Development of Deep Learning Software to Improve HPLC and GC Predictions using a new Crown-ether based Mesogenic Stationary Phase and Beyond

Warda Fella Belaid, Azeddine Dekhira, Philippe Lesot, Ouassila Ferroukhi

Development of Deep Learning Software to Improve HPLC and GC Predictions using a New Crown-Ether based Mesogenic Stationary Phase and Beyond

Warda Fella Belaid[a], Azeddine Dekhira[b], Philippe Lesot[c,d] and Ouassila Ferroukhi[a,*]

[a]Laboratory of Chromatography, Faculty of Chemistry, University of Sciences and Technology Houari Boumedienne, USTHB, B.P. 32 El-Alia, Bab-Ezzouar, Algiers.
16111, Algeria.

[b]Laboratory of Computational Theoretical Chemistry and Photonics, Faculty of Chemistry, University of Sciences and Technology Houari Boumedienne, USTHB, B.P. 32 El-Alia, Bab-Ezzouar, Algiers 16111, Algeria.

[c]Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO), UMR-CNRS 8182, Faculté des Sciences d'Orsay, Equipe RMN en Milieu Orienté, Université Paris-Saclay, Site Henri Moissan (HM-1), Bureau 0209, RDC, 17-19, Avenue des Sciences, Orsay 91400, France.

[d]Centre National de la Recherche Scientifique (CNRS), 3, Rue Michel Ange, Paris 75016, France.

*Corresponding author : Ouassila Ferroukhi : oferroukhi@usthb.dz

Abstract

*The application of AI to analytical and separative sciences is a recent challenge that offers new perspectives in terms of data prediction. In this work, we report an AI-based software, named **Chrompredict 1.0**, which based on chromatographic data of a novel mesogenic crown ether stationary phase (CESP). Its molecular design represents a significant advancement due to the unique combination of properties and binding capabilities, including the formation of a cavity, mesogenic behavior via mobile chains, and a range of polar and non-polar interactions (aromatic rings, N=N and C=O double bonds, alkyl chains, π–π interactions, and hydrogen bonding). The mesogenic phase is effective in both normal and reversed-phase chromatography, enhancing the software's adaptability across diverse datasets. Here we introduce for the first time an unprecedented scientific approach, integrating deep learning techniques with the novel CESP, which demonstrates exceptional thermal and analytical performance in both liquid chromatography modes, especially in the separation of complex hydrocarbon isomers. This ability enables the results obtained with CESP to extend across various types of stationary phases. Leveraging these insights, a comprehensive chromatographic dataset on a series of aromatic and polyaromatic molecules interacting with our CESP was used to train a Deep Learning Model (DLM). This model is embedded within a user-friendly software, **Chrompredict 1.0**, designed for predicting chromatographic parameters (MAE = 0.042, R^2 = 0.95) by selecting chemical descriptors directly from SMILES notation. It offers a deeper understanding of molecular structure and interactions through exploratory data analysis, identifying key factors affecting model accuracy and chromatographic behavior. Users can configure hyperparameters, choose from six machine learning models, and compare their performance with DLM. **Chrompredict 1.0** excels in retention behavior prediction for compounds with known structures, and it accurately predicts chromatographic retention and thermal characteristics for different temperatures in HPLC and GC. The model has been successfully tested with METLIN database of 1,023 small molecules of diverse structures and polarities (R^2 > 0.75, error range ±7.8 s). Overall, the CESP, combined with **Chrompredict 1.0**, offers a robust tool for intelligent chromatographic analysis, encompassing chemo-informatics, statistical analysis, and graphical capabilities across a broad range of compounds and stationary phases.*

Keywords: Liquid and gas chromatography, Crown ether, Mesogenic stationary phase, Deep learning software, Retention time prediction

1. Introduction

Chromatography is a key separation technique in analytical chemistry, which is widely used to identify, quantify, and separate compounds in complex mixtures across a range of applications, such as drug discovery, environmental analysis [1], and quality control [2]. However, some traditional chromatographic methods face the challenges of low separation efficiency, low-speed analysis time, and high solvent consumption [3]. To address these issues, Quantitative Structure-Retention Relationship (**QSRR**) models using Machine Learning (**ML**) algorithms to predict retention behavior based on their molecular structure have been recently developed [4,5,6], This approach requires fewer experimental runs, leading to significant solvent savings. While QSRR models have been valuable in understanding the molecular properties contributing to retention, they struggle to capture complex nonlinear relationships between molecular structures and retention behavior [7].

Advanced Deep Learning (**DL**) techniques have become increasingly important in the physical sciences research toolkit [8], including fields such as physics [9,10,11], chemistry [12,13,14], and material sciences [15,16,17,18,19,20]. This proliferation of machine learning methods, coupled with the growing accessibility of extensive datasets, has been acknowledged as the "Fourth Paradigm of Science" [21] and the "Fourth Industrial Revolution" [22]. These advancements hold great promise for substantially increasing the impact of computational approaches in both applied and fundamental research [23]. In this context, proposing new DL-based tools in the frame of chromatographic analysis appear to be an exciting challenge that deserves attention.

Traditional C18 columns, although widely used due to their reliable hydrophobic interactions, are often limited when it comes to separating complex mixtures, particularly those involving structurally similar compounds. These columns primarily rely on non-polar interactions, which can hinder their ability to achieve high-resolution separations in more intricate analyses [24]. This emphasizes the importance of thorough optimization in method development to enhance separation effectiveness. Incorporating liquid crystals into the stationary phase (SP) is known to improve the separation process by increasing resolution [3,25,26]. These materials exhibit unique physical and chemical properties, such as anisotropy and mesogenicity, and retain some ordered crystalline structure while flowing like liquids [27].

Due to their elongated molecular structure, liquid crystals enable for fine-tuning of separation selectivity and improved analytical efficiency [3]. They are particularly effective for separating narrow-boiling polycyclic aromatic hydrocarbons that are generally difficult to separate with conventional stationary phases [26,27]. Interestingly, liquid crystals are also

capable of molecular shape recognition, which is essential for the distinction between planar and non-planar solutes, thus improving chromatographic separation performance [28].

If the design of new stationary phase in chromatography is a noteworthy task, and necessary to extend the arsenal of chromatographic tools, it can be advantageously combined with DL techniques to robustly analyze and model series of experimental results from a collection of analytes. Indeed, during the last decade, the unprecedented growth of ML applications in analytical sciences has revolutionized the analysis of data [6], and several areas are taking full advantage of these computational approaches, including chromatography. These tools can be used, for example, to predict the retention time and retention factor for new chemical solutes.

For this purpose, we have designed and developed a new DL-based software named **Chrompredict 1.0** with a user-friendly interface. It addresses this gap by providing a powerful tool to model their experimental results for future predictions of retention time and retention factor for new chemical solutes. This software facilitates investigations related to thermal and separation performance of chromatography columns by offering a variety of useful options. They include, for instance, the injection temperatures, calculation of molecular descriptors, as well as providing other data analysis, graphical, and evaluation tools. The deep learning model implemented in this software was trained using experimental data (data collection), enabling accurate predictions of retention behavior for future experiments with the specified column. As we will see, this approach overcomes the limitations of traditional modeling techniques by capturing subtle patterns even when the datasets are relatively small and by understanding the thermal properties of chromatography columns. From practical point of view, the software's user-friendly interface ensures accessibility for researchers with diverse expertise levels in deep learning techniques, enabling them to leverage the methodology devised in this study and thereby enhances the quality of their chromatography experiments.

In this study, a Crown-Ether based Liquid Crystal (**CELC**) was synthesized and grafted onto a silica gel and its performances in HPLC have been explored on a series of 33 analytes. The experimental results obtained will be discussed in terms geometric parameters and chromatographic performances. In a second part, we will describe the integration of liquid crystal-based stationary phases with the aforementioned Graphical User Interface (**GUI**) based deep learning solution and the capability of this tool to predict retention and thermal behavior will be examined. This synergistic approach not only improves efficiency but

also expands the applicability of chromatography, rendering it a compelling choice for a wider range of applications.

2. Description of the Chrompredict 1.0 software

2.1. Deep learning software development tools

In this work, an extensive set of tools and libraries were employed for data separation and deep learning tasks. The primary programming language utilized was Python 3.11 in the Scientific Development Environment (Spyder 5.4.3, https://spyder-ide.org/). Key libraries imported for data handling, manipulation, and creating visually appealing and informative statistical graphics encompassed pandas (2.0.3, https://pandas.pydata.org/), NumPy (1.5, https://numpy.org/), matplotlib (https://matplotlib.org/) and seaborn (0.12.2, https://seaborn.pydata.org/). These four libraries played a crucial role in facilitating efficient data processing, analysis, and visualization throughout the study. The scikit-learn library (1.3.0, https://scikit-learn.org/) incorporates an extensive array of algorithms and tools, including RandomForest Regressor, ExtraTrees Regressor, Support Vector, Ridge, and Lasso Regression. These algorithms offer robust solutions for regression and prediction purposes. In order to evaluate the performance and accuracy of the models, several evaluation metrics including Mean Squared Error (**MSE**), Mean Absolute Error (**MAE**), Median Absolute Error, Explained Variance Score (**EVS**), and R-squared were employed. Furthermore, TensorFlow 2.9.1 (https://tensorflow.org/) enables the implementation of neural networks and advanced deep learning techniques. TensorBoard, a valuable companion tool, facilitates the visualization, analysis, and monitoring of TensorFlow models. It facilitates the tracking of training progress and the exploration of model performance and behavior. RDKit (https://www.rdkit.org/) and Mordred (https://mordred-descriptor.github.io/documentation/master/descriptors.html) libraries were harnessed for molecular descriptors calculation, while Qt designer and PyQt 5 frameworks were leveraged for creating the graphical user interface (GUI) of the application, along with custom classes such as TableModel, MplCanvas, and PandasModel.

2.2. Interactive user interface

The interactive user interface of the chromatographic prediction system is thoughtfully engineered to provide a seamless and user-friendly experience, enabling effortless access and utilization of advanced functionalities. Developed in Python, the graphical user interface is constructed on Qt-designer and PyQt5 framework, allowing smooth integration with the

TensorFlow 2.0 neural network framework. This combination ensures efficient neural network construction and interface realization, while also facilitating intuitive interaction with the software. The main menu of the software offers a dedicated network training interface tailored specifically for chromatographic prediction applications (see **Figure 1**).

(a) (b)

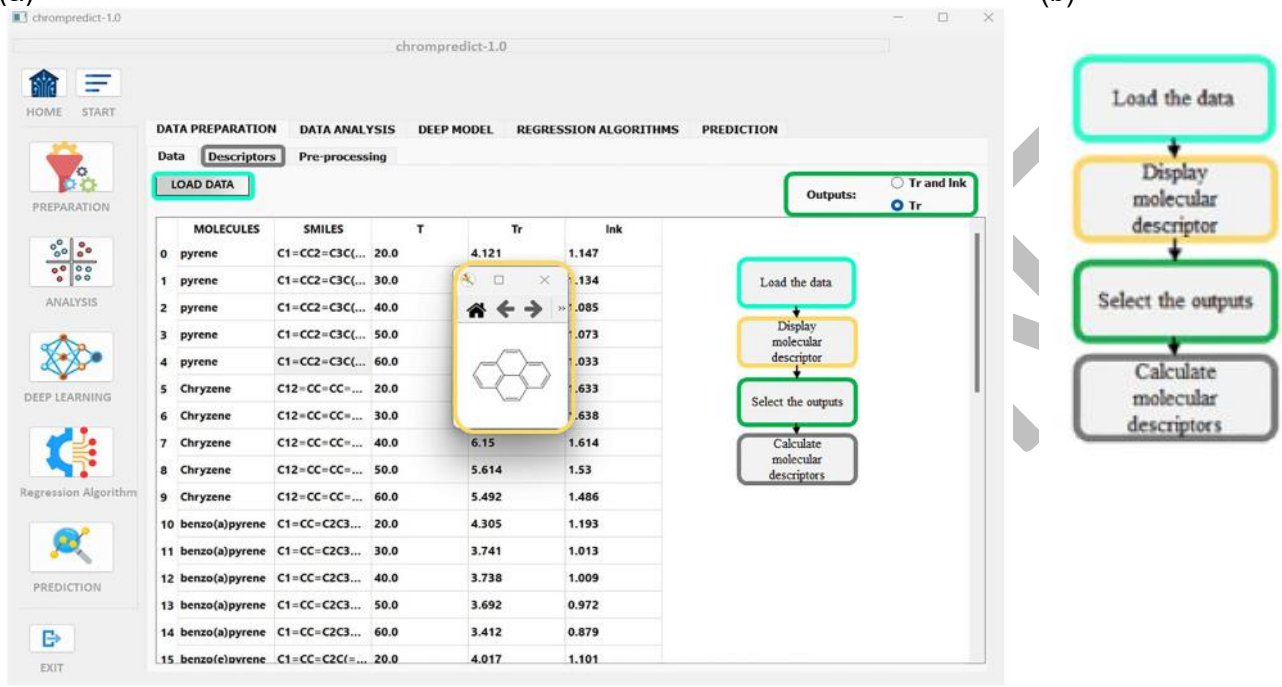

Figure 1. *(a) Example of the main menu of a data preparation tab or widget. This image provides an overview of the options, guiding users through data loading, visualizing molecular structures and selecting the desired input features. (b) Description of the four successive steps involved in preparing data for model training and optimization.*

This interface enables users to manage training data, access existing data, create new deep neural networks, and save network configurations. By providing a comprehensive interface, the software simplifies the process of setting up and managing neural network models, thereby enhancing efficiency in predicting chromatographic properties. Validation serves as a pivotal phase in model construction, facilitating a rigorous assessment of the Artificial Neural Network's (**ANN**) efficiency. This involves deploying specific ANN performance measures to gauge the network architecture's superiority and its predictive capabilities concerning chromatographic behavior. The main interface of the chromatographic prediction software assumes paramount importance as it provides crucial insights into the neural network's characteristics. This encompasses comprehensive details concerning the network structure, error metrics, and accuracy measures that pertain specifically to chromatographic predictions.

