):69–76

18. Webb R, Doble P, Dawson M (2009) Optimisation of hplc gradient sepa-rations using artificial neural networks (anns): application to benzodiaz-epines in post-mortem samples. J Chromatogr B 877(7):615–620

19. Chatterjee S (2013) QBD considerations for analytical methods—FDA perspective. Paper presented at the US IFPAC annual meeting

20. Burnett K, Harrington B, Graul T, Fanalis S, Haddad P, Poole C (2013) Qbd in liquid chromatographic applications. Elsevier

21. Kaliszan R (2000) Chapter 11 recent advances in quantitative structure-retention relationships (QSRR). In: Valkó K (ed) Handbook of analytical separations. Elsevier Science, pp 503–534

22. Héberger K (2007) Quantitative structure–(chromatographic) reten-tion relationships. J Chromatogr A 1158(1):273–305. https://doi.org/10.1016/j.chroma.2007.03.108

23. Amos RIJ, Haddad PR, Szucs R, Dolan JW, Pohl CA (2018) Molecular modeling and prediction accuracy in quantitative structure-retention relationship calculations for chromatography. TrAC, Trends Anal Chem 105:352–359. https://doi.org/10.1016/j.trac.2018.05.019

24. De Matteis CI, Simpson DA, Doughty SW, Euerby MR, Shaw PN, Barrett DA (2010) Chromatographic retention behaviour of n-alkylbenzenes and pentylbenzene structural isomers on porous graphitic carbon and octadecyl-bonded silica studied using molecular modelling and QSRR. J Chromatogr A 1217(44):6987–6993

25. MacNeil JD (2012) Analytical difficulties facing today's regulatory labo-ratories: issues in method validation. Drug Test Anal 4(Suppl 1):17–24. https://doi.org/10.1002/dta.1358

26. Volta ESL, Gonçalves R, Menezes JC, Ramos A (2021) Analytical method lifecycle management in pharmaceutical industry: a review. AAPS PharmSciTech 22(3):128. https://doi.org/10.1208/s12249-021-01960-9

27. Yang W, Qian W, Yuan Z, Chen B (2022) Perspectives on the flexibility analysis for continuous pharmaceutical manufacturing processes. Chin J Chem Eng 41:29–41. https://doi.org/10.1016/j.cjche.2021.12.005

28. Akash MSH, Rehman K (2020) Introduction to pharmaceutical analysis. In: Akash MSH, Rehman K (eds) Essentials of pharmaceutical analysis. Springer Nature Singapore, Singapore, pp 1–18

29. Cadinoska M, Popstefanova N, Ilievska M, Karadzinska E, Jovanoska M, Glavas Dodov M (2019) Trending and out-of-trend results in pharma-ceutical industry. Maced Pharm Bull 65:39–60. https://doi.org/10.33320/maced.pharm.bull.2019.65.01.005

30. Appleton T, Bryan P, Contos D et al (2012) Nonclinical dose formulation: out of specification investigations. Aaps J 14(3):523–529. https://doi.org/10.1208/s12248-012-9347-4

31. Martinez Calatayud J (2005) Spectrophotometry | pharmaceutical applications. In: Worsfold P, Townshend A, Poole C (eds) Encyclopedia of analytical science, 2nd edn. Elsevier, Oxford, pp 373–383

32. Simundic AM, Lippi G (2012) Preanalytical phase–a continuous chal-lenge for laboratory professionals. Biochem Med 22(2):145–149. https://doi.org/10.11613/bm.2012.017

33. Paré GKS (2017) Handbook of ehealth evaluation: an evidence-based approach. University of Victoria, Victoria

34. Redrup MJ, Igarashi H, Schaefgen J et al (2016) Sample management: recommendation for best practices and harmonization from the global bioanalysis consortium harmonization team. Aaps J 18(2):290–293. https://doi.org/10.1208/s12248-016-9869-2

35. Piskunov DP, Danilova LA, Pushkin AS, Rukavishnikova SA (2020) Influence of exogenous and endogenous factors on the quality of the preanalytical stage of laboratory tests (review of litera-ture). Klin Lab Diagn 65(12):778–784. https://doi.org/10.18821/0869-2084-2020-65-12-778-784

36. Krčmová LK, Melichar B, Švec F (2020) Chromatographic methods development for clinical practice: requirements and limitations. Clin Chem Lab Med 58(11):1785–1793. https://doi.org/10.1515/cclm-2020-0517

37. Patil R, Bhaskar R, Ola M, Pingale D, Chalikwar SS (2019) Bioanalytical method development and method validation in human plasma by using LC MS/MS

38. Khamis MM, Adamko DJ, El-Aneed A (2021) Strategies and challenges in method development and validation for the absolute quantifica-tion of endogenous biomarker metabolites using liquid chromatog-raphy-tandem mass spectrometry. Mass Spectrom Rev 40(1):31–52. https://doi.org/10.1002/mas.21607

39. Ragoisha G (2020) Challenge for electrochemical impedance spectroscopy in the dynamic world. J Solid State Electrochem 24:2171–2172

40. Pierce KM, Trinklein TJ, Nadeau JS, Synovec RE (2021) Chapter 20 - data analysis methods for gas chromatography. In: Poole CF (ed) Gas chro-matography, 2nd edn. Elsevier, Amsterdam, pp 525–546

41. Oliveri P, Malegori C, Simonetti R, Casale M (2019) The impact of signal pre-processing on the final interpretation of analytical outcomes - a tutorial. Anal Chim Acta 1058:9–17. https://doi.org/10.1016/j.aca.2018.10.055

42. Cobas C (2020) NMR signal processing, prediction, and structure verification with machine learning techniques. Magn Reson Chem 58(6):512–519

43. Ito K, Xu X, Kikuchi J (2021) Improved prediction of carbonless NMR spectra by the machine learning of theoretical and fragment descrip-tors for environmental mixture analysis. Anal Chem 93(18):6901–6906

44. Kern S, Liehr S, Wander L, Bornemann-Pfeiffer M, Müller S, Maiwald M, Kowarik S (2020) Artificial neural networks for quantitative online NMR spectroscopy. Anal Bioanal Chem 412(18):4447–4459

45. Jonas E, Kuhn S (2019) Rapid prediction of NMR spectral properties with quantified uncertainty. J Cheminf 11(1):1–7

46. Jia W, Yang Z, Yang M, Cheng L, Lei Z, Wang X (2021) Machine learning enhanced spectrum recognition based on computer vision (SRCV) for intelligent NMR data extraction. J Chem Inf Model 61(1):21–25. https://doi.org/10.1021/acs.jcim.0c01046

47. Lin M, Xiong J, Su M et al (2022) A machine learning protocol for reveal-ing ion transport mechanisms from dynamic NMR shifts in paramag-netic battery materials. Chem Sci 13(26):7863–7872. https://doi.org/10.1039/D2SC01306A

48. Chemaxon (2021) Chemaxon software solution service for chemistry and biology. Accessed May 2021 from https://chemaxon.com/products/calculators-and-predictors

49. Mamede R, Pereira F, Aires-de-Sousa J (2021) Machine learning predic-tion of UV–vis spectra features of organic compounds related to

photoreactive potential. Sci Rep 11(1):23720. https://doi.org/10.1038/s41598-021-03070-9

50. Chan B, Hirao K (2020) Rapid prediction of ultraviolet-visible spectra from conventional (non-time-dependent) density functional theory calculations. J Phys Chem Lett 11(18):7882–7885. https://doi.org/10.1021/acs.jpclett.0c02146

51. Urbina F, Batra K, Luebke KJ et al (2021) UV-advisor: attention-based recurrent neural networks to predict UV–vis spectra. Anal Chem 93(48):16076–16085. https://doi.org/10.1021/acs.analchem.1c03741

52. ORCA (2021) Orca. Basis sets. Orca input library. Accessed May 2021 from https://sites.google.com/site/orcainputlibrary/basis-sets