2.3. Main functional structure

The functional structure of **Chrompredict 1.0** software was specifically designed to optimize the workflow for predicting retention time and retention factor in chromatography experiments. It offers a comprehensive set of tools that enable users to make accurate predictions and gain insights into the relationships between input and target variables. The main flowchart of the software is schematically described in **Figure 2**.

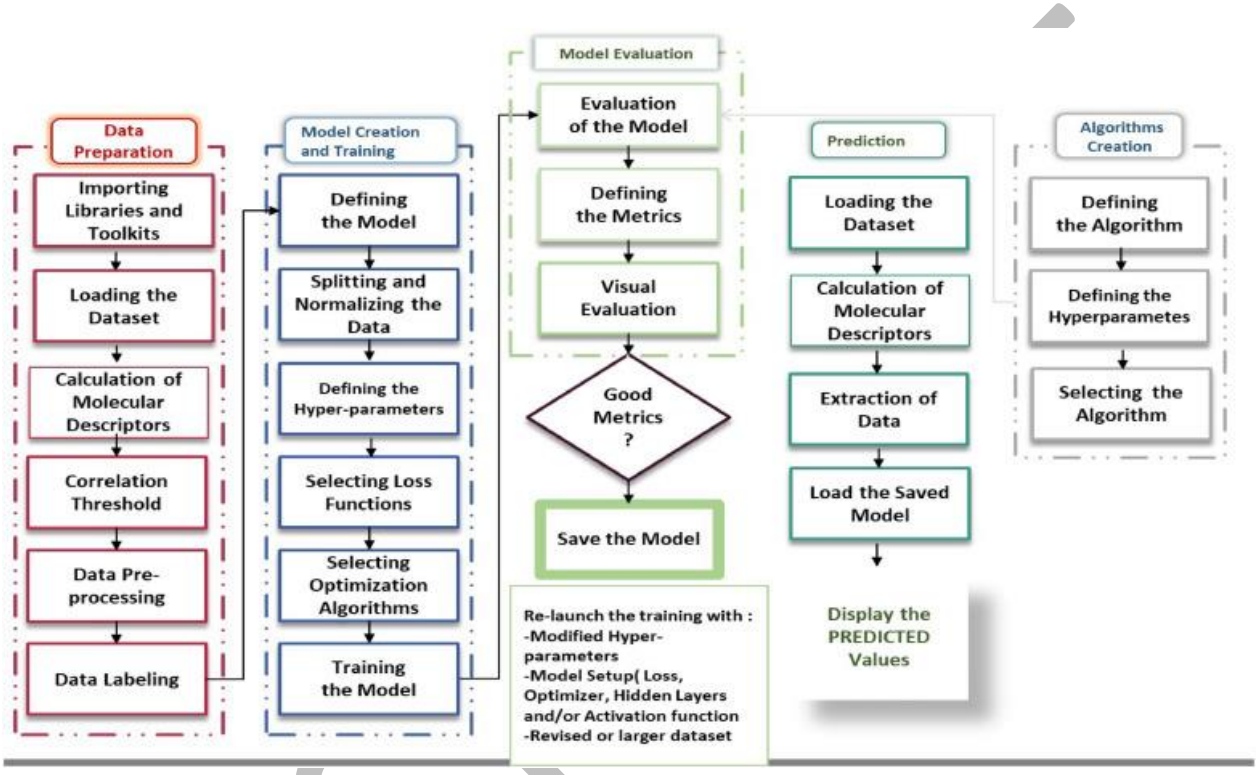

Figure 2. *Flowchart of the* ***Chrompredict 1.0*** *software. This flowchart encapsulates the sequential steps executed by the program to predict outputs. Beginning with data input, it navigates through preprocessing, feature extraction, model training, and validation. The chart succinctly outlines the essential steps culminating in accurate predictions).*

The data preparation section allows the following functions: loading the experimental dataset, calculating molecular descriptors, data pre-processing, and data labeling. The goal is to prepare the dataset automatically and feed it into the deep neural network. Subsequently, a suite of analytical tools, including summary statistics and various plots, is offered. These features elucidate the relationships between input and target variables, improving the robustness and accuracy of predictive models. The model training section offers custom model creation, parameter adjustment, optimization algorithm selection, and model preservation, ensuring stable training and improved generalization. The software also incorporates a selection of six ML models as options, helping users with tasks of comparison and benchmarking. Model evaluation involves the trained model and regression algorithms,

using the test dataset. Various evaluation metrics were implemented to assess the model's accuracy and its predictive capability. Users can visualize the performance of the selected models using different regression functionalities plots, aiding in the interpretation and understanding of the predictive models.

The prediction component encompasses: (i) model loading, (ii) displaying retention time and retention factor predictions, enabling users to compare and (iii) assess the performance of different prediction methods, including both deep learning and machine learning models.

2.4. Data preparation function

2.4.1. Calculation of descriptors. During the data preparation phase, a crucial step involves computing molecular descriptors, which are the result of a logical and mathematical process that transforms chemical information from a molecule's symbolic representation into a useful number or the outcome of a standardized experiment for the injected solutes (see **Figure 3**). This is achieved using the RDKit and Mordred libraries with the descriptors being calculated from SMILES representations of the molecules. These descriptors serve as input features for deep learning models and regression algorithms, offering a comprehensive representation of the molecular structures and their properties. Consequently, this enables an in-depth analysis of the molecule behavior and interactions with the stationary phase [**29,30**].

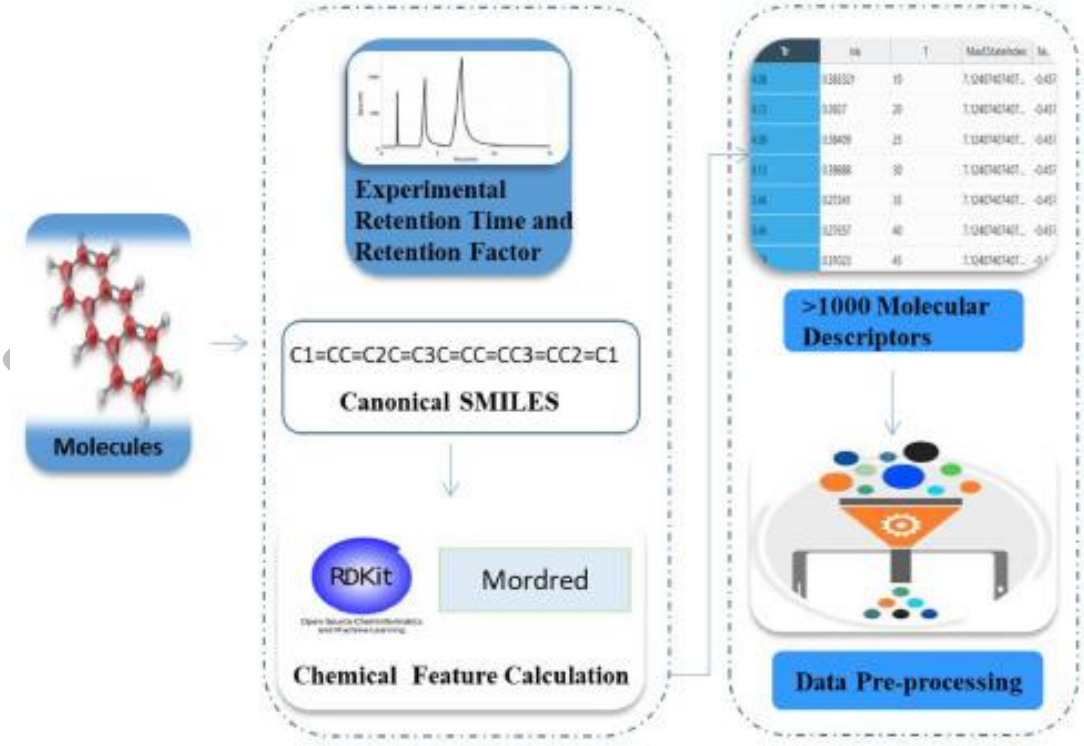

***Figure 3**. Detailed flowchart of data preparation in the software workflow illustrating the systematic data preparation process, a pivotal step in the software workflow. The essential features of the model's input are shown here.*

Over 1000 molecular descriptors were calculated, covering molecular topological, geometrical, and electronic properties. From a quantitative perspective, these numerical values depict the physical and chemical characteristics of the molecules, such as molecular weight, Log *P*, number of rings, and hydrophilic factor.

2.4.2. Data pre-processing. In order to enhance the model's ability to automatically extract features, the data pre-processing part will perform automatic feature selection and transformation, ensuring that the deep learning model is trained on the most relevant and informative features (see **Figure SI-1**). This is a crucial stage in any ML or DL study, as it directly impacts the model's performance and generalization capabilities.

During the pre-processing stage, several key steps are carried out to curate the dataset and retain only the most significant features. These steps include: (i) dropping columns with fewer than five unique values, as they may not provide sufficient information for the model to learn effectively, (ii) selecting columns with a standard deviation greater than 0.01 [**30**], as they are likely to have additional variation and further contribute to the model's learning process, (iii) filtering the Data Frame to only include highly correlated features and the target variable; users can choose the correlation threshold with the target variable during the filtration process, ensuring that the selected features have a strong relationship with the target variable, (iv) dropping correlated features with a correlation threshold of 0.96 [**31,32**], reducing multicollinearity and ensuring that the model is trained on independent features.

By performing these automatic pre-processing steps, the resulting Data Frame contains only the most relevant and informative features. This allows the deep learning model to focus on learning the underlying patterns and relationships in the data more effectively, ultimately improving its performance and ability to generalize to new data.

2.5. Data analysis

The chromatographic prediction software presents an extensive collection of data analysis instruments, enabling users to visualize and comprehend the underlying patterns and features of their data effortlessly. By offering different types of graphical representations, the software allows users to delve into data skewness, pair plots, box plots, heat maps, histograms, and data types, rendering the data analysis process highly efficient and user-friendly. Additionally, users can register their files, make any relevant changes manually, and re-upload them if necessary, ensuring flexibility and control over their data analysis workflow (see **Figure SI-2**).

2.6. Data training

After processing, input data undergoes normalization with the Min-Max Scaler. The split function is implemented to diyvide the dataset into training and testing subsets based on the user-defined split percentage.

2.6.1. Adaptive deep neural network architecture. The Deep Neural Network **(DNN**) model consists of an input layer, user-defined hidden layers with varying neuron counts, and an output layer with one or two neurons for dual output predictions. The activation function for the hidden layers is also user-defined (see **Figure SI-3**). The model is compiled using the user-selected loss function and optimizer algorithm, with early stopping implemented to prevent overfitting. The training progress is logged using the TensorBoard callback. The model is trained on the training data for the user-defined number of epochs and batch size, which can better meet the scope of application of the system to rotating machinery. Howevery, the percentage for splitting the data is specified by the user, with the tested dataset selected randomly to ensure variability. After training, the user can save the model for future use in predictions (see **Figure 4**).

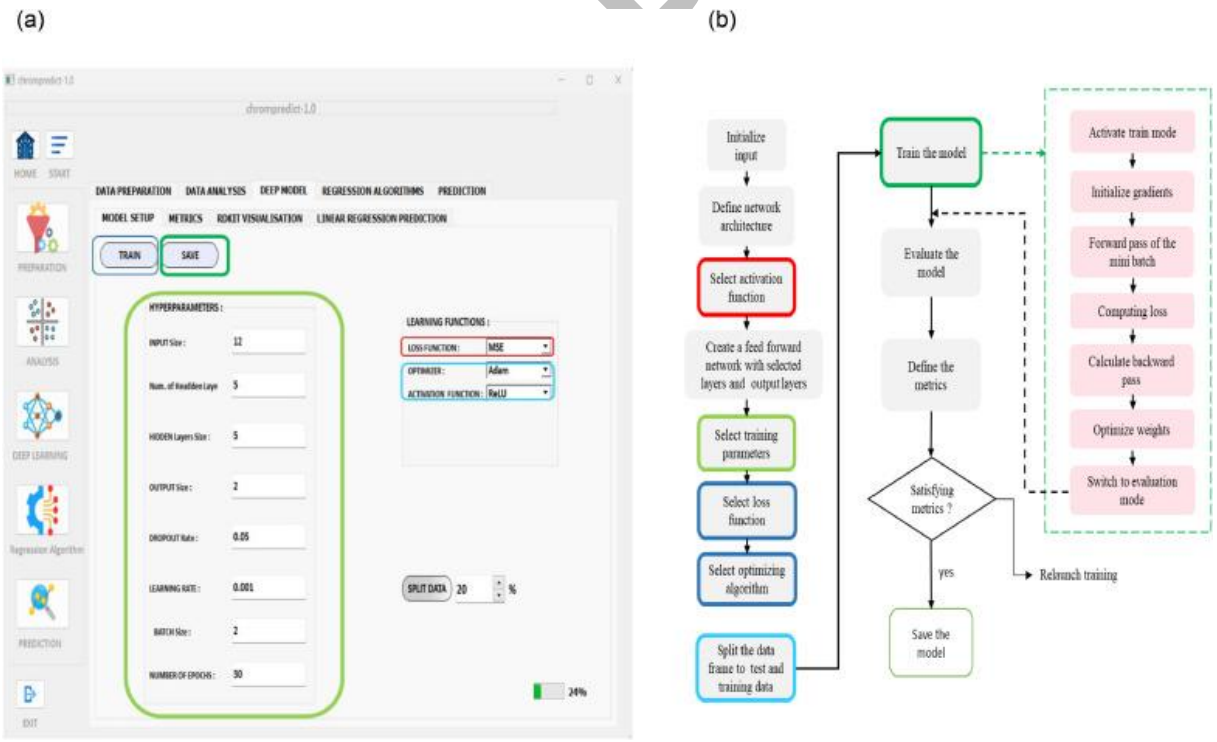

Figure 4. Model set up and training. (a) A screenshot, (b) Flowchart serving as a visual guide and displaying the software's capabilities in model setup and training. By providing specialized tools and features, the software streamlines the process of model configuration and training, enhancing user experience and facilitating accurate predictions.

2.6.2. Machine learning algorithms. This section aims to compare the performance of various regression algorithms (see **Figure 5**), such as Linear Regression (**LR**), Random Forest Regression (**RFR**), and Support Vector Regression (**SVR**), with the performance of DNN models. Users can choose from these algorithms and define their parameters. The selected algorithm with high evaluation metrics can be saved for future use. The goal is to assess the potential for achieving higher evaluation metrics using these algorithms and to determine their scalability for use in predicting small molecule behavior in chromatography.

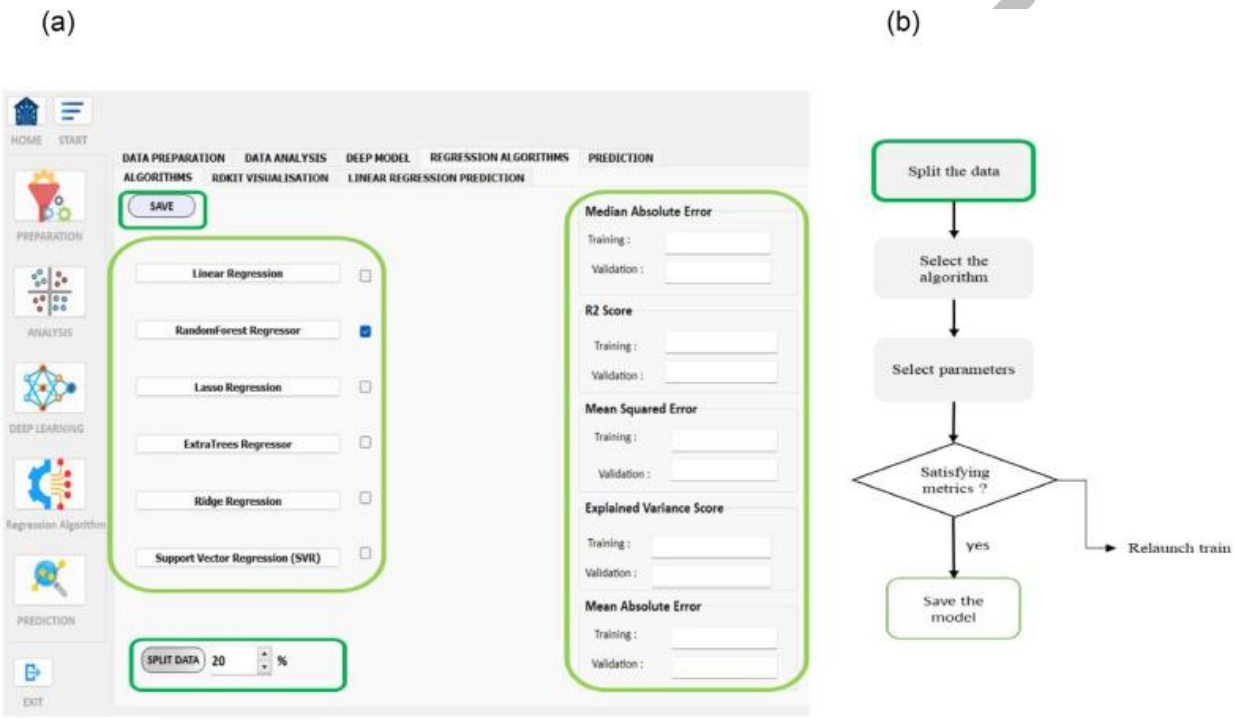

Figure 5. *Machine learning regression algorithms training. (a) A screenshot capturing the Machine Learning Algorithms window in the software, featuring a flowchart that serves as a guide. (b) Flowchart elucidating the sequential steps leading to the development and saving of the final model, offering valuable insights into the algorithmic processes.*

2.7. Model performance insights

The user interface provides an extensive selection of evaluation visualization tools, streamlining the assessment of the model's performance and allowing users to identify potential issues related to overfitting or underfitting. Feedback on the model's outputs is displayed through TensorBoard, evaluation regression plots, or alternative performance metrics,enabling users to closely monitor the training progress and make well-informed adjustments as necessary. By offering such transparency, the software ensures optimal model performance and heightens the precision of chromatographic predictions (see **Figure 6**).

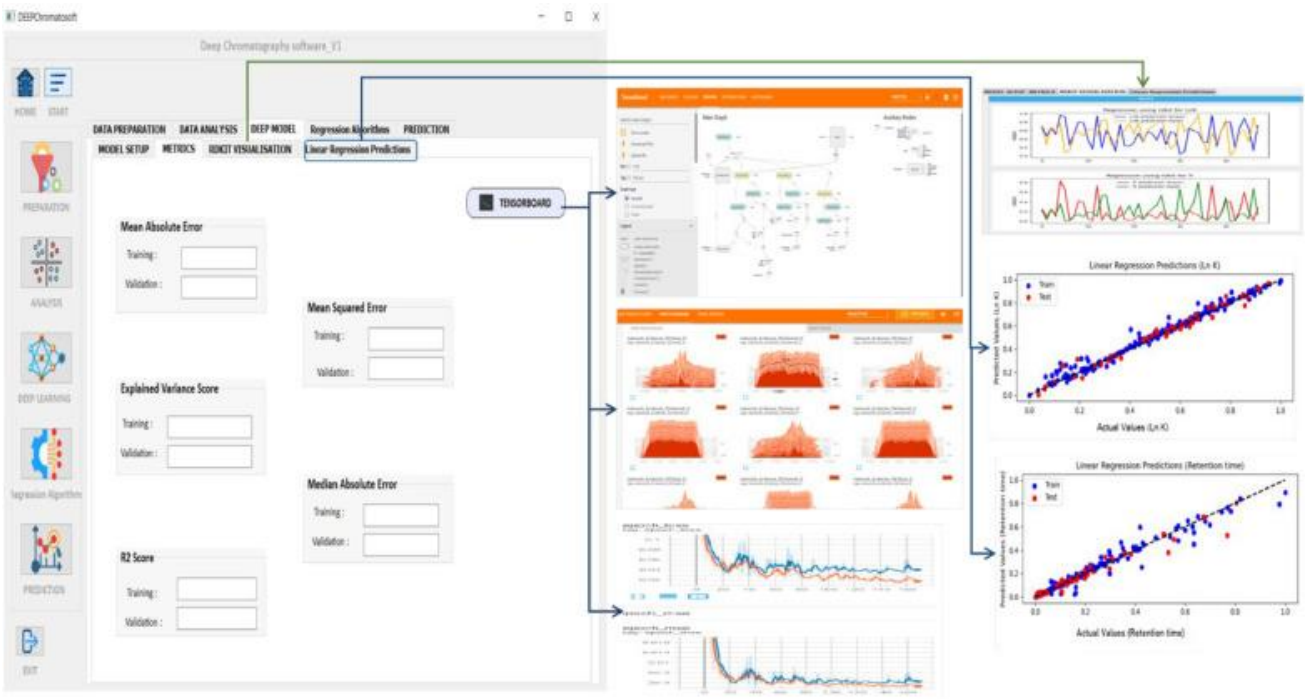

Figure 6. *Screen shot of the evaluation tab Widget main menu. Users can access crucial information pertaining to model evaluation and performance metrics, enabling informed decision-making and refinement of predictive models. This figure serves as a key reference point, highlighting the significance of the software interface in evaluating and optimizing chromatographic predictions.*

2.8. Prediction process

The software displays predicted outputs for both DNN and ML saved models, providing a comprehensive comparison of their respective performances. In the prediction phase, users have the ability to specify the injection temperature, allowing for greater control over the analysis process (see **Figure 7**). Additionally, this model can be further enhanced to predict retention times based on the composition of the mobile phase and other parameters influencing separation. Notably, it also supports simultaneous predictions for multiple molecules, thereby offering a more comprehensive understanding of chromatographic behavior across diverse analyte mixtures.

3. Experimental section

3.1. Chemicals and reagents

All reagents used in the different syntheses were purchased from across without any further purification. The totally porous silica particles (specific surface area: 425 $m^2.g^{-1}$), particle diameter: 5 µm, pore diameter: 100 Å were obtained from Merck (Darmstadt, Germany). All solvents were chromatography grade and purchased from Fluka (Buchs, Switzerland).

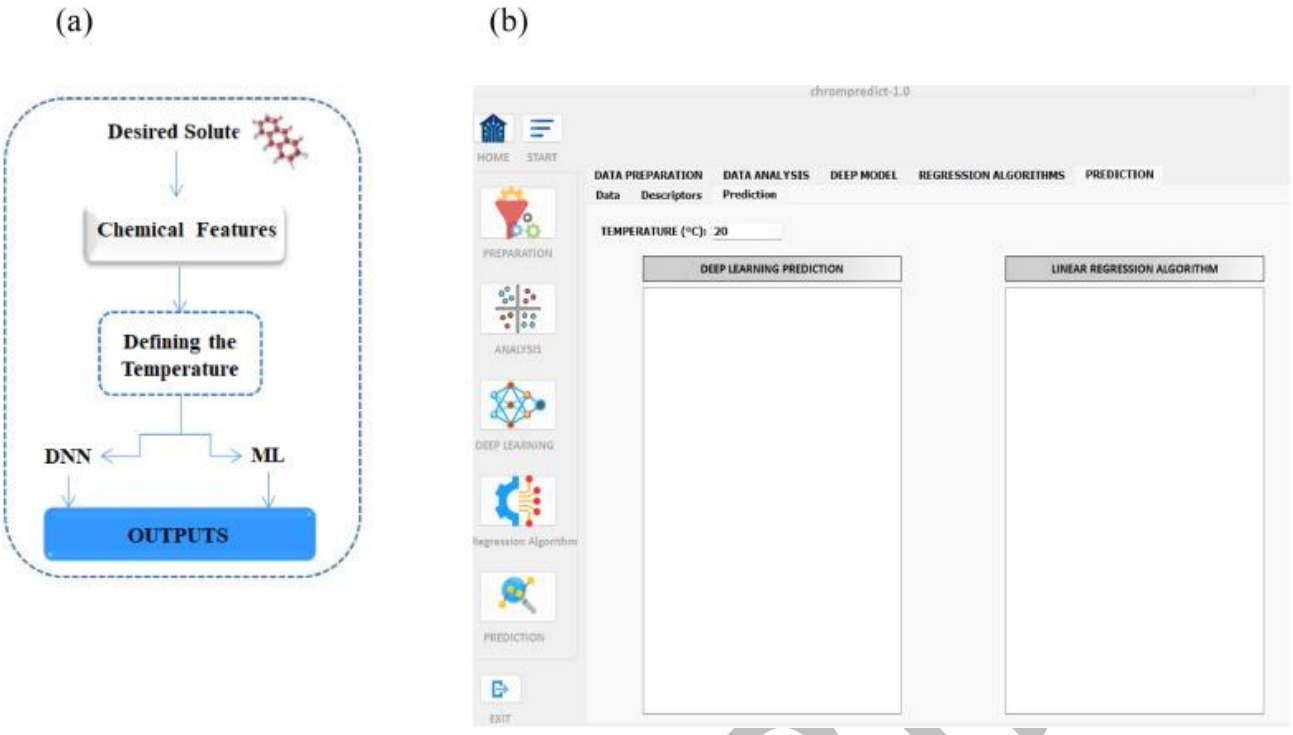

Figure 7. *(a) Prediction function process involved in* ***Chrompredict 1.0****. (b) Associated menu in the user interface.*

3.2. Experimental apparatus

A Waters ALC/GPC 244 chromatograph with a 6000 A pump, a 7125 Rheodyne syringe with a 20 µl sample loop, and a Waters UV240 detector operating at 254 nm for detection and analysis purposes. A 15 × 0.4 cm column was packed with CELC using the slurry technique [33]. Differential Scanning Calorimetry (**DSC**) was carried out using a Mettler FP85 device while elemental analyses were performed at the Microanalysis Lab (Service de microanalyse) of the French National Center for Scientific Research (Gif-sur-Yvette, France). CESP specific surface area was determined using the nitrogen BET method on a Micromeritics Accusorb 200 instrument. ^{1}H NMR analysis was performed on a 250 MHz (5.8 T) Bruker NMR spectrometer.