53. (2021) UV-vis spectroscopy orca tutorials 5.0 documentation. Accessed May 2021 from https://www.orcasoftware.de/tutorials_orca/spec/uvvis.html

54. Orca A (2021) Avogadro orca: an open-source molecular builder and visualization tool, version 4.2. Accessed May 2021 from https://avogadro.cc/

55. Neese F (2012) The orca program system. Wiley Interdiscip Rev Comput Mol Sci 2(1):73–78

56. Avogadro (2021) Avogadro program manual. Accessed Jun 2021 from https://avogadro.cc/docs

57. Neese F, Wennmohs F, Becker U, Riplinger C (2020) The orca quantum chemistry program package. J Chem Phys 152(22):224108. https://doi.org/10.1063/5.0004608

58. Stanstrup J, Neumann S, Vrhovšek U (2015) Predret: prediction of retention time by direct mapping between multiple chromatographic systems. Anal Chem 87(18):9421–9428. https://doi.org/10.1021/acs.analchem.5b02287

59. McGill C, Forsuelo M, Guan Y, Green WH (2021) Predicting infrared spectra with message passing neural networks. J Chem Inf Model 61(6):2594–2609. https://doi.org/10.1021/acs.jcim.1c00055

60. Maciel EVS, Pereira Dos Santos NG, Vargas Medina DA, Lanças FM (2022) Electron ionization mass spectrometry: Quo vadis? Electrophoresis 43(15):1587–1600. https://doi.org/10.1002/elps.202100392

61. Smith RW (2013) Mass spectrometry. In: Siegel JA, Saukko PJ, Houck MM (eds) Encyclopedia of forensic sciences, 2nd edn. Academic Press, Waltham, pp 603–608

62. Bauer CA, Grimme S (2016) How to compute electron ionization mass spectra from first principles. J Phys Chem A 120(21):3755–3766. https://doi.org/10.1021/acs.jpca.6b02907

63. Zhou Z, Zare RN (2017) Personal information from latent fingerprints using desorption electrospray ionization mass spectrometry and machine learning. Anal Chem 89(2):1369–1372. https://doi.org/10.1021/acs.analchem.6b04498

64. Wei JN, Belanger D, Adams RP, Sculley D (2019) Rapid prediction of electron-ionization mass spectrometry using neural networks. ACS Cent Sci 5(4):700–708. https://doi.org/10.1021/acscentsci.9b00085

65. Wang S, Kind T, Tantillo DJ, Fiehn O (2020) Predicting in silico electron ionization mass spectra using quantum chemistry. J Cheminform 12(1):63. https://doi.org/10.1186/s13321-020-00470-3

66. Allen F, Pon A, Greiner R, Wishart D (2016) Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. Anal Chem 88(15):7689–7697. https://doi.org/10.1021/acs.analchem.6b01622

67. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS (2021) Cfm-id 4.0: more accurate ESI-MS/MS spectral prediction and compound identification. Anal Chem 93(34):11692–11700. https://doi.org/10.1021/acs.analchem.1c01465

68. Kannan R, Solaimalai A, Jayakumar M, Surendran U (2022) Chapter 26 - advance molecular tools to detect plant pathogens. In: Rakshit A, Meena VS, Abhilash PC, Sarma BK, Singh HB, Fraceto L, Parihar M, Singh AK (eds) Biopesticides. Woodhead Publishing, pp 401–416

69. Zhu G, Bian Y, Hursthouse AS et al (2017) Application of 3-d fluorescence: characterization of natural organic matter in natural water and water purification systems. J Fluoresc 27(6):2069–2094. https://doi.org/10.1007/s10895-017-2146-7

70. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105(8):2999–3093. https://doi.org/10.1021/cr9904009

71. Shavitt I, Bartlett RJ (2009) Many-body methods in chemistry and physics: Mbpt and coupled-cluster theory

72. Pavošević F, Hammes-Schiffer S (2019) Multicomponent equation-of-motion coupled cluster singles and doubles: theory and calculation of excitation energies for positronium hydride. J Chem Phys 150(16):161102. https://doi.org/10.1063/1.5094035

73. Caricato M (2012) Absorption and emission spectra of solvated molecules with the eom–ccsd–pcm method. J Chem Theory Comput 8(11):4494–4502. https://doi.org/10.1021/ct3006997

74. Powell J, Heider EC, Campiglia A, Harper JK (2016) Predicting accurate fluorescent spectra for high molecular weight polycyclic aromatic hydrocarbons using density functional theory. J Mol Spectrosc 328:37–45. https://doi.org/10.1016/j.jms.2016.06.015

75. Ye Z-R, Huang I-S, Chan Y-T et al (2020) Predicting the emission wavelength of organic molecules using a combinatorial qsar and machine learning approach. RSC Adv 10:23834–23841

76. Shams-Nateri A, Piri N (2016) Prediction of emission spectra of fluorescence materials using principal component analysis. Color Res Appl 41(1):16–21. https://doi.org/10.1002/col.21959

77. Mai S, Atkins AJ, Plasser F, González L (2019) The influence of the electronic structure method on intersystem crossing dynamics. The case of thioformaldehyde. J Chem Theory Comput 15(6):3470–3480

78. Ohto T, Dodia M, Xu J et al (2019) Accessing the accuracy of density functional theory through structure and dynamics of the water-air interface. J Phys Chem Lett 10(17):4914–4919. https://doi.org/10.1021/acs.jpclett.9b01983

79. Wang S, Zhang J, Gharbi O, Vivier V, Gao M, Orazem ME (2021) Electrochemical impedance spectroscopy. Nat Rev Methods Prim 1:41

80. Magar HS, Hassan RYA, Mulchandani A (2021) Electrochemical impedance spectroscopy (EIS): Principles, construction, and biosensing applications. Sensors. https://doi.org/10.3390/s21196578

81. Krukiewicz K (2020) Electrochemical impedance spectroscopy as a versatile tool for the characterization of neural tissue: a mini review. Electrochem Commun 116:106742. https://doi.org/10.1016/j.elecom.2020.106742

82. Pasqual JAR, Freisleben LC, Colpo JC, Egea JRJ, dos Santos LAL, de Sousa VC (2021) In situ drug release measuring in α-tcp cement by electrochemical impedance spectroscopy. J Mater Sci Mater Med 32(4):38. https://doi.org/10.1007/s10856-021-06507-9

83. Heijne A et al (2018) Quantification of bio-anode capacitance in bio-electrochemical systems using electrochemical impedance spectroscopy. J Power Sour 400:533

84. Vadhva P, Hu JX, Johnson MJ, Stocker R, Braglia M, Brett DJL, Rettie AJE (2021) Electrochemical impedance spectroscopy for all-solid-state batteries: theory, methods and future outlook. ChemElectroChem 8(11):1930–1947

85. Meyer Q, Zeng Y, Zhao C (2019) Electrochemical impedance spectroscopy of catalyst and carbon degradations in proton exchange membrane fuel cells. J Power Sourc 437:226922

86. Pajkossy T, Jurczakowski R (2017) Electrochemical impedance spectroscopy in interfacial studies. Curr Opin Electrochem 1(1):53–58. https://doi.org/10.1016/j.coelec.2017.01.006

87. Maradesa A, Py B, Quattrocchi E, Ciucci F (2022) The probabilistic deconvolution of the distribution of relaxation times with finite gaussian processes. Electrochim Acta 413:140119. https://doi.org/10.1016/j.electacta.2022.140119

88. Schulz E, Speekenbrink M, Krause A (2018) A tutorial on gaussian process regression: modelling, exploring, and exploiting functions. J Math Psychol 85:1–16. https://doi.org/10.1016/j.jmp.2018.03.001

89. Liu J, Ciucci F (2020) The gaussian process distribution of relaxation times: a machine learning tool for the analysis and prediction of electrochemical impedance spectroscopy data. Electrochim Acta 331:135316. https://doi.org/10.1016/j.electacta.2019.135316