3.3. HPLC experiments

To investigate the thermal and chromatographic properties of CESP, a controlled temperature water bath with a thermocouple was set up to regulate the column temperature. Both the column and the mobile phase tank were immersed in the water bath throughout the experiment. Normal phase chromatography was performed using 100 % hexane as the mobile phase, followed by 100 % isooctane at higher temperatures (above 69 °C). While reversed phase chromatography utilized a 35/65 (v/v) mixture of acetonitrile and water. The

high-performance liquid chromatography (HPLC) analysis was conducted under isocratic conditions with a flow rate of 1 mL/min. To switch from normal mode to reversed mode, the column was rinsed sequentially with isopropanol and isopropanol-acetonitrile at a low flow rate for 2 h, followed by rinsing with HPLC grade acetonitrile for 1 h. Finally, the acetonitrile-water mobile phase was introduced until the chromatographic system reached equilibrium.

3.4. Synthesis and NMR characterization of the mesogenic material

The synthesis of intermediate products, noted (**1**), (**2**) and, (**3**) is given in **Figure 8**. The final product CELC is synthesized through series of reactions that involve assembling these intermediates in a specific manner (see **Figure 9**). The process is complex and requires careful control of reaction conditions and purification steps to ensure high product yield and purity. The resulting product is further analyzed using ^{1}H NMR in $CDCl_3$ to confirm its structure and purity (see **Figure SI-4**).

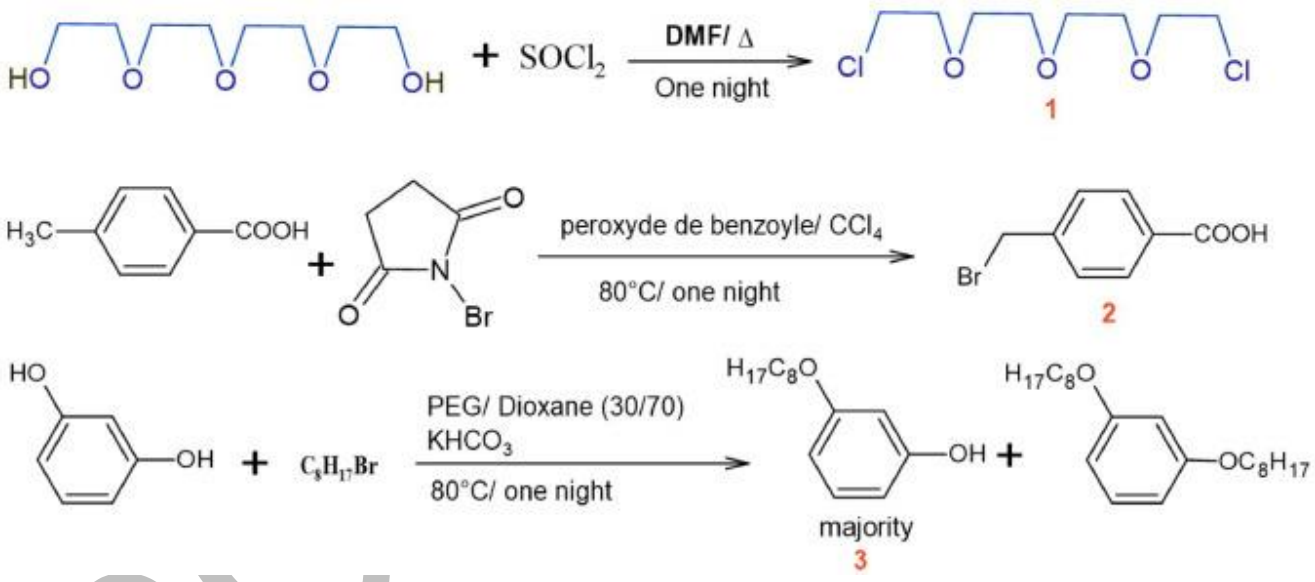

Figure 8. *Synthesis scheme of three key intermediates (**1, 2, 3**) of the crown ether liquid crystal.*

3.4.1. Synthesis of intermediate products. Three crucial presynthetic steps were followed to form key intermediates: 1-chloro-2-{2-[2-(2-chloroethoxy)ethoxy]ethoxy}ethane (**1**), alpha-bromo-para-toluic acid (**2**), and 3-octaoxyphenyl (**3**). These intermediates were subsequently utilized in the final product synthesis of crown ether liquid crystal. The synthetic procedure for the intermediate **(1)** involved evaporating 2,2′-[oxybis(ethane-2,1-diyloxy)]di(ethan-1-ol) evaporate under vacuum of $SOCl_2$ and extracting the formed product with dichloromethane. The resulting product was then washed three times with water, dried

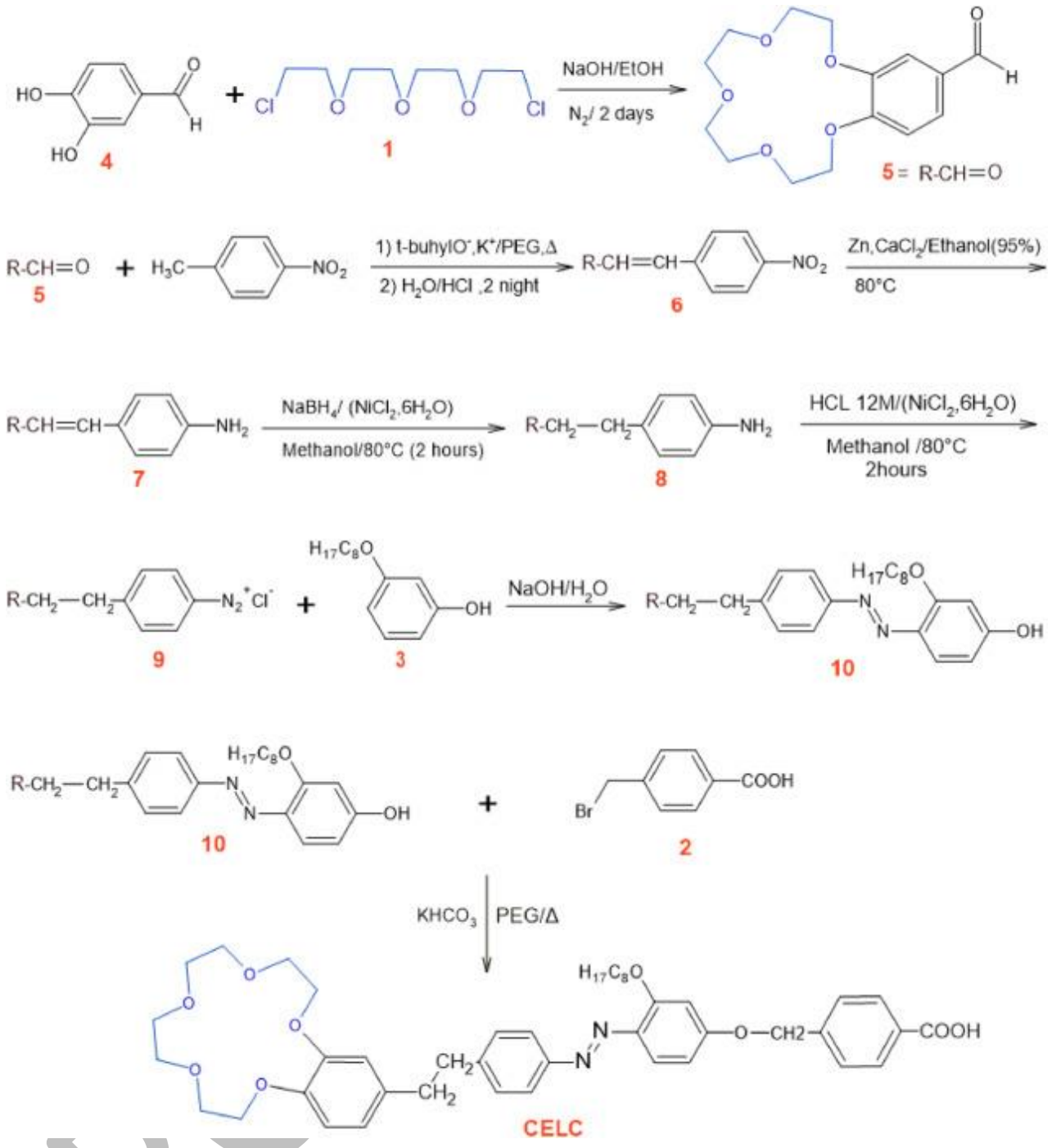

Figure 9. *Complete synthetic route to 4-(4-{4- [2-(6,7,9,10,12,13,15,16-octahydro-5,8,11,14,17-pentaoxabenzo-cyclopentadecen-2-yl)-ethyl]-phenylazo}−3-octyloxy-phenoxymethyl)-benzoic acid liquid crystal, the CELC investigated here.*

over $MgSO_4$, and filtered. The obtained product's purity was verified by conducting ^{1}H-NMR analysis in $CDCl_3$. This intermediate material was subsequently used in the synthesis of the final product. The second intermediate was obtained by reacting sand 1-bromopyrrolidine-2,5-dione in benzene peroxide/CCl_4 at 80 °C for a duration of one night. The product obtained was 4-(bromomethyl)benzoic acid (**2**). Finally, the third intermediate (**3**) was synthesized by reacting benzene-1,3-diol and 1-bromooctane under the following conditions: PEG/Dioxane

(30/70) as a solvent and $KHCO_3$ as a base, heated at 80 °C over a night. This intermediate was obtained in high yield (89 %) and was the major product of the reaction.

3.4.2. Synthesis of mesogenic material. The synthesis of crown ether part of the final molecule was achieved by alkylation of aldehydes through the reaction of 3,4-dihydroxybenzaldehyde (**4**) with (**1**) in the presence of sodium hydroxide and ethanol under a nitrogen atmosphere for two days. The crude yield (**5**) was purified by chromatography on silica gel using a mixture of ethyl acetate and dichloromethane (50:50) as the mobile phase. Subsequently, the aldehyde (**5**) was treated with 1-methyl-4-nitrobenzene in the presence of a mixture of potassium tert-butoxide and polyethylene glycol, followed by the addition of dilute hydrochloric acid for two nights to afford the desired nitro compound (**6**) in good yields. The crude yield was then partially reduced with a mixture of Zn, $CaCl_2$, and 95 % ethanol by heating at 80 °C (**7**), followed by a total reduction of the nitro group in the presence of $NaBH_4$ and $NiCl_2/6H_2O$, and then with methanol at 80 °C for two hours to afford the corresponding product (**8**) in moderate to good yields.

To form the diazo compound (**9**), two steps were required. First, the amine group was selectively ionized using an HCl 12M/dioxane system and $NaNO_2/H_2O$. Second, a selective reduction reaction with $NaOH/H_2O$ was performed in the presence of 3-(octyloxy)phenol (**3**) to alkylate in the *meta* position, resulting in the desired product (**10**). The organic phase was then extracted with dichloromethane after evaporating the dioxane, washed in acidified water multiple times, dried with $MgSO_4$, filtered, and the solvent was eventually evaporated. The resulting residue was chromatographed on silica gel (70–230 mesh) using ethyl acetate as the mobile phase, and the pure compound was collected in the final fraction. Finally, the mesogenic acid was formed by reacting the solid product (**10**) with 4-(bromomethyl) benzoic acid in PEG as a solvent, heating, and using $KHCO_3$ as a co-solvent. The acid formed in the solid state was recovered by adding acidulated H_2O, which allowed it to be precipitated. The crude solid was then filtered through sintered glass. The recrystallized final CELC (see **Figure SI-5**) obtained in an appropriate solvent, was dried under vacuum.

3.4.3. Amidification reaction and grafting procedure. In the last stage, the amidification reaction (see **Figure 10**). was performed by combining the ABDMS-silica with the liquid crystal (**CELC**), using a microwave-assisted grafting approach [**34,35**], ABDMS-silica was meticulously prepared, following the guidelines outlined in relevant scientific literature [**35**], after that, a 24 h drying process was applied to the ABDMS-silica, with the temperature being kept at 110 °C and under vacuum. In the final phase, a 50 mL Pyrex vessel was utilized to

combine15 mL of trifluorotoluene with the ABDMS-silica, and the contents were stirred. Simultaneously, in another 50 mL Pyrex vessel, a mixture consisting of 1 g of CELC, 1.63 g of (Benzotriazol-1-yloxy) tripyrrolidinophosphonium hexafluorophosphate (PyBOP), and 1.09 mL of *N,N*-diisopropylethylamine (DIPEA) in a suitable solvent was also being agitated. Afterward, both mixtures were amalgamated and placed within a microwave environment, wherein they underwent a grafting process lasting for a duration of 1 h, with the temperature set at 97 °C. Upon the completion of the grafting process, a concluding step involved the introduction of 2.4 mL of an end-capping reagent, namely N-(trimethylsilyl) dimethylamine. This reagent was introduced into the mixture, which was then subjected to an additional microwave treatment, lasting 1 h at a temperature of 90 °C.

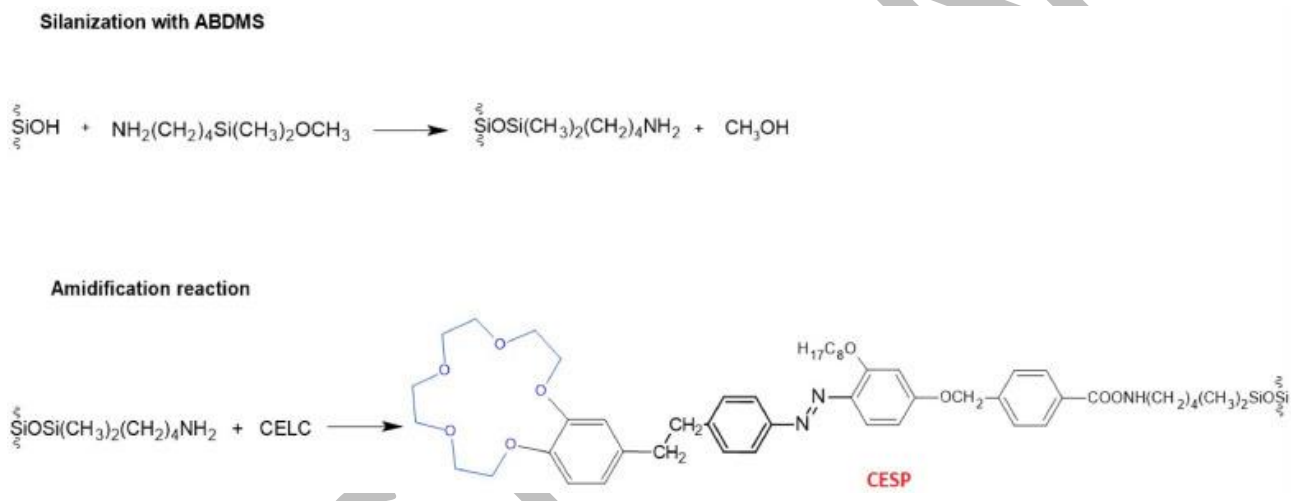

Figure 10. *Grafting procedure of the CELC on the inert support of the column.*

Following this, any surplus reagent was effectively removed through a series of rinses, involving 30 mL of trifluorotoluene, 30 mL of tetrahydrofuran, and 30 mL of acetonitrile. The resultant CESP product was subsequently exposed to a drying process at a temperature of 110 °C, under vacuum conditions, extending for a period of 24 h. This material is then packed into an HPLC column using a slurry technique [**33**]. The mesogenic group can provide selectivity for certain types of analytes, while the silica material provides a stable and inert support for the stationary phase.

3.4.4. Structural characterization. The ^{1}H NMR spectrum of the synthesized CELC was recorded in $CDCl_3$ (see **Figure SI-4**). All δ values are given in ppm. The peaks correspond to different methylene groups within the ether-linked alkyl chain, each with a distinct chemical environment. The peak at δ = 0.80 (d, 3H, CH_3) represents the terminal methyl group whereas other peaks at δ = 1.25 (qd, 8H, 4 × CH_3-CH_2-$\underline{CH_2}$-), δ = 1.5 (m, 2H, CH_3-$\underline{CH_2}$-), δ =

1.75 (qd, 2H, -Ar-O-CH_2-$\underline{CH_2}$-) and 3.75 (t, 2H, Ar-O-$\underline{CH_2}$-) are attributed to methylene groups in various positions along the alkyl chain. The peak at δ = 2.75 (-Ar-CH_2) is due to the methylene groups adjacent to the aromatic rings, indicating their distinct chemical shift caused by the influence of the aromatic system. The methylene of the crown ether cycle is represented by peaks at δ = 3.35, 3.60, and 4.05. The protons bound to the aromatic ring (-Ar-O-CH_2-) are represented by a peak at δ = 5.35 ppm. Singlet peaks at δ = 6.40 and doublets 6.75 and 7.71 (N–H-aromatic-O) correspond to the proton attached to the aromatic ring. The peaks corresponding to the protons on the ring adjacent to the acid function are observed at δ = 6.55 and δ = 7.55 (d, 2H, H-aromatic). Protons attached to the aromatic ring between oxygen and the alkyl group are represented by peaks at δ = 7.88 and 7.32 (d, 2H, R–H-aromatic-O). The proton of the aromatic ring adjacent to the crown ether cycle resonates at 6.90 (d, 2H, H-aromatic).

4. Results and discussion

4.1. Characterization of CESP

The elemental analysis shows the calculated percentages of carbon, hydrogen, and nitrogen in the compound, indicating good purity of the synthesized product. Transition temperatures of CELC determined by DSC are given in **Table 1**, showing a small nematic range at high temperatures.

Table 1. Temperature transitions (determined by DSC).

Transition	Solid → Nematic	Nematic → Liquid
Temperature	142 °C	189 °C

According to the literature [**30**], the following equation (**Eq. (1)**) can be used to determine the coverage density (τ) of CESP:(1). The variables in the equation **Eq. (1)** have the following definitions: pc is the carbon percentage by weight of the bonded material, Mc is the atomic mass of carbon, Mw is the molecular weight of the grafted molecule, nc is the total number of carbon atoms in the bonded organic group, and S_{BET} is the specific surface area ($m^2.g^{-1}$ silica). The analysis of **Table 2** reveals a τ value of 2.06 μmol.m^{-2} for CESP. This value may appear relatively low when considering the surface area, but it's important to note that 76.5 % of the NH groups in ABDMS-silica were successfully bonded, indicating a highly effective coverage of the stationary phase.

Table 2. Elemental analysis of ABDMS-silica and CESP.

Empty Cell	ABDMS-silica	CESP
Carbon (%)	7.4	27.3
Hydrogen (%)	1.86	3.11
Nitrogen (%)	1.38	1.21
τ NH_2, μmol.m^{-2}	2.7	2.7
τ ELCP, μmol.m^{-2}		2.06

4.2. Thermal and chromatographic properties of the CESP

The CESP column integrates both a crown ether ring and a liquid crystal state, these two attributes enhance chromatographic performance by providing shape selectivity and a variety of interaction capabilities with the analytes (see **Figures SI-6** and **SI-7**), thus affecting the observed selectivity characteristics.

To elucidate the impact of temperature on the behavior of the grafted liquid crystal, and based on the transitions identified through differential scanning calorimetry (DSC) analysis, a thermal study of the CESP was conducted in both normal and reversed-phase modes. This study aimed to confirm the presence of the mesogenic state and to investigate potential modifications. For each injection, three independent experiments were performed, and the averages were calculated to ensure precision and reliability. The Van't Hoff equation (**Eq. (2)**) was utilized for this purpose:(2)

In **Eq. (2),** the parameters k, ΔH, ΔS, R, T, and Φ represent the retention factor, the enthalpy of transfer, the entropy of transfer, the ideal gas constant, the absolute temperature, and the phase ratio of the column, respectively. The Van't Hoff plots of ln k *versus.* $1/T$ show a linear trend. A deviation in the curve is observed when structural changes occur in the stationary phase due to its mesogenic state. As a result, the interactions between the solute, mobile phase and stationary phase vary.

4.2.1. Normal-phase conditions. **Figure 11** illustrates the temperature dependence of ln k for a series of aromatic hydrocarbons in normal-phase of CESP. The change in slope of the ln k *versus* temperature curve indicates a change in the enthalpy (ΔH) of the system. The first and second transition temperatures were located at **323** K (50 °C) and **307** K (34 °C). Additionally, the thermal curves reveal a variation in the selectivity of aromatic compounds during heating (**Figure SI.6**). Around the inflection points in the Van't Hoff plots, selectivity increases and falls below 1, indicating a reversal in the elution order of the compounds. This phenomenon is likely due to the presence of the mesogenic state, which enables increased

mobility of the stationary phase chains during heating, thereby making the crown ether cavity more accessible to the analyzed solutes. These properties result in varying efficiency, selectivity, and resolution of the isomers as a function of temperature, emphasizing the importance of understanding thermodynamic properties in chromatographic systems for effective separation strategies.

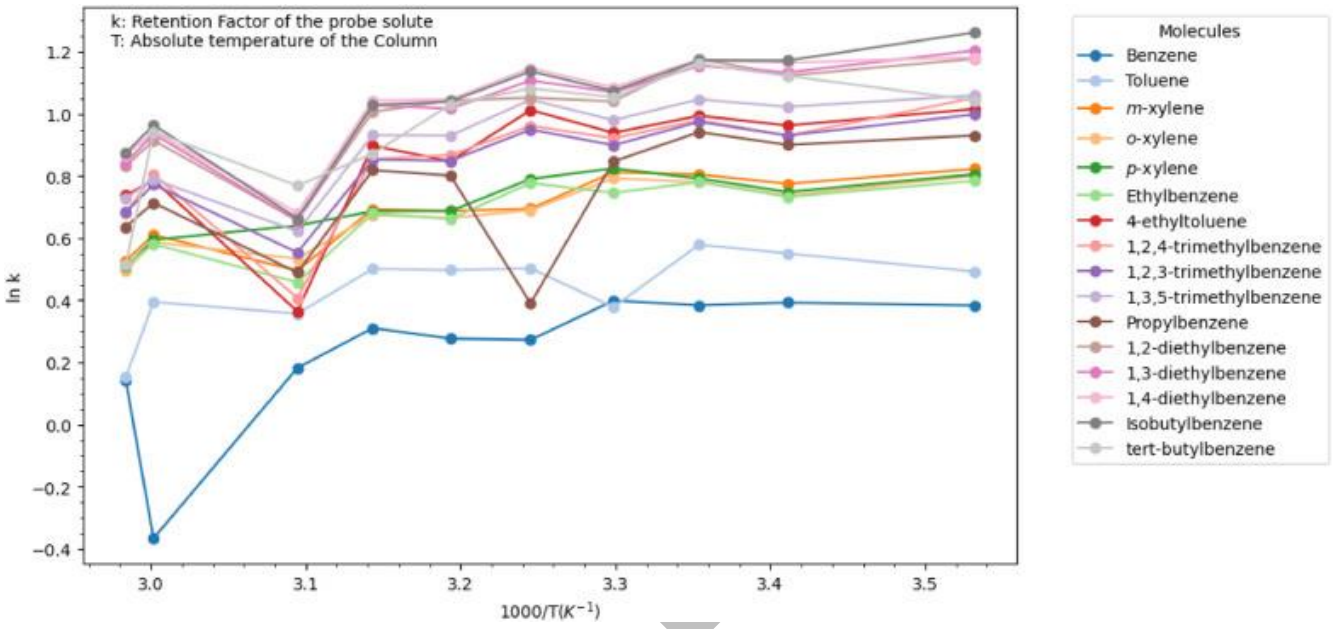

Figure 11. *Temperature dependence of ln k for aromatic hydrocarbons in normal phase. Under normal-phase conditions, the new stationary phase exhibited satisfactory separation efficiency and selectivity for various analytes, including benzene, toluene, p-xylene, m-xylene, ethyl-4-toluene, and trimethyl 1,2,4-benzene.*

For shape selectivity, xylene isomers: *m*-xylene/*o*-xylene (α = 1.05), *m*-xylene/*p*-xylene (α = 1.2), and *o*-xylene/*p*-xylene (α = 1.11), see **Figure SI-8**. Para- and meta-xylenes, known for being challenging to separate, CESP showed a high selectivity as well as for other aromatics solutes. The more linear solute, such as *p*-xylene is retained more strongly than *m*-xylene within the crown ether cavity. That might be due to synergetic effect of the mesogenic chain and the crown ether cavity.

4.2.2. Reversed-phase conditions. The liquid crystal phase influenced by the temperature enhances the interactions between the stationayry phase and the analytes [**14**], adding another dimension to the separation capabilities of the stationary phase.

The Van't Hoff plot in **Figure 12** shows two distinct breakpoints at 293 K (20 °C) and 312 K (39 °C), indicating that CESP phase exhibits unique temperature-dependent mesophase properties. This demonstrates that the stationary phase undergoes a morphological change probably due to the molecular motions of the long mesogenic chains.

This underscores the fact that the liquid crystal stationary phase maintains its mesogenic properties across a range of mobile phase conditions. When the mobile phase is changed going from the normal to reversed phase, it essentially modifies the separation environment. In comparison to the normal phase, a noticeable shift in transition temperatures is observed, with the transitions being less pronounced. This can likely be attributed to the hydrophobic effect of the long-grafted chains in the stationary phase, which, in the presence of water in the mobile phase, restricts their mobility and hinders structural rearrangement.