90. Lu Y, Zhao CZ, Huang J, Zhang QK (2022) The timescale identification decoupling complicated kinetic processes in lithium batteries. Joule. 6:1172

91. Liu J, Wan TH, Ciucci F (2020) A bayesian view on the hilbert transform and the kramers-kronig transform of electrochemical impedance data: probabilistic estimates and quality scores. Electrochim Acta 357:136864

92. Ciucci F (2020) The gaussian process hilbert transform (gp-ht): TESTING the consistency of electrochemical impedance spectroscopy data. J Electrochem Soc 167:126503

93. Kiss FL, Corbet BP, Simeth NA, Feringa BL, Crespi S (2021) Predicting the substituent effects in the optical and electrochemical properties of n, n′-substituted isoindigos. Photochem Photobiol Sci 20(7):927–938. https://doi.org/10.1007/s43630-021-00071-5

94. Li Y, Ullrich CA (2011) Time-dependent transition density matrix. Chem Phys 391:157–163

95. Li H-W, Guan Z, Cheng Y, Lui T, Yang Q, Lee C-S, Tsang S-W (2016) On the study of exciton binding energy with direct charge generation in photovoltaic polymers. Adv Electr Mater. https://doi.org/10.1002/aelm.201600200

96. Albuquerque LS, Arias JJR, Santos BPS, Marques M et al (2020) Synthesis and characterization of novel conjugated copolymers for application in third generation photovoltaic solar cells. J Market Res 9:7975–7988

97. Min K, Choi B, Park K, Cho E (2018) Machine learning assisted optimization of electrochemical properties for ni-rich cathode materials. Sci Rep 8(1):15778. https://doi.org/10.1038/s41598-018-34201-4

98. Min K, Park K, Park SY, Seo S-W, Choi B, Cho E (2017) Improved electrochemical properties of lini0.91co0.06mn0.03o2 cathode material via li-reactive coating with metal phosphates. Sci Rep 7(1):7151. https://doi.org/10.1038/s41598-017-07375-6

99. Jalili-Jahani N, Zeraatkar E (2021) Fuzzy wavelet network based on extended kalman filter training algorithm combined with least square weight estimation: efficient and improved chromatographic QSRR/QSPR models. Chemom Intell Lab Syst 208:104191. https://doi.org/10.1016/j.chemolab.2020.104191

100. Naylor BC, Catrow JL, Maschek JA, Cox JE (2020) Qsrr automator: a tool for automating retention time prediction in lipidomics and metabolomics. Metabolites 10(6):237

101. Wen Y, Amos RIJ, Talebi M, Szucs R, Dolan JW, Pohl CA, Haddad PR (2018) Retention index prediction using quantitative structure-retention relationships for improving structure identification in nontargeted metabolomics. Anal Chem 90(15):9434–9440. https://doi.org/10.1021/acs.analchem.8b02084

102. Daghir-Wojtkowiak E, Studzińska S, Buszewski B, Kaliszan R, Markuszewski MJ (2014) Quantitative structure–retention relationships of ionic liquid cations in characterization of stationary phases for hplc. Anal Methods 6(4):1189–1196. https://doi.org/10.1039/C3AY41805G

103. Goryński K, Bojko B, Nowaczyk A, Buciński A, Pawliszyn J, Kaliszan R (2013) Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds. Anal Chim Acta 797:13–19. https://doi.org/10.1016/j.aca.2013.08.025

104. Bodzioch K, Durand A, Kaliszan R, Baczek T, Vander Heyden Y (2010) Advanced QSRR modeling of peptides behavior in RPLC. Talanta 81(4–5):1711–1718. https://doi.org/10.1016/j.talanta.2010.03.028

105. Filipic S, Nikolic K, Krizman M, Danica A (2008) The quantitative structure–retention relationship (QSRR) analysis of some centrally acting antihypertensives and diuretics. QSAR Comb Sci 27:1036–1044. https://doi.org/10.1002/qsar.200710161

106. Bahmani A, Saaidpour S, Rostami A (2017) Quantitative structure–retention relationship modeling of morphine and its derivatives on ov-1 column in gas–liquid chromatography using genetic algorithm. Chromatographia 80(4):629–636. https://doi.org/10.1007/s10337-017-3273-7

107. Zhang DX, Si HZ, Liu X (2014) Quantitative structure-retention time relationship for retention time of coffee flavor compounds. Adv Mater Res 926–930:1010–1013. https://doi.org/10.4028/www.scientific.net/AMR.926-930.1010

108. Paritala J, Peraman R, Kondreddy VK, Subrahmanyam CVS, Ravichandiran V (2021) Quantitative structure retention relationship (QSRR) approach for assessment of chromatographic behavior of antiviral drugs in the development of liquid chromatographic method. J Liq Chromatogr Relat Technol 44(13–14):637–648. https://doi.org/10.1080/10826076.2022.2025827

109. Akbar J, Iqbal S, Batool F, Karim A, Chan KW (2012) Predicting retention times of naturally occurring phenolic compounds in reversed-phase liquid chromatography: a quantitative structure-retention relationship (qsrr) approach. Int J Mol Sci 13(11):15387–15400. https://doi.org/10.3390/ijms131115387

110. Parinet J (2021) Prediction of pesticide retention time in reversed-phase liquid chromatography using quantitative-structure retention relationship models: a comparative study of seven molecular descriptors

111. Maljurić N, Golubović J, Otašević B, Zečević M, Protić A (2018) Quantitative structure –retention relationship modeling of selected antipsychotics and their impurities in green liquid chromatography using cyclodextrin mobile phases. Anal Bioanal Chem 410(10):2533–2550. https://doi.org/10.1007/s00216-018-0911-3

112. Ji C, Li Y et al (2009) Quantitative structure-retention relationships for mycotoxins and fungal metabolites in LC-MS/MS. J Sep Sci 32:3967–3979. https://doi.org/10.1002/jssc.200900441

113. Szucs R, Brown R, Brunelli C, Heaton JC, Hradski J (2021) Structure driven prediction of chromatographic retention times: applications to pharmaceutical analysis. Int J Mol Sci 22(8):3848

114. Duarte A, Capelo S (2006) Application of chemometrics in separation science. J Liq Chromatogr Relat Technol 29(7–8):1143–1176

115. Bos TS, Knol WC, Molenaar SR, Niezen LE, Schoenmakers PJ, Somsen GW, Pirok BW (2020) Recent applications of chemometrics in one-and two-dimensional chromatography. J Sep Sci 43(9–10):1678–1727

116. Komsta Ł (2012) Chemometrics in fingerprinting by means of thin layer chromatography. Chromatogr Res Int 2012:893246. https://doi.org/10.1155/2012/893246

117. BWIQ (2021). Chemometric tool BWIQ-software package. Accessed Jun 2021 from https://bwtek.Com/support and https://bwtek.Com/videos-applications

118. Whittaker ET (1922) On a new method of graduation. Proc Edinb Math Soc 41:63–75

119. Suzuki T, Yoshida N (2020) Penalized least squares approximation methods and their applications to stochastic processes. Jpn J Stat Data Sci 3(2):513–541. https://doi.org/10.1007/s42081-019-00064-w

120. Carlos Cobas J, Bernstein MA, Martín-Pastor M, Tahoces PG (2006) A new general-purpose fully automatic baseline-correction procedure for 1d and 2d NMR data. J Magn Reson 183(1):145–151. https://doi.org/10.1016/j.jmr.2006.07.013

121. Zhang Z, Chen S, Liang Y et al (2010) An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. J Raman Spectrosc 41:659–669

122. Eilers PH (2003) A perfect smoother. Anal Chem 75(14):3631–3636. https://doi.org/10.1021/ac034173t

123. Zhang ZM, Chen S, Liang YZ (2010) Baseline correction using adaptive iteratively reweighted penalized least squares. Analyst 135(5):1138–1146. https://doi.org/10.1039/b922045c

124. Baek SJ, Park A, Ahn YJ, Choo J (2015) Baseline correction using asymmetrically reweighted penalized least squares smoothing. Analyst 140(1):250–257. https://doi.org/10.1039/c4an01061b

125. Chen L, Wu Y, Li T, Chen Z (2018) Collaborative penalized least squares for background correction of multiple Raman spectra. J Anal Methods Chem 2018:9031356. https://doi.org/10.1155/2018/9031356

126. Pérez Y, Casado M, Raldúa D et al (2020) Mcr-als analysis of (1)h NMR spectra by segments to study the zebrafish exposure to acrylamide. Anal Bioanal Chem 412(23):5695–5706. https://doi.org/10.1007/s00216-020-02789-0

127. Felten J, Hall H, Jaumot J, Tauler R, de Juan A, Gorzsás A (2015) Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (MCR-ALS). Nat Protoc 10(2):217–240. https://doi.org/10.1038/nprot.2015.008

128. Smith JP, Holahan EC, Smith FC, Marrero V, Booksh KS (2019) A novel multivariate curve resolution-alternating least squares (MCR-ALS) methodology for application in hyperspectral raman imaging analysis. Analyst 144(18):5425–5438. https://doi.org/10.1039/C9AN00787C

129. Nagai Y, Sohn WY, Katayama K (2019) An initial estimation method using cosine similarity for multivariate curve resolution: application to NMR spectra of chemical mixtures. Analyst 144:5986

130. Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. Philos Magaz J Sci 2(11):559–572

131. Cserháti T (2010) Data evaluation in chromatography by principal component analysis. Biomed Chromatogr 24(1):20–28. https://doi.org/10.1002/bmc.1294

132. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2(1–3):37–52

133. Pang T, Zhang H, Wen L et al (2021) Quantitative analysis of a weak correlation between complicated data on the basis of principal

component analysis. J Anal Methods Chem 2021:8874827. https://doi.org/10.1155/2021/8874827

134. Soares EJ, Clifford AJ, Brown CD, Dean RR, Hupp AM (2019) Balancing resolution with analysis time for biodiesel–diesel fuel separations using GC, PCA, and the mahalanobis distance. Separations 6(2):28

135. Sudol PE, Gough DV, Prebihalo SE, Synovec RE (2020) Impact of data bin size on the classification of diesel fuels using comprehensive two-dimensional gas chromatography with principal component analysis. Talanta 206:120239

136. Cook DW, Rutan SC (2014) Chemometrics for the analysis of chromatographic data in metabolomics investigations. J Chemom 28:681–987

137. Smilde AK, Bro R, Geladi P (2004) Multi-way analysis with applications in the chemical sciences. John Wiley & Sons

138. Khakimov B, Amigo JM, Bak S, Engelsen SB (2012) Plant metabolomics: resolution and quantification of elusive peaks in liquid chromatography–mass spectrometry profiles of complex plant extracts using multi-way decomposition methods. J Chromatogr A 1266:84–94. https://doi.org/10.1016/j.chroma.2012.10.023

139. Kumar K (2019) Optimizing parallel factor (PARAFAC) assisted excitation-emission matrix fluorescence (EEMF) spectroscopic analysis of multifluorophoric mixtures. J Fluoresc 29(3):683–691. https://doi.org/10.1007/s10895-019-02379-z

140. Kumar K, Kumar Mishra A (2015) Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation–emission matrix fluorescence (EEMF) layout: certain practical aspects. Chemom Intell Lab Syst 147:121–130. https://doi.org/10.1016/j.chemolab.2015.08.008

141. Dramićanin T, Zeković I, Periša J, Dramićanin MD (2019) The parallel factor analysis of beer fluorescence. J Fluoresc 29(5):1103–1111. https://doi.org/10.1007/s10895-019-02421-0

142. Peng N, Wang K, Tu N, Liu Y, Li Z (2020) Fluorescence regional integration combined with parallel factor analysis to quantify fluorescent spectra for dissolved organic matter released from manure biochars. RSC Adv 10(52):31502–31510. https://doi.org/10.1039/D0RA02706E

143. Barker M, Rayens W (2003) Partial least squares for discrimination. J Chemom 17(3):166–173. https://doi.org/10.1002/cem.785

144. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G (2020) So you think you can pls-da? BMC Bioinf 21(1):2. https://doi.org/10.1186/s12859-019-3310-7

145. Bayci AWL, Baker DA, Somerset AE et al (2018) Metabolomic identification of diagnostic serum-based biomarkers for advanced stage melanoma. Metabolomics 14(8):105. https://doi.org/10.1007/s11306-018-1398-9

146. Dittgen CL, Hoffmann JF, Chaves FC, Rombaldi CV, Filho JMC, Vanier NL (2019) Discrimination of genotype and geographical origin of black rice grown in brazil by LC-MS analysis of phenolics. Food Chem 288:297–305. https://doi.org/10.1016/j.foodchem.2019.03.006

147. Tsanaktsidou E, Karavasili C, Zacharis CK, Fatouros DG, Markopoulou CK (2020) Partial least square model (PLS) as a tool to predict the diffusion of steroids across artificial membranes. Molecules. https://doi.org/10.3390/molecules25061387

148. Wang X, Wang P, Zhang A, Sun H (2015) Chapter 11 - metabolic profiling and potential biomarkers of Shenyinxu syndrome and the therapeutic effect of liuweidihuang wan. In: Wang X, Zhang A, Sun H (eds) Chinmedomics. Academic Press, Boston, pp 175–194

149. Zhang J, Yu Q, Cheng H, Ge Y-Q, Liu H, Ye X, Chen Y (2018) Metabolomic approach for the authentication of berry fruit juice by liquid chromatography quadrupole time-of-flight mass spectrometry coupled to chemometrics. J Agric Food Chem 66(30):8199–8208

150. Cao S, Du H, Tang B, Xi C, Chen Z (2021) Non-target metabolomics based on high-resolution mass spectrometry combined with chemometric analysis for discriminating geographical origins of Rhizoma Coptidis. Microchem J 160:105685. https://doi.org/10.1016/j.microc.2020.105685

151. Pisner DA, Schnyer DM (2020) Chapter 6 - support vector machine. In: Mechelli A, Vieira S (eds) Machine learning. Academic Press, pp 101–121

152. Zheng S (2016) Smoothly approximated support vector domain description. Pattern Recogn 49:55–64. https://doi.org/10.1016/j.patcog.2015.07.003

153. Képeš E, Vrábel J, Adamovsky O, Střítežská S, Modlitbová P, Pořízka P, Kaiser J (2022) Interpreting support vector machines applied in laser-induced breakdown spectroscopy. Anal Chim Acta 1192:339352. https://doi.org/10.1016/j.aca.2021.339352

154. Khanmohammadi Khorrami M et al (2021) Genetic algorithm based support vector machine regression for prediction of sara analysis in crude oil samples using ATR-FTIR spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 245:118945. https://doi.org/10.1016/j.saa.2020.118945

155. Hwang H, Jeong HK, Lee HK et al (2020) Machine learning classifies core and outer fucosylation of n-glycoproteins using mass spectrometry. Sci Rep 10(1):318. https://doi.org/10.1038/s41598-019-57274-1

156. Usman AG et al (2020) Artificial intelligence-based models for the qualitative and quantitative prediction of a phytochemical compound using hplc method. Turk J Chem 44(5):1339–1351. https://doi.org/10.3906/kim-2003-6

157. Mendez KM, Broadhurst DI, Reinke SN (2020) Migrating from partial least squares discriminant analysis to artificial neural networks: a comparison of functionally equivalent visualisation and feature contribution tools using jupyter notebooks. Metabolomics 16(2):17. https://doi.org/10.1007/s11306-020-1640-0

158. D'Archivio AA, Incani A, Ruggieri F (2011) Retention modelling of polychlorinated biphenyls in comprehensive two-dimensional gas chromatography. Anal Bioanal Chem 399:903–913

159. Malenović A, Jančić-Stojanović BS, Kostić N, Ivanović D, Medenica M (2011) Optimization of artificial neural networks for modeling of atorvastatin and its impurities retention in micellar liquid chromatography. Chromatographia 73:993–998