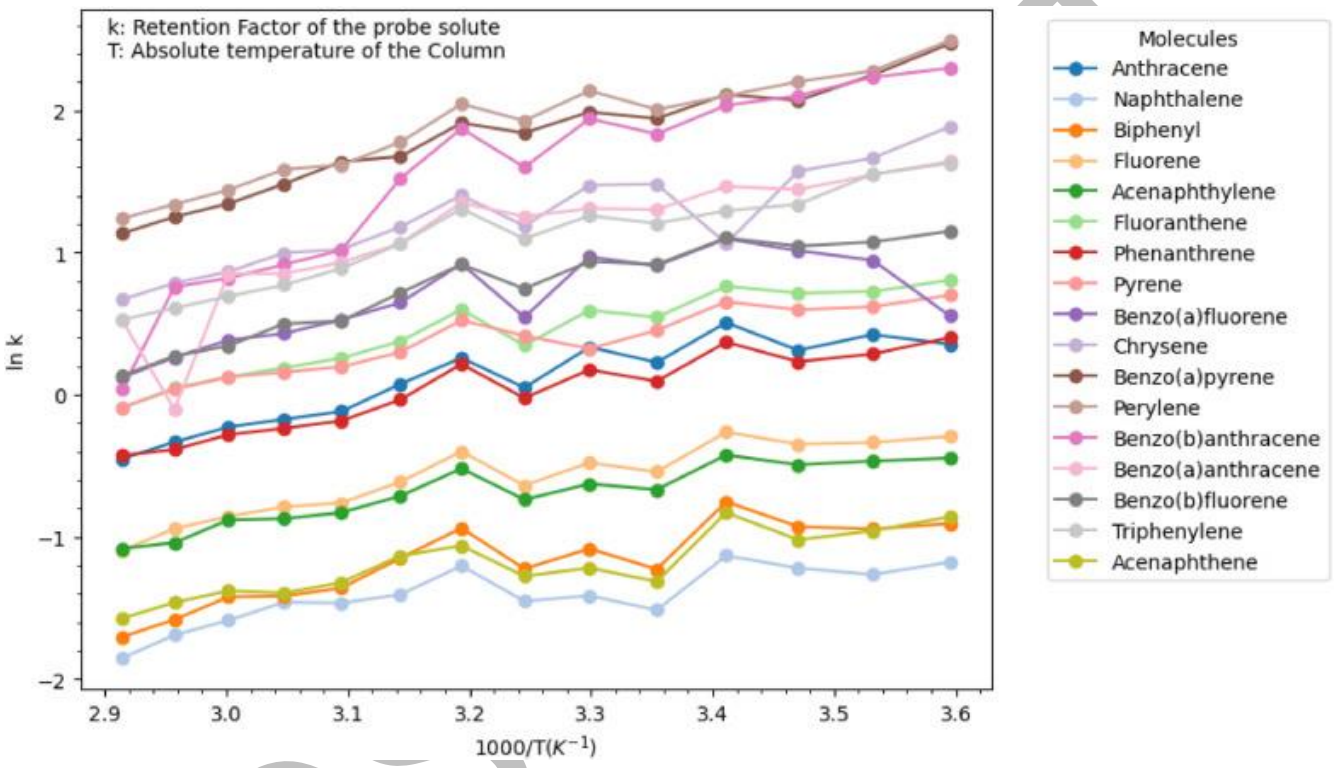

Figure 12. *Temperature dependence of ln k for PAH in the reversed phase.*

To evaluate shape selectivity under reversed-phase conditions, linear polycyclic aromatic hydrocarbons (PAHs) are commonly used to assess aromatic selectivity and retention characteristics of stationary phases across different chromatographic conditions. In this study, seventeen hydrophobic PAHs with diverse geometries were selected as probe molecules. Utilizing a mobile phase composed of 65 % water, the CESP demonstrates exceptional separation of 17 polycyclic aromatic hydrocarbons (PAHs), each characterized by a structure containing 2 to 5 aromatic rings. Notably, five pairs of PAHs were successfully separated, offering valuable insights into the column's selectivity (see **Figure SI-7**). For instance, the CESP column exhibited significant selectivity between phenanthrene and anthracene (α = 1.6), a level of separation rarely achieved with other stationary phases [**3,36**]. Furthermore, anthracene, characterized by its elongated structure, demonstrates greater retention than

phenanthrene, likely attributable to the shape selectivity of the stationary phase. The separation of isomers is also influenced by temperature, as illustrated in **Figure SI-9**. This dependence arises because temperature affects the kinetic energy of the molecules, which in turn impacts their interactions with the stationary phase.

Overall, the new crown-ether based mesogenic stationary phase exhibits promising chromatographic performance in both normal-phase and reversed-phase modes, considering the influence of temperature on the thermodynamics of retention. It excels in separating traditionally challenging isomers and holds significant implications for the development of future advanced stationary phases.

4.3. Applications of Chrompredict 1.0 based on performance of the CESP and several known chromatographic databases

4.3.1. Selected descriptors for the CESP. A systematic comparison of the selected attributes in the reversed and normal phases (see **Figure SI-10** and **Figure SI-11**), revealed that geometric and shape descriptors play a key role in the chromatographic behavior of compounds and their interactions with the stationary phase. Among the evaluated descriptors, Ipc (information of atom-pair connectivity) captures the overall connectivity pattern of a molecule, while Average Atom-Type E-State Index of order 5 for polarizability (**AATS5p**) considers the local environment of each atom. Moran Auto-Transformation of Structure (**MATS**) accounts for the distribution of atoms in a molecule, while Global Graph Index 4 (**GGI4**) and the second Zagreb Index capture the overall shape and complexity of a molecule. These descriptors provide a comprehensive representation of molecular structure and shape, enabling a deeper understanding of their chromatographic behavior.

The unique characteristics of reversed-phase chromatography are highlighted by the dissimilarities in chosen features such as **BalabanJx** (topological index), **Kappa3** (3D molecular shape index), and **ATS6p** (Atom-type topological state) (see **Figure SI-12**). In reversed-phase chromatography, where the stationary phase is nonpolar and the mobile phase is polar, descriptors associated with hydrophobicity, form, and size become more pertinent. These features reflect the interaction between the hydrophobic stationary phase and the analyte molecules, as well as the competition with the polar mobile phase, influencing chromatographic behavior.

In normal-phase chromatography, where the stationary phase is polar and the mobile phase is nonpolar, descriptors connected to polarity, hydrogen bonding, and electronic interactions are more significant. Features such as **GATS5c** (Geary autocorrelation) and **EState_VSA8_y** (E-State van der Waals surface area) illustrate the interaction between the

polar stationary phase and solutes, alongside the competition with the nonpolar mobile phase, dictating chromatographic behavior in this mode.

4.3.2. Selected descriptors for different chromatographic databases. Additionally, four auxiliary datasets were used, including the application of liquid crystals in normal-phase chromatography (LC-LCC8 column) [3] and gas chromatography (LCC1 and GC-3-CH_3 columns) [29,37], featuring a relatively smaller dataset of not > 38 small molecules. Furthermore, the METLIN Small Molecule Retention Time (**SMRT**) dataset [38], providing experimentally acquired reverse-phase chromatography retention time data for 1023 small molecules, was also incorporated as an external dataset thus covering a wide range of molecule types with different chemical structures (see **Table 3**). These diverse datasets provide a comprehensive evaluation of the software's performance and adaptability across a wide range of chromatographic conditions and different data scales.

Table 3. Classification of injected molecules by solute families.

Dataset	Molecule Families
LC-LCC8 [3], (Rev. mode)	Polycyclic Aromatic. Hydrocarbons
LCC1 [31]	Aromatic Hydrocarbons
GC-3-CH_3 [37]	Aromatic Hydrocarbons, Naphthalenes, Terpenes, Terpenoids
METLIN [38]	Aromatic Hydrocarbons, Polycyclic Aromatic Hydrocarbons (PAHs), Heterocyclic Compounds, Alkanes, Alkenes, Alkynes, Alcohols, Aldehydes, Ketones, Carboxylic Acids, Amines, Amides, Ethers, Epoxides, Halogenated Compounds

The selection of these descriptors generated and selected after the pre-processing by **Chrompredict 1.0** suggests that the retention time is influenced by a combination of factors related to the physicochemical properties and structural characteristics of the molecules.

For the METLIN dataset [38], the selected descriptors focus more on specific atomic or molecular properties and their distribution within the molecule. They include measures of hydrogen bonding potential (NHOHCount, nHBDon), acidity (nAcid), autocorrelation properties (AATS3p, AATS1i, AATS3i, AATS4i), and solubility (FilterItLogS). This suggests that the interactions in the METLIN database might be greatly influenced by the presence of specific functional groups, the molecular acidity, hydrogen bonding, and solubility (**Table 4**).

Table 4. Comparison of selected descriptors from multiple databases after pre-processing.

Data base	Data's dimen.	Correl. threshold	Selected Descriptors[a]						
LC-LCC8, Rev. mode	39	65	FpDensity Morgan2	BCUT2D_CHGLO	AATS4s	ATSC4dv	ATSC2i	AATSC1v	MATS3c
			GATS1c	GATS1s	BCUTdv-1l	MINaaaC	CIC2	CIC3	MIC2
			MATS3s	MATS5s	ZMIC2	ZMIC3			
METLIN	1023	67	AATS3p	AATS1i	AATS3i	AATS4i	ATSC1are	AATSC1c	AATSC1are
			BCUTm-1l	Mi	ETA_epsilon_2	ETA_dEpsilon_D	nHBDon	FilterItLogS	AMID_C
			NHOHCount	nAcid	MATS1are	GATS1pe	SLogP		
GC-3-CH_3	38	46	AATS3are	AATSC0dv	ETA_dEpsilon_A	IC1	BIC1	MIC2	PEOE_VSA10_y
			AATS1s	AATS3s	EState_VSA6_y				
			AATS4p	MATS2dv	MATS2d	Xpc-6dv	Xp-5dv	MAXaaCH	MAXaasC
LCC1	23	65	MPC5	Zagreb2	mZagreb1	mZagreb2	Ipc	ATS5d	PEOE_VSA7_y
			BsIC1						

[a] With: **FpDensityMorgan2**: Morgan fingerprint density at radius 2. **BCUT2D_CHGLO**: Burden Modified Eigenvalues 2D - lowest atomic partial charge. **AATS4s**: Average Broto-Moreau autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities. **ATSC4dv**: Autocorrelation of lag 4 / weighted by atomic van der Waals volumes. **ATSC2i**: Autocorrelation of lag 2 / weighted by atomic ionization potentials. **AATSC1v**: Average autocorrelation of lag 1 / weighted by atomic van der Waals volumes. **MATS3c**: Moran autocorrelation of lag 3 / weighted by atomic charges. **MATS3s**: Moran autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities. **MATS5s**: Moran autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities. **GATS1c**: Geary autocorrelation of lag 1 / weighted by atomic charges. **GATS1s**: Geary autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities. **BCUTdv-1l**: Burden Modified Eigenvalues - lowest eigenvalue n.1 / weighted by atomic van der Waals volumes. **MINaaaC**: Minimum atom-type E-state - aromatic carbons. **CIC2**: Information content index (neighborhood symmetry of 2-order). **CIC3**: Information content index (neighborhood symmetry of 3-order). **MIC2**: Mean information content on 2-order neighborhood. **ZMIC2**: Mean information content index (neighborhood symmetry of 2-order). **ZMIC3**: Mean information content index (neighborhood symmetry of 3-order).

The descriptors selected for the liquid crystal columns are more varied, reflecting a broader range of molecular properties. They include measures of molecular structure and electronic properties (FpDensityMorgan2, BCUT2D_CHGLO, AATS4s, etc.), autocorrelation descriptors (AATS1s, AATS3s, AATS3are, etc.), and information content of the molecular structure (Ipc, ATS5d, AATS4p, etc.). This suggests that the interactions in the liquid crystal columns might be influenced by a wider range of factors, including the shape, size, and branching of the molecules, the spatial distribution of their atoms, their electronic states, and the complexity of their structure.

In summary, the selected descriptors in the METLIN database and the liquid crystal columns reflect the different types of interactions that might be occurring in each case, providing insights into the factors that might influence the retention time in chromatography (**Table 4**). The relevance of the selected molecular descriptors lies in their ability to encapsulate fundamental characteristics of molecular structure. They provide insights into key aspects such as molecular geometry, charge distribution, ionization potential, and van der Waals volumes. Consequently, they are instrumental in predicting retention values, as they effectively represent the molecular properties that govern chromatographic behavior.

4.3.3. Exploratory data analysis for different database. The boxplots of the pre-processed data (see **Figure 13**). in both modes (reverse and normal) illustrate the effectiveness of the pre-processing offered by the software. Despite the residual presence of some outliers, the overall reduction in their prevalence suggests that the pre-processing process has significantly improved the quality of the data. These observations indicate that pre-processing plays a crucial role in obtaining cleaned data that is ready for further analysis. For a more comprehensive view, boxplots for other chromatographic databases, which also display the successful pre-processing of the data by the software, can be found in the annex (see **Figure SI-13**).

4.3.4. Evaluation of the deep learning models. The evaluation of the model's performance on the six datasets reveals promising results for both the training and testing sets. The high R-squared values (see **Table 5**) indicate a strong linear relationship between the predicted and actual values, emphasizing the model's ability to capture a significant portion of the data's variance. Moreover, the low mean squared error demonstrates the model's accurate predictions, with the average predictions closely aligned with the true values. The low median reinforces the model's precision, highlighting its consistency in making predictions across various data points. Furthermore, the variance analysis showcases the model's stability in delivering consistent predictions throughout the datasets.

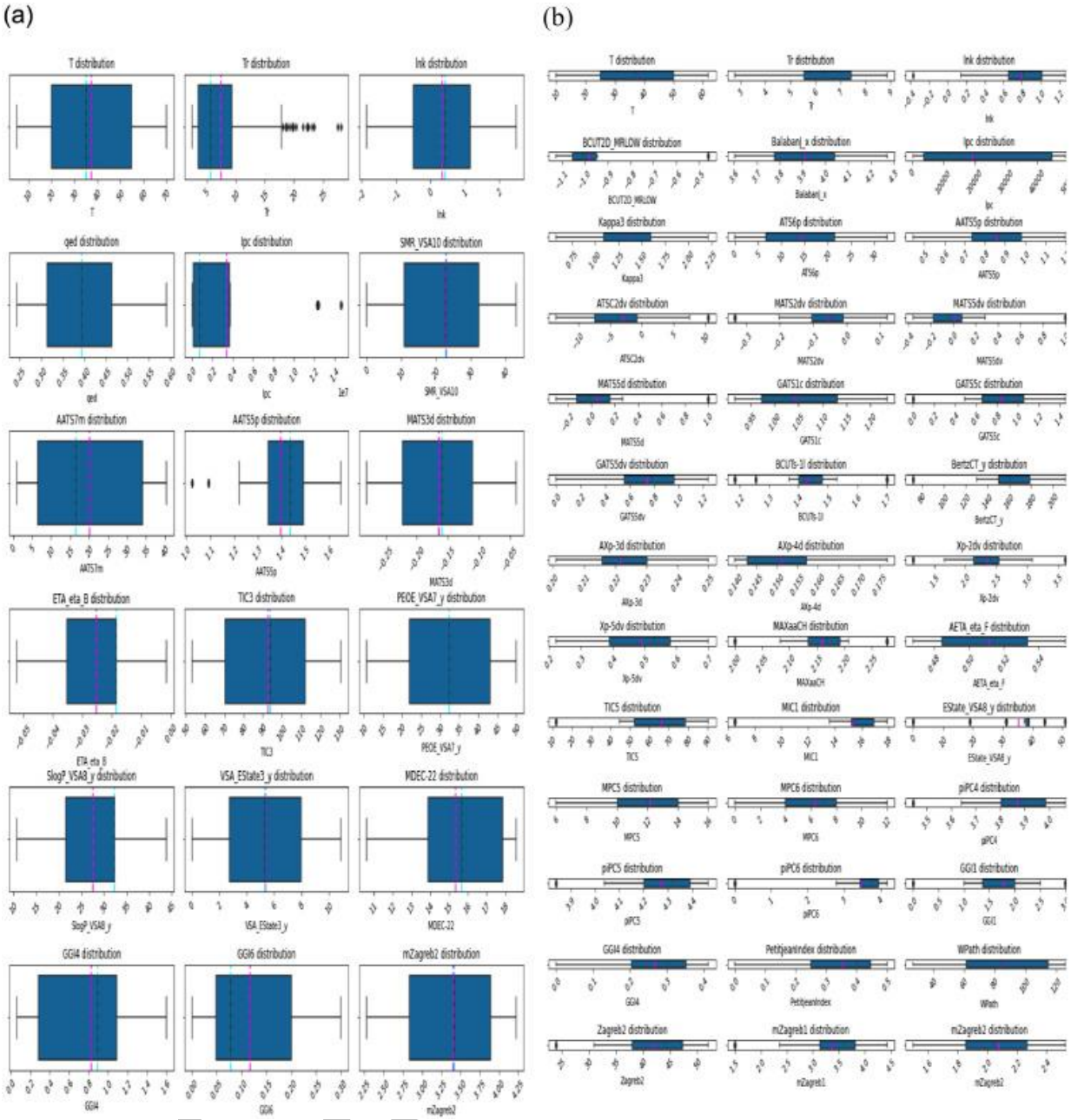

***Figure 13**. Boxplots generated by **Chrompredict 1.0** after automatical preprocessing of two data sets descriptors distribution of ECSP in two different modes: (a) reversed mode and (b) normal mode.*

Table 5. Model evaluation metrics for different datasets.

Phase	CESP Rev. mode				CESP Normal mode				LC-LCC8 [3] Rev. mode				GC-3-CH_3 [37]				METLIN [38]				LCC1 [31]			
AI model	**DL** [a]		**ML** [b]		**DL**		**ML**		**DL**		**ML**		**DL**		**ML**		**DL**		**ML**		**DL**		**ML**	
Data	T [c]	V [d]	T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V
R-squared %	97	96	97	94	97	88	97	87	99	98	97	97	95	37	97	43	95	75	94	66	99.7	98.9	99.9	99.3
EVS % [e]	97	97	97	94	96	88	96	87	99	98	97	98	93	42	97	49	95	63	94	66	99.8	99.7	99.9	99.3
MAE % [f]	1.26	1.45	1.8	3.6	3.16	4.53	1.5	3.6	0.56	3.94	3.15	3.52	4.03	14.3	1.6	10	4.2	12.8	5.1	14.2	1.15	1.6	0.7	1.3
Med-AE % [g]	0.15	0.18	0.6	2.3	2.11	3.58	0.8	2.5	0.07	0.56	2.23	2.66	2.5	9.48	0.14	5.4	2.5	7.4	3.6	11.3	0.8	1.5	0.7	1.7
Mean-SE % [h]	1.51	1.61	0.1	0.3	0.22	0.35	0.07	0.4	2.31	1.28	0.18	0.18	4.2	4.39	0.1	2.5	0.4	3.6	0.5	3.4	0.02	0.03	7.1	0.02

[a] DL: Deep learning. [b] ML: Machine learning. [c] T: Training.
[d] V: Validation. [e] EVS: Explained variance score. [f] MAE: Mean absolute error.
[g] Med-AE: Median absolute error. [h] Mean-SE: Mean squared error.

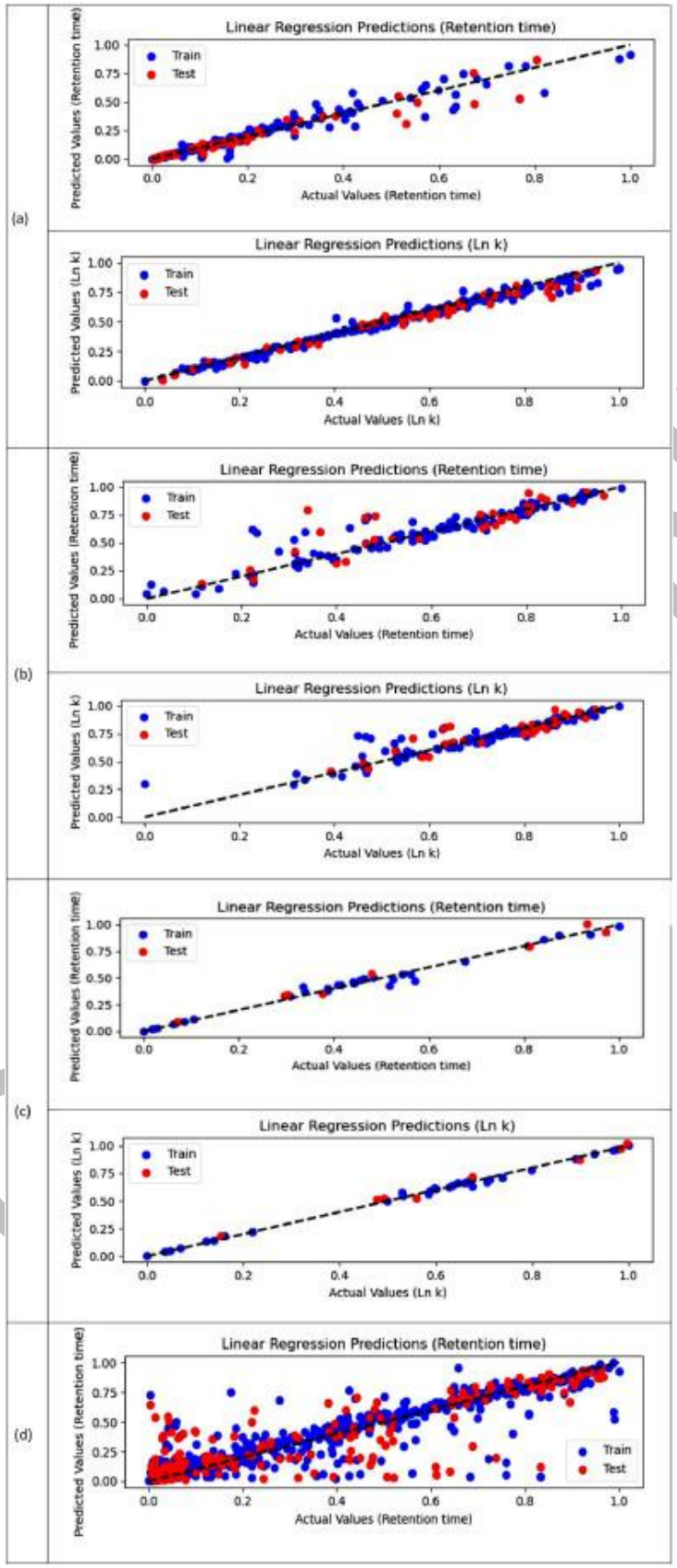

***Figure 14**. Linear regression evaluation of T_R (min) and ln k by **Chrompredict 1.0** (deep model prediction) applied to different HPLC datasets: (a) CESP-Reversed mode, (b) CESP-Normal mode, (c) LC-LCC8 and (d) METLIN. Blue dots and red dots correspond to training experimental and testing data, respectively.*

In the linear regression evaluation (see **Figures 14** and 15) the majority of data points closely align with the predicted values, indicating that the model effectively captures the retention behavior across the dataset. The 20 % of data used for testing was selected randomly, ensuring an unbiased assessment of the model's generalization ability. Although minor deviations were observed for certain individual molecules, these deviations fall within acceptable error margins (< 0.25). Overall, the results confirm the robustness and reliability of the model in predicting retention times for diverse compounds.

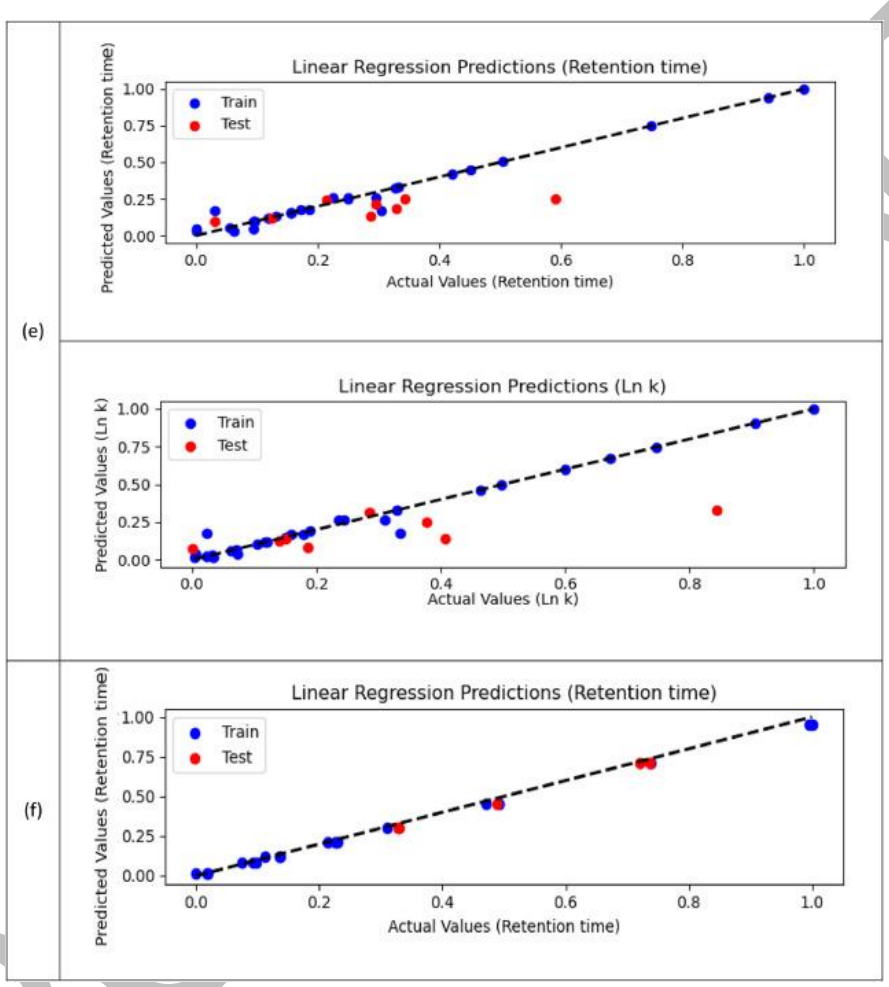

***Figure 15**. Linear regression evaluation of T_r (min) and ln k by **Chrompredict 1.0** (deep model prediction) applied to different GC datasets: (e) GC-3-CH_3 and (f) LCC1. Blue dots and red dots correspond to training experimental and testing data, respectively.*

The results from **Table 5** underscore the accurate predictive capabilities of both ML and DL algorithms for chromatographic behavior. Notably, ML and DL algorithms demonstrated comparable performance with smaller datasets. However, as the dataset size increases, DL algorithms exhibit superior performance, reflected by higher R-squared values and lower error metrics. These findings indicate the potential advantages of using DL algorithms when

working with larger chromatographic datasets, as they effectively capture intricate relationships and patterns within the data.

To further investigate the effectiveness of models in addressing real-world problems, data was collected and compared to the predicted and actual values (**Table 6**). The deep learning model demonstrated a close agreement between predicted and experimental results, highlighting its practical applicability. The minor variations between experimental and predicted values, as commonly observed in predictive modeling, are influenced by factors such as column type, solvent composition, and mobile phase [39,40]. Additionally, the selected hyperparameters, which govern the model's learning process, may also contribute [41].

Table 6. Comparison between real and predicted values for retention time (min) and retention factors across four HPLC datasets.

Technique	**High-pressure liquid chromatography**															
Dataset	**CESP Rev. mode**				**CESP Normal mode**				**LC-LCC8 [3] Rev. mode**				**METLIN [38]**			
Molecule	**Acenaphthylene**				**Tert-butylbenzene**				**Benzo [a]fluroene**				**Cetyltrimethyl ammonium-palmitate / l-alanine**			
AI Model [a]	**DNN** [a]		**ML** [b]		**DNN**		**ML**		**DNN**		**ML**		**DNN**		**ML**	
R/P [b]	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
Tr [c]	2.56	3.64	2.56	3.71	7.12	6.95	7.12	6.67	2.16	2.15	2.16	2.21	2.96/ 7.24	2.83/ 7.74	2.96/ 7.24	3.50/ 6.42
ln *k* [d]	−1.57	−0.36	-1.57	−0.32	0.94	0.90	0.94	0.84	0.15	0.15	0.15	0.18	/	/	/	/

[a] AI model: Deep neural network (DNN) or machine learning (ML). [b] R: Real value, P: Predicted value.
[c] Tr: Retention time (min). [d] ln *k*: Retention factor.