160. Xu Y, Chen J, Yang D, Hu Y et al (2021) Development of lc-ms/ms determination method and backpropagation artificial neural networks pharmacokinetic model of febuxostat in healthy subjects. J Clin Pharm Ther 46(2):333–342. https://doi.org/10.1111/jcpt.13285

161. Huang S, Liu Y, Sun X, Li J (2021) Application of artificial neural network based on traditional detection and GC-MS in prediction of free radicals in thermal oxidation of vegetable oil. Molecules. https://doi.org/10.3390/molecules26216717

162. Mert Ozupek N, Cavas L (2021) Modelling of multilinear gradient retention time of bio-sweetener rebaudioside a in HPLC analysis. Anal Biochem 627:114248. https://doi.org/10.1016/j.ab.2021.114248

163. Dinç E, Yücesoy C, Onur F (2002) Simultaneous spectrophotometric determination of mefenamic acid and paracetamol in a pharmaceutical preparation using ratio spectra derivative spectrophotometry and chemometric methods. J Pharm Biomed Anal 28(6):1091–1100. https://doi.org/10.1016/S0731-7085(02)00031-6

164. Dinç E (2003) Linear regression analysis and its application to the multivariate spectral calibrations for the multiresolution of a ternary mixture of caffeine, paracetamol and metamizol in tablets. J Pharm Biomed Anal 33(4):605–615. https://doi.org/10.1016/s0731-7085(03)00260-7

165. Niazi A, Goodarzi M (2008) Orthogonal signal correction-partial least squares method for simultaneous spectrophotometric determination of cypermethrin and tetramethrin. Spectrochim Acta A Mol Biomol Spectrosc 69(4):1165–1169. https://doi.org/10.1016/j.saa.2007.06.017

166. Dinç E, Baydan E, Kanbur M, Onur F (2002) Spectrophotometric multi-component determination of sunset yellow, tartrazine and allura red in soft drink powder by double divisor-ratio spectra derivative, inverse least-squares and principal component regression methods. Talanta 58(3):579–594. https://doi.org/10.1016/S0039-9140(02)00320-X

167. Chen K, Park J, Li F, Patil SM, Keire DA (2018) Chemometric methods to quantify 1d and 2d NMR spectral differences among similar protein therapeutics. AAPS PharmSciTech 19(3):1011–1019. https://doi.org/10.1208/s12249-017-0911-1

168. Wasim M, Brereton RG (2005) Application of multivariate curve resolution methods to on-flow LC-NMR. J Chromatogr A 1096(1–2):2–15. https://doi.org/10.1016/j.chroma.2005.05.101

169. Gargallo R, Tauler R, Cuesta-Sánchez F, Massart DL (1996) Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation. TrAC Trends Anal Chem 15(7):279–286. https://doi.org/10.1016/0165-9936(96)00048-9

170. Peré-Trepat E, Lacorte S, Tauler R (2005) Solving liquid chromatography mass spectrometry coelution problems in the analysis of environmental samples by multivariate curve resolution. J Chromatogr A 1096(1):111–122. https://doi.org/10.1016/j.chroma.2005.04.089

171. Bylund D, Danielsson R, Malmquist G, Markides KE (2002) Chromatographic alignment by warping and dynamic programming as a preprocessing tool for parafac modelling of liquid chromatography–mass spectrometry data. J Chromatogr A 961(2):237–244. https://doi.org/10.1016/S0021-9673(02)00588-5

172. Deng X, Liao Q, Xu X et al (2014) Analysis of essential oils from cassia bark and cassia twig samples by GC-MS combined with multivariate data analysis. Food Anal Methods 7(9):1840–1847. https://doi.org/10.1007/s12161-014-9821-y

173. Fraga CG (2003) Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography–selected-ion mass spectrometry data. J Chromatogr A 1019(1–2):31–42. https://doi.org/10.1016/s0021-9673(03)01329-3

174. Lukitaningsih E et al (2012) Quantitative analysis of lard in cosmetic lotion formulation using FTIR spectroscopy and partial least square calibration. J Am Oil Chem Soc 89(8):1537–1543. https://doi.org/10.1007/s11746-012-2052-8

175. Rohman A, Che Man YB (2011) Application of fourier transform infrared (FT-IR) spectroscopy combined with chemometrics for authentication of cod-liver oil. Vib Spectrosc 55(2):141–145. https://doi.org/10.1016/j.vibspec.2010.10.001

176. Yan F, Liang Z, Jianna C, Zhengtao W, Losahan X, Zhengxing Z (2001) Analysis of Cnidium Monnieri fruits in different regions of china. Talanta 53(6):1155–1162. https://doi.org/10.1016/S0039-9140(00)00594-4

177. Comas E, Gimeno RA, Ferré J, Marcé RM, Borrull F, Rius FX (2004) Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods. J Chromatogr A 1035(2):195–202. https://doi.org/10.1016/j.chroma.2004.02.069

178. Detroyer A, Schoonjans V, Questier F, Vander Heyden Y, Borosy AP, Guo Q, Massart DL (2000) Exploratory chemometric analysis of the classification of pharmaceutical substances based on chromatographic data. J Chromatogr A 897(1):23–36. https://doi.org/10.1016/S0021-9673(00)00803-7

179. Lu X-F, Bi K-S, Zhao X, Chen X-H (2012) Authentication and distinction of shenmai injection with hplc fingerprint analysis assisted by pattern recognition techniques. J Pharm Anal 2(5):327–333. https://doi.org/10.1016/j.jpha.2012.07.009

180. Abdollahi H, Nazari F (2003) Rank annihilation factor analysis for spectrophotometric study of complex formation equilibria. Analytica Chimica Acta ANAL CHIM ACTA 486:109–123. https://doi.org/10.1016/S0003-2670(03)00471-9

181. Kong W-J, Zhao Y-L, Xiao X-H, Jin C, Li Z-L (2009) Quantitative and chemical fingerprint analysis for quality control of Rhizoma Coptidis-chinensis based on UPLC-pad combined with Chemometrics methods. Phytomedicine 16(10):950–959. https://doi.org/10.1016/j.phymed.2009.03.016

182. O'Connell M-L, Ryder AG, Leger MN, Howley T (2010) Qualitative analysis using raman spectroscopy and chemometrics: A comprehensive model system for narcotics analysis. Appl Spectrosc 64(10):1109–1121

183. Porfire A, Tomuta I, Tefas L, Leucuta SE, Achim M (2012) Simultaneous quantification of l-α-phosphatidylcholine and cholesterol in liposomes using near infrared spectrometry and chemometry. J Pharm Biomed Anal 63:87–94. https://doi.org/10.1016/j.jpba.2012.01.017

184. Candolfi A, De Maesschalck R, Massart DL, Hailey PA, Harrington AC (1999) Identification of pharmaceutical excipients using NIR spectroscopy and simca. J Pharm Biomed Anal 19(6):923–935. https://doi.org/10.1016/s0731-7085(98)00234-9

185. Rodríguez Cáceres MI, Durán Merás I, Ornelas Soto NE, López de Alba PL, López Martinez L (2008) Determination of anticarcinogenic and rescue therapy drugs in urine by photoinduced spectrofluorimetry using multivariate calibration: comparison of several second-order methods. Anal Bioanal Chem 391(4):1119–1127. https://doi.org/10.1007/s00216-008-2069-x

186. Ter Horst JP, Turimella SL, Metsers F, Zwiers A (2021) Implementation of quality by design (QBD) principles in regulatory dossiers of medicinal products in the European Union (EU) between 2014 and 2019. Ther Innov Regul Sci 55(3):583–590. https://doi.org/10.1007/s43441-020-00254-9

187. Gurrala S, Raj S, Cvs S, Anumolu PD (2022) Quality-by-design approach for chromatographic analysis of metformin, Empagliflozin and Linagliptin. J Chromatogr Sci 60(1):68–80