The relevance of the selected molecular descriptors, including FpDensityMorgan2 (associated with fingerprint density), BCUT2D_CHGLO (pertaining to atomic partial charge), and various autocorrelation descriptors (such as AATS4s and MATS3c), lies in their ability to encapsulate fundamental characteristics of molecular structure. These descriptors provide insights into key aspects such as molecular geometry, charge distribution, ionization potential, and van der Waals volumes. Consequently, they are instrumental in predicting retention values, as they effectively represent the molecular properties that govern chromatographic behavior. While the selected descriptors obtained from the RDKit and Mordred libraries provide valuable insights into key physicochemical interactions [40], expanding the dataset with additional molecular descriptors from other sources can significantly enhance the model's predictive performance.

In this context, the retention factors of volatile compounds were successfully predicted, illustrating the versatility of the deep learning model. Moreover, this AI-based approach can be extended beyond HPLC approach, and can be effectively applied to other chromatographic techniques, such as gas chromatography (GC-3-CH_3 [37] and LCC1 [29]) (**Table 7**). In this table, the predictions for molecules not included in the training data were selected for testing, providing a more rigorous evaluation of the model using HPLC data.

Table 7. Comparison betwseen real and predicted values for retention time and retention factors across two GC datasets.

Technique	**Gas chromatography**							
Dataset	**GC-3-CH_3 [37]**				**LCC1 [29]**			
Molecule	***α*-terpineol**				**1,2,4,5-tetramethylbenzene**			
AI Model [a]	**DNN**		**ML**		**DNN**		**ML**	
R/P [b]	R	P	R	P	R	P	R	P
Tr [c]	9.9	11.12	9.9	9.70	8.8	8.97	8.8	8.45
ln *k* [d]	1.41	1.50	1.41	1.48	/	/	/	/

[a] AI model: Deep neural network (DNN) or machine learning (ML). [b] R: Real value, P: Predicted value.
[c] Tr: Retention time (min). [d] ln *k*: Retention factor.

5. Conclusion

A novel AI-based tool called **Chrompredict 1.0**, has been developed to enhance the understanding of chromatographic interactions and to predict retention parameters based on data derived from a newly synthesized mesogenic crown ether stationary phase (CESP) grafted onto ABDMS-silica. This CESP was specifically designed for its unique combination of properties, including a molecular cavity, mesogenic behavior through a mobile chain, and a unique range of polar and non-polar interactions. These distinctive features enable its use in both normal and reversed-phase modes, significantly increasing the versatility and applicability across diverse datasets of our AI-based software's. As expected, it also offers exceptional chromatographic performance, particularly in the separation of linear polycyclic aromatic hydrocarbons (PAHs) and other hydrocarbons, demonstrating its powerful selectivity. Analytical and thermal evaluations confirmed the phase's ability to separate traditionally challenging isomers with a high degree of precision in both normal and reversed modes

Data collected from the chromatographic experiments, including retention times (Tr) and capacity factors (*k*), were integrated into the **Chrompredict 1.0** software, which uses SMILES-derived chemical descriptors to establish relationships between solutes and the stationary phase. After preprocessing, these data were used to train machine learning and

deep learning models, which were evaluated using a range of metrics and regression analyses. The software has been successfully applied to HPLC datasets from both normal and reversed-phase modes, as well as to GC data. Analysis of results have demonstrated a robust stability and strong generalization, achieving prediction accuracies for Tr and ln k values exceeding 90 %. The model was also able to predicted retention factors for volatile compounds, highlighting its flexibility and effectiveness in handling complex chromatographic dynamics. Tested on the METLIN database containing 1023 small molecules of varying structures and polarities, the software obtained an $R^2 > 0.75$ with an error margin of ±7.8 s, confirming its universal applicability.

As the field of chromatography advances, the integration of deep learning models, such as those implemented in **Chrompredict 1.0**, alongside novel stationary phases like CESP, is poised to transform chromatographic methodologies. This innovative approach offers unprecedented precision in predicting retention behaviors and optimizing chromatographic processes, paving the way for significant advancements in the field.

Acknowledgments

We express our thanks to Dr. Jean-Pierre Bayle (ICMMO, University Paris-Sud, France) who designed the liquid crystal and guided and assisted us in the synthesis of the material. We would also like to express our gratitude to the reviewers for their invaluable feedback, which has significantly improved the quality of our manuscript We extend our thanks to Nouredine. Krimat and Rayane Ferroukhi for providing language assistance.

Appendix. Supplementary materials

Data availability
Data will be made available on request.

References

[1] S. Ahuja
Handbook of Analytical Chromatography
(2nd Ed.), Wiley, New York (2002).

[2] L.R. Snyder, J.J. Kirkland, J.W. Dolan
Introduction to Modern Liquid Chromatography
John Wiley & Sons (2011).

[3] N. Mermat, O. Ferroukhi, V. Peulon-Agasse, J.-P. Bayle, M.H. Guermouche, P. Cardinael
Original mesogenic citronellol-based stationary phase for both normal-and reversed-phase HPLC modes: properties and applications
Chromatographia, 83 (2020), pp. 1495-1508.

[4] Y. Wen, M. Talebi, R.I.J. Amos, R. Szucs, J.W. Dolan, C.A. Pohl, P.R. Haddad
Retention prediction in reversed phase high performance liquid chromatography using quantitative structure-retention relationships applied to the hydrophobic subtraction mode
J. Chromatogr. A, 1541 (2018), pp. 1-11.

[5] Y. Driouche, D. Messadi
Quantitative structure–retention relationship model for predicting retention indices of constituents of essential oils of Thymus vulgaris (Lamiaceae)
J. Serb. Chem. Soc., 84 (4) (2019), pp. 405-416.

[6] R. Bouwmeester, L. Martens, S. Degroeve
Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction
Anal. Chem., 91 (5) (2019), pp. 3694-3703.

[7] J. Ghasemi, S. Saaidpour
QSRR prediction of the chromatographic retention behavior of painkiller drugs
J. Chromatogr. Sci., 47 (2) (2009), pp. 156-163.

[8] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová
Machine learning and the physical sciences
Rev. Mod. Phys., 91 (4) (2019), Article 045002.

[9] P. Baldi, P. Sadowski, D. Whiteson
Searching for exotic particles in high-energy physics with deep learning
Nat. Commun., 5 (1) (2014), p. 430.

[10] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, N.S. Philip
Transient classification in LIGO data using difference boosting neural network
Phys. Rev. D, 95 (10) (2017), Article 104059.

[11] C. Dreissigacker, R. Sharma, C. Messenger, R. Zhao, R. Prix
Deep-learning continuous gravitational wave
Phys. Rev. D, 100 (4) (2019), Article 044009.

[12] G.B. Goh, C. Siegel, A. Vishnu, N.O. Hodas, N. Baker, Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR model, arXiv preprint arXiv:1706.06689, 2017. doi:10.48550/arXiv.1706.06689.

[13] W. Ma, F. Cheng, Y. Liu
Deep-learning-enabled on-demand design of chiral metamaterials
ACS Nano, 12 (6) (2018), pp. 6326-6334.

[14] Z. Zhang, J.A. Schott, M. Liu, H. Chen, X. Lu, B.G. Sumpter, J. Fu, S. Dai
Prediction of carbon dioxide adsorption via deep learning
Angew. Chem., Int. Ed., 131 (1) (2019), pp. 265-269.

[15] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh
Machine learning for molecular and materials science
Nature, 559 (7715) (2018), pp. 547-555.

[16] D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton, A. Agrawal
ElemNet: deep learning the chemistry of materials from only elemental composition
Sci. Rep., 8 (1) (2018), p. 17593.

[17] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli
Insightful classification of crystal structures using deep learning
Nat. Commun., 9 (1) (2018), p. 2775.

[18] J. Wei, X. Chu, X. Sun, K. Xu, H. Deng, J. Chen, Z. Wei, M. Lei
Machine learning in materials science
InfoMat, 1 (3) (2019), pp. 338-358.

[19] J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K.A. Persson, G. Ceder, A. Jain
Unsupervised word embeddings capture latent knowledge from materials science literature
Nature, 571 (7763) (2019), pp. 95-98.

[20] A. Agrawal, A. Choudhary
Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science
APL Mater., 4 (5) (2016), Article 053208.

[21] K. Schwab, "The fourth industrial revolution. Foreign Affair", 2015. [Online]. Available: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

[22] J. Prost, P.G. de Gennes
The Physics of Liquid Crystals, Oxford University Press (1995).

[23] R.S. Zola, L. Evangelista, Y.C. Yang, D.K. Yang
Surface induced phase separation and pattern formation at the isotropic interface in chiral nematic liquid crystals
Phys. Res. Lab., 110 (5) (2013), Article 057801.

[24] Chaotropic chromatography
Advances in Chromatography (1st Ed.), CRC Press (2024), pp. 1-43.

[25] H.Y. Sigaki, E.K. Lenzi, R.S. Zola, M. Perc, H.V. Ribeiro
Learning physical properties of liquid crystals with deep convolutional neural networks
Sci. Rep., 10 (1) (2020), p. 7664

[26] F.A. Khodja, P. Sassiat, M. Hanafi, D. Thiebaut, J. Vial
A promising metastable liquid crystal stationary phase for gas chromatography
J. Chromatogr. A, 1616 (2020), Article 460786.

[27] M. Dahmane, F. Athman, S. Boudah, S. Sebih, M.H. Guermouche, J.-P. Bayle
End group effect on the thermal and gas chromatographic properties of some para-substituted non-symmetrical nematogens
Chromatographia, 79 (2016), pp. 885-902.

[28] F. Athman, M. Dahmane, S. Boudah, M.H. Guermouche, J.-P. Bayle, S. Sebih
Evaluation of thermal and analytical properties of two liquid crystals in capillary GC
Chromatographia, 70 (2009), pp. 503-510.

[29] A. Addoun, O. Ferroukhi, M. Dahmane, S. Guermouche, J.-P. Bayle, M.H. Guermouche
Three nematogen azobenzene-based stationary phases for capillary GC: synthesis and comparative study
Chromatographia, 77 (2014), pp. 1367-1377.

[30] E.S. Fedorova, D.D. Matyushin, I.V. Plyushchenko, A.N. Stavrianidi, A.K. Buryak
Deep learning for retention time prediction in reversed-phase liquid chromatograph
J*. Chromatogr. A*, 1664 (2022), Article 462792.

[31] Y. Kim, C. Lim, J. Lee, S. Kim, S. Kim, D.H. Seo
Chemistry-informed machine learning: using chemical property features to improve gas classification performance
Chemom. Intell. Lab. Syst., 237 (2023), Article 104808.

[32] M. Obradović, A. Stavrianidi, E. Fedorova, A. Bogojević, O. Shpigun, A. Buryak, S. Lazović
A comparative study of the predictive performance of different descriptor calculation tools: molecular-based elution order modeling and interpretation of retention mechanism for isomeric compounds from METLIN database
J. Chromatogr. A, 1719 (2024), Article 46473.

[35] H.P. Keller, F. Erni, H.R. Linder, R.W. Frei
Dynamic slurry packing technique for liquid chromatography columns
Anal. Chem., 49 (13) (1977), pp. 1958-1963.

[34] M. Migno, A. Tchapla, O. Mercie, N. Couvrat, S. Tisse, P. Cardinael
High-density octadecyl chemically bonded core–shell silica phases for HPLC: comparison of microwave-assisted and classical synthetic routes, structural characterization and chromatographic evaluation
Chromatographia, 77 (2014), pp. 1577-1588.

[35] M. Mignot, C. De Saint Jores, A. Tchapla, F. Boyer, P. Cardinael, V. Peulon-Agasse
New anthracenyl polar embedded stationary phases with enhanced aromatic selectivity, a combined experimental and theoretical study: part 1-experimental study
J*. Chromatogr. A*, 1512 (2017), pp. 9-21.

[36] Y. Yang, J. Chen, X. Liang, B. Liu, K. Quan, X. Liu, H. Qiu
Adjustable chromatographic performance of silica-based mixed-mode stationary phase through the control of co-grafting amounts of imidazole and C18 chain
J. Chromatogr. A, 1722 (2024) Article 464889.

[37] D. Belaïdi, S. Sebih, M.H. Guermouche, J.-P. Bayle, S. Boudah
Étude par chromatographie en phase gazeuse de nouveaux composés cristaux liquides nématiques à chaîne latéral
C. R. Chim., 5 (8–9) (2002), pp. 591-598.

[38] X. Domingo-Almenara, C. Guijas, E. Billings, J.R. Montenegro-Burke, W. Uritboonthai, A.E. Aisporna, E. Chen, H.P. Benton, G. Siuzdak
The METLIN small molecule dataset for machine learning-based retention time prediction
Nat. Commun., 10 (1) (2019), p. 5811.

[39] R. Disela, T. Neijenhuis, O.Le Bussy, G. Geldhof, M. Klijn, M. Pabst, M. Ottens
Experimental characterization and prediction of Escherichia coli host cell proteome retention during preparative chromatography
Biotechnol. Bioeng., (2024), 121 (12), pp. 3848-3859.

[40] M. Ouabane, K. Tabti, H. Hajji, M. Elbouhi, A. Khaldan, K. Elkamal, A. Sbai, M.A. Ajana, C. Sekkate, M. Bouachrine, T. Lakhlifi
Structure-odor relationship in pyrazines and derivatives: a physicochemical study using 3D-QSPR, HQSPR, Monte Carlo, molecular docking, ADME-Tox, and molecular dynamics
Arab. J. Chem., 16 (11) (2023), Article 105207.

[41] A. Ilemobayo, O. Justus Durodola, O. Alade, J.O. Awotunde, A.T. Olanrewaju, O. Falana, A. Ogungbire, A. Osinuga, D. Ogunbiyi, A. Ifeanyi, I.E. Odezuligbo, O. Edu