188. Zacharis CK, Vastardi E (2018) Application of analytical quality by design principles for the determination of alkyl p-toluenesulfonates impurities in aprepitant by HPLC. Validation using total-error concept. J Pharm Biomed Anal 150:152–161

189. Almeida J, Bezerra M, Markl D, Berghaus A, Borman P, Schlindwein W (2020) Development and validation of an in-line API quantification method using AQBD principles based on UV-vis spectroscopy to monitor and optimise continuous hot melt extrusion process. Pharmaceutics 12(2):150

190. Žigart N, Časar Z (2020) Development of a stability-indicating analytical method for determination of venetoclax using AQBD principles. ACS Omega 5(28):17726–17742

191. Bandopadhyay S, Beg S, Katare O, Sharma T, Singh B (2020) Integrated analytical quality by design (AQBD) approach for the development and validation of bioanalytical liquid chromatography method for estimation of valsartan. J Chromatogr Sci 58(7):606–621

192. Bossunia MTI, Urmi KF, Chironjit Kumar S (2017) Quality-by-design approach to stability indicating RP-HPLC analytical method development for estimation of canagliflozin API and its validation. Pharm Methods 8:2

193. Peraman R, Bhadraya K, Reddy YP, Reddy CS, Lokesh T (2015) Analytical quality by design approach in RP-HPLC method development for the assay of etofenamate in dosage forms. Indian J Pharm Sci 77(6):751–757. https://doi.org/10.4103/0250-474x.174971

194. Tumpa A, Stajić A, Jančić-Stojanović B, Medenica M (2017) Quality by design in the development of hydrophilic interaction liquid chromatography method with gradient elution for the analysis of olanzapine. J Pharm Biomed Anal 134:18–26

195. Peraman R, Kalva B, Shanka S, Reddy YP (2014) Analytical quality by design (AQBD) approach to liquid chromatographic method for quantification of acyclovir and hydrocortisone in dosage forms. Anal Chem Lett 4(5–6):329–342

196. Peraman R, Bandi J, Kondreddy VK et al (2021) Analytical quality by design approach versus conventional approach: development of HPLC-dad method for simultaneous determination of etizolam and propranolol hydrochloride. J Liquid Chromatogr Relat Technol 44(3–4):197–209

197. Palakurthi AK, Dongala T, Katakam LNR (2020) QBD based development of HPLC method for simultaneous quantification of telmisartan and hydrochlorothiazide impurities in tablets dosage form. Pract Lab Med 21:e00169

198. Krishna MV, Dash RN, Reddy BJ, Venugopal P, Sandeep P, Madhavi G (2016) Quality by design (QBD) approach to develop HPLC method for eberconazole nitrate: application oxidative and photolytic degradation kinetics. J Saudi Chem Soc 20:S313–S322

199. Thakur D, Jain A, Ghoshal G, Shivhare U, Katare O (2017) RP-HPLC method development using analytical QBD approach for estimation of cyanidin-3-o-glucoside in natural biopolymer based microcapsules and tablet dosage form. J Pharm Investig 47(5):413–427

200. Sandhu PS, Beg S, Kumar R, Katare O, Singh B (2017) Analytical QBD-based systematic bioanalytical HPLC method development for estimation of quercetin dihydrate. J Liq Chromatogr Relat Technol 40(10):506–516

201. Alruwaili NK (2021) Analytical quality by design approach of reverse-phase high-performance liquid chromatography of atorvastatin: method development, optimization, validation, and the stability-indicated method. Int J Anal Chem. https://doi.org/10.1155/2021/8833900

202. Fink DW (1988) Ivermectin analytical profiles of drug substances. Elsevier, pp 155–184

203. Abdel-Moety EM, Al-Khamees HA (1990) Analytical profile of azintamide. In: Analytical profiles of drug substances, vol 18, pp 1–32. Elsevier

204. Zubair MU, Hassan MM, Al-Meshal IA (1986) Caffeine. In: Analytical profiles of drug substances, vol 15, pp 71–150. Elsevier

205. Florey K (1979) Aspirin. In: Analytical profiles of drug substances, vol 8, pp 1–46. Elsevier

206. Townley ER (1979) Griseofulvin. In: Analytical profiles of drug substances, vol 8, pp 219–249. Elsevier

207. Pitré D (1986) Iodamide. In: Analytical profiles of drug substances, vol 15, pp 337–365. Elsevier.

208. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B et al (2009) Hmdb: a knowledgebase for the human metabolome. Nucl Acids Res 37(suppl_1):D603–D610

Malarvannan *et al. Future Journal of Pharmaceutical Sciences*   (2023) 9:86

Page 40 of 42

209. SDBSWeb (2014) Sdbsweb: National institute of advanced industrial science and technology. Accessed May 2021 from http://sdbs.Riodb.Aist.Go.Jp

210. Hassan MM, Elazzouny AA (1982) Clofibrate. In: Analytical profiles of drug substances, vol 11, pp 197–224. Elsevier

211. Piskorik HG (1985) Tripelennamine hydrochloride. In: Analytical profiles of drug substances. vol 14, pp 107–133. Elsevier

212. Pitrè D (1985) Iopoanoic acid. In Analytical profiles of drug substances, vol. 14, pp 181–206. Elsevier

213. Brittain HG (2001) Malic acid. Profiles Drug Subst Excip Relat Methodol 28:153–195

214. Al-Badr AA, El-Obeid HA (1988) Analytical profile of primidone. In: Analytical profiles of drug substances, vol 17, pp 749–795. Elsevier

215. Fairbrother JE (1974) Acetaminophen. In: Analytical profiles of drug substances, vol 3, pp 1–109. Elsevier

216. Ali SL (1983) Benzocaine. In: Analytical profiles of drug substances, vol 12, pp 73–104. Elsevier

217. Eisner U, Gore PH (1958) The light absorption of pyrroles. Part i. Ultraviolet spectra. J Chem Soc 186:922–927. https://doi.org/10.1039/JR9580000922

218. Filpponen I, Sadeghifar H, Argyropoulos DS (2011) Photoresponsive cellulose nanocrystals. Nanomater Nanotechnol 1:7. https://doi.org/10.5772/50949

219. Arayne MS, Sultana N, Zuberi MH, Siddiqui FA (2009) Spectrophotometric quantitation of metformin in bulk drug and pharmaceutical formulations using multivariate technique. Indian J Pharm Sci 71(3):331–335. https://doi.org/10.4103/0250-474x.56022

220. Dunn DL, Jones WJ, Dorsey ED (1983) Analysis of chlorobutanol in ophthalmie ointments and aqueous solutions by reverse-phase high-performance liquid chromatography. J Pharm Sci 72(3):277–280. https://doi.org/10.1002/jps.2600720317

221. Pubchem (2022) Pubchem database. Accessed May 2021 from https://pubchem.Ncbi.Nlm.Nih.Gov.

222. Fairbrother JE (1973) Chloral hydrate. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 85–143

223. Daley RD (1972) Halothane. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 119–147

224. Orzech CE, Nash NG, Daley RD (1979) Halcinonide. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 283–314

225. Stober HC (1986) Lithium carbonate. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 367–391

226. Chao MKC, Albert KS, Fusari SA (1978) Phenobarbital. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 359–399

227. Newman AW, Vitez IM, Mueller RL, Kiesnowski CC et al (1999) Sorbitol. In: Brittain HG (ed) Analytical profiles of drug substances and excipients. Academic Press, pp 459–502

228. Manius GJ (1978) Trimethoprim. In: Florey K (ed) Analytical profiles of drug substances. Academic Press, pp 445–475

229. Coates J (2000) Interpretation of infrared spectra: a practical approach. John Wiley & Sons Ltd, Chichester

230. Vilas S, Thilagar S (2021) Formulation and optimisation of lamivudine-loaded eudragit(®) s 100 polymer-coated pectin microspheres for colon-specific delivery. IET Nanobiotechnol 15(1):90–99. https://doi.org/10.1049/nbt2.12010

231. Bansal R, Guleria A, Acharya PC (2013) FT-IR method development and validation for quantitative estimation of zidovudine in bulk and tablet dosage form. Drug Res 63(4):165–170. https://doi.org/10.1055/s-0032-1333297

232. Anbazhagan S, Indumathy N, Shanmugapandiyan P, Sridhar SK (2005) Simultaneous quantification of stavudine, lamivudine and nevirapine by UV spectroscopy, reverse phase HPLC and HPTLC in tablets. J Pharm Biomed Anal 39(3–4):801–804

233. Chandramowli B, Kumar BS, Bhikshapathi D, Rajkamal BB (2018) A new quantitative analytical method development and validation for the analysis of boceprevir in bulk and marketed formulation. Int J Pharm Sci Drug Res 10(3):201–205

234. Snehal DJ, Prafulla SC, Vishal SM (2018) HPLC method development of cidofovir as bulk drug and its formulation. Int J Eng Devel Res 6(2):870–874

235. Abdul K, Milon M, Mohammad A, Parvin M, Rafiquzzaman M, Kundu S (2017) Development and validation of a new RP-HPLC method for the determination of daclatasvir dihydrochloride in bulk and pharmaceutical dosage forms. Int J Pharm 7:7–13

236. Satyanarayana L, Naidu S, Rao MN, Ayyanna C, Kumar A (2011) The estimation of RALTIGRAVIR in tablet dosage form by RP-HPLC. Asian J Pharm Anal 1(3):56–58

237. Tiruveedhi VBG, Battula VR, Bonige KB (2021) Rp-hplc (stability-indicating) based assay method for the simultaneous estimation of doravirine, tenofovir disoproxil fumarate and lamivudine. Int J Appl Pharm 13(1):153–159

238. Tejaswi KD et al (2019) Reverse-phase high-performance liquid chromatography method development and validation for simultaneous estimation and forced degradation studies of Emtricitabine, Rilpivirine, and Tenofovir Alafenamide in solid dosage form. Asian J Pharm Clin Res 12:112. https://doi.org/10.22159/ajpcr.2018.v12i1.28765

239. Mastanamma S, Chandini S, Reehana S, Saidulu P (2018) Development and validation of stability indicating RP-HPLC method for the simultaneous estimation of Sofosbuvir and Ledipasvir in bulk and their combined dosage form. Future J Pharm Sci 4(2):116–123

240. Dey S, Patro SS, Babu NS, Murthy P, Panda S (2017) Development and validation of a stability-indicating RP–HPLC method for estimation of atazanavir sulfate in bulk. J Pharm Anal 7(2):134–140

241. Rathnasamy R, Karuvalam R, Pakkath R, Kamalakannan P, Sivasubramanian A (2018) RP-HPLC method development and method validation of lopinavir and ritonavir in pharmaceutical dosage form. Am J Clin Microbiol Antimicrob 1(1):1002

242. Ezzeldin E, Abo-Talib NF, Tammam MH, Asiri YA, Amr AE-GE, Almehizia AA (2020) Validated reversed-phase liquid chromatographic method with gradient elution for simultaneous determination of the antiviral agents: sofosbuvir, ledipasvir, daclatasvir, and simeprevir in their dosage forms. Molecules 25(20):4611

243. Ahmad SAR, Patil L, Usman MRM, Imran M, Akhtar R (2018) Analytical method development and validation for the simultaneous estimation of abacavir and lamivudine by reversed-phase high-performance liquid chromatography in bulk and tablet dosage forms. Pharm Res 10(1):92. https://doi.org/10.4103/pr.pr_96_17

244. Haneef M, Rajkamal B, Goud VM (2013) Development and validation by RP-HPLC method for estimation of zidovudine in bulk and its pharmaceutical dosage form. Asian J Res Chem 6(4):341–344

245. Godela R, Gummadi S (2021) A simple stability indicating RP-HPLC-dad method for concurrent analysis of tenofovir disoproxil fumarate, doravirine and lamivudine in pure blend and their combined film coated tablets. Paper presented at the Annales Pharmaceutiques Françaises.

246. Harikrishnan N, Prasad VV et al (2019) Stability indicating RP-HPLC method development and validation for the simultaneous estimation of pibrentasvir and glecaprevir in bulk and pharmaceutical dosage form. Int J Res Pharm Sci 10(3):1841–1846

247. Ramya V et al (2019) Simultaneous estimation of amantadine hydrochloride and oseltamivir phosphate using precolumn derivatization technique. Int J Pharm Sci Res 10(12):5443–5449. https://doi.org/10.13040/IJPSR.0975-8232.10(12).5443-49

248. García J, Márquez A, Ruiz R, López LF, Claro C, Lucero MJ (2006) Determination of foscarnet (trisodium phosphonoformate) in pharmaceutical preparations by high-performance liquid chromatography with ultraviolet detection. Biomed Chromatogr 20(10):1024–1027

249. Ramesh P, Basavaiah K, Vinay K, Xavier CM (2012) Development and validation of Rp-HPLC method for the determination of ganciclovir in bulk drug and in formulations. Int Schol Res Not. https://doi.org/10.5402/2012/894965

250. Nekkala K, Kumar JS, Ramachandran D (2019) Stability indicating RP-HPLC method for quantification of impurities in fosamprenavir calcium drug substance. J Pharm Sci Res 11(3):712–717

251. Bulduk İ (2021) HPLC-UV method for quantification of favipiravir in pharmaceutical formulations. Acta Chromatogr 33(3):209–215

252. Asha E, Surendra Babu K (2020) A new selective separation method development and validation of trifluridine and tipiracil and its degradents were characterized by LC-MS/MS/QTOF. J Pharm Sci Res 12(1):199–205

253. Reddy Y, Harika K, Sowjanya K, Swathi E, Soujanya B, Reddy S (2011) Estimation of zanamivir drug present in tablets using RP-HPLC method. Int J PharmTech Res 3:180–186

254. Dalmora SL, Sangoi M et al (2010) Validation of a stability-indicating RP-HPLC method for the determination of entecavir in tablet dosage form. J AOAC Int 93(2):523–530

255. Prakash KV, Rao JV, Raju NA (2007) RP-HPLC method for the estimation of nelfinavir mesylate in tablet dosage form. E J Chem 4(3):302–306

256. Jing Q, Shen Y, Tang Y, Ren F, Yu X, Hou Z (2006) Determination of nelfinavir mesylate as bulk drug and in pharmaceutical dosage form by stability indicating HPLC. J Pharm Biomed Anal 41(3):1065–1069

257. Attaluri T, Seru G, Varanasi SNM (2021) Development and validation of a stability-indicating RP-HPLC method for the simultaneous estimation of bictegravir, emtricitabine, and tenofovir alafenamide fumarate. Turk J Pharm Sci 18(4):410–419. https://doi.org/10.4274/tjps.galenos.2020.70962

258. Jitta SR, Salwa K et al (2021) Development and validation of high-performance liquid chromatography method for the quantification of remdesivir in intravenous dosage form. Assay Drug Dev Technol 19(8):475–483. https://doi.org/10.1089/adt.2021.074

259. Shuangjin C, Fang F, Han L, Ming M (2007) New method for high-performance liquid chromatographic determination of amantadine and its analogues in rat plasma. J Pharm Biomed Anal 44(5):1100–1105

260. Kaliszan R (2007) Qsrr: Quantitative structure-(chromatographic) retention relationships. Chem Rev 107(7):3212–3246

261. Venkat Kumar C, Anantahakumar D, Rao J (2010) A new validated rp-hplc method for determination of penciclovir in human plasma. Int J Chem Sci 2:95–102

262. Raees Ahmad SA, Patil L, Mohammed Usman MR, Imran M, Akhtar R (2018) Analytical method development and validation for the simultaneous estimation of abacavir and lamivudine by reversed-phase high-performance liquid chromatography in bulk and tablet dosage forms. Pharmacognosy Res 10(1):92–97. https://doi.org/10.4103/pr.pr_96_17

263. Hussain Shah SS, Nasiri MI, Sarwar H, Ali A et al (2021) RP-HPLC method development and validation for quantification of daclatasvir dihydrochloride and its application to pharmaceutical dosage form. Pak J Pharm Sci 34(3):951–956

264. Hiremath SN, Bhirud CH (2015) Development and validation of a stability indicating HPLC method for the simultaneous analysis of lopinavir and ritonavir in fixed-dose combination tablets. J Taib Univ Med Sci 10(3):271–277. https://doi.org/10.1016/j.jtumed.2014.11.006

265. dos Santos JV, de Carvalho LA, Pina ME (2011) Development and validation of a RP-HPLC method for the determination of zidovudine and its related substances in sustained-release tablets. Anal Sci 27(3):283–289. https://doi.org/10.2116/analsci.27.283

266. Higashi Y, Uemori I, Fujii Y (2005) Simultaneous determination of amantadine and rimantadine by hplc in rat plasma with pre-column derivatization and fluorescence detection for pharmacokinetic studies. Biomed Chromatogr 19(9):655–662. https://doi.org/10.1002/bmc.492

267. Mamatha J, Devanna N (2017) Development and validation of a rp-hplc method for analysis of cidofovir in medicinal form. Indian J Sci Technol 10:1–5. https://doi.org/10.17485/ijst/2017/v10i34/117108

268. Patel J, Bedi H, Middha A, Prajapati L, Parmar V (2012) Rp-hplc method for estimation of nitazoxanide in oral suspension formulation. Der Pharm Chem 4:1140–1144

269. Kanagala V, Anusha M, Reddy S (2015) Rapid rp-hplc method development and validation for analysis of raltegravir in bulk and pharmaceutical dosage form. Asian J Pharm Anal. https://doi.org/10.5958/2231-5675.2015.00003.4

270. Babu G, Atmakuri LR, Rao J (2015) A rapid RP-HPLC method development and validation for the quantitative estimation ribavirin in tablets. Int J Pharm Pharm Sci 7:60–63

271. Chandrasekaran B, Abed SN, Al-Attraqchi O, Kuche K, Tekade RK (2018) Chapter 21 - computer-aided prediction of pharmacokinetic (admet) properties. In: Tekade RK (ed) Dosage form design parameters. Academic Press, pp 731–755

272. Wu Y-J, Zhan T, Hou Z, Fang L, Xu Y (2020) Physical and chemical descriptors for predicting interfacial thermal resistance. Sci Data 7:36. https://doi.org/10.1038/s41597-020-0373-2

273. Zhang C, Idelbayev Y, Roberts N, Tao Y et al (2017) Small molecule accurate recognition technology (smart) to enhance natural products research. Sci Rep 7(1):14243–14243. https://doi.org/10.1038/s41598-017-13923-x

274. Stagner WC, Haware RV (2019) Qbd innovation through advances in pat, data analysis methodologies, and material characterization. AAPS PharmSciTech 20(7):295. https://doi.org/10.1208/s12249-019-1506-9

275. Skoog DA et al (2014) Fundamentals of analytical chemistry, 9th edn. Brooks/Cole, Belmont CA

276. Booksh KS, Kowalski BR (1994) Theory of analytical chemistry. Anal Chem 66(15):782A-791A. https://doi.org/10.1021/ac00087a718

277. Wise BM, Gallagher NB (1996) The process chemometrics approach to process monitoring and fault detection. J Process Control 6(6):329–348

278. Fu H, Yin Q, Xu L, Wang W, Chen F, Yang T (2017) A comprehensive quality evaluation method by ft-nir spectroscopy and chemometric: fine classification and untargeted authentication against multiple frauds for chinese ganoderma lucidum. Spectrochimica acta. Part A Mol Biomol Spectrosc 182:17–25. https://doi.org/10.1016/j.saa.2017.03.074

279. Rady AM, Guyer DE (2015) Evaluation of sugar content in potatoes using nir reflectance and wavelength selection techniques. Postharvest Biol Technol 103:17–26. https://doi.org/10.1016/j.postharvbio.2015.02.012

280. Hong X, Wang J, Qi G (2015) E-nose combined with chemometrics to trace tomato-juice quality. J Food Eng 149:38–43. https://doi.org/10.1016/j.jfoodeng.2014.10.003

281. Canizo BV, Escudero LB, Pérez MB, Pellerano RG, Wuilloud RG (2017) Intra-regional classification of grape seeds produced in mendoza province (argentina) by multi-elemental analysis and chemometrics tools. Food Chem 242:272–278

282. Liu W, Liu C, Hu X, Yang J, Zheng L (2016) Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics. Food Chem 210:415–421. https://doi.org/10.1016/j.foodchem.2016.04.117

283. Barbosa RM, de Paula ES, Paulelli AC, Moore AF, Souza JMO et al (2016) Recognition of organic rice samples based on trace elements and support vector machines. J Food Compos Anal 45:95–100. https://doi.org/10.1016/j.jfca.2015.09.010

284. Zheng W, Fu X, Ying Y (2014) Spectroscopy-based food classification with extreme learning machine. Chemom Intell Lab Syst 139:42–47. https://doi.org/10.1016/j.chemolab.2014.09.015

285. PK H (1985) Receptor modling in environmental chemistry. John Wiley Sons, New York

286. Golubović J, Protić A, Otašević B, Zečević M (2016) Quantitative structure-retention relationships applied to development of liquid chromatography gradient-elution method for the separation of sartans. Talanta 150:190–197. https://doi.org/10.1016/j.talanta.2015.12.035

287. Khosrokhavar R, Ghasemi JB, Shiri F (2010) 2d quantitative structure-property relationship study of mycotoxins by multiple linear regression and support vector machine. Int J Mol Sci 11(9):3052–3068. https://doi.org/10.3390/ijms11093052

288. Fu T, Sieng IDY (2021) A comparative study between pcr, plsr, and lw-pls on the predictive performance at different data splitting ratios. Chem Eng Commun. https://doi.org/10.1080/00986445.2021.1957853

289. El-Gindy A, Emara S, Mostafa A (2005) Hplc and chemometric-assisted spectrophotometric methods for simultaneous determination of atenolol, amiloride hydrochloride and chlorthalidone. Il Farmaco 60(3):269–278. https://doi.org/10.1016/j.farmac.2004.11.013

290. Luis ML, Fraga JMG, Jiménez F, Jiménez AI, Arias JJ (2001) Simultaneous spectrophotometric determination of diuretics by using multivariate calibration methods. Talanta 53(4):761–770. https://doi.org/10.1016/S0039-9140(00)00538-5

291. Cramer RD (1993) Partial least squares (pls): its strengths and limitations. Perspect Drug Discov Des 1(2):269–278. https://doi.org/10.1007/BF0217 4528

292. El-Gindy A, Emara S, Mostafa A (2006) Application and validation of chemometrics-assisted spectrophotometry and liquid chromatography for the simultaneous determination of six-component pharmaceuticals. J Pharm Biomed Anal 41(2):421–430. https://doi.org/10.1016/j.jpba.2005.12.005

293. Koleini F, Balsini P, Parastar H (2022) Evaluation of partial least-squares regression with multivariate analytical figures of merit for determination of 10 pesticides in milk. Int J Environ Anal Chem 102(8):1900–1910. https://doi.org/10.1080/03067319.2020.1745198

294. Pate ME, Turner MK, Thornhill NF, Titchener-Hooker NJ (1999) The use of principal component analysis for the modelling of high performance liquid chromatography. Bioprocess Eng 21(3):261–272. https://doi.org/10.1007/s004490050674
295. Singh VD, Daharwal SJ (2017) Development and validation of multivariate calibration methods for simultaneous estimation of paracetamol, enalapril maleate and hydrochlorothiazide in pharmaceutical dosage form. Spectrochim Acta A Mol Biomol Spectrosc 171:369–375. https://doi.org/10.1016/j.saa.2016.08.028

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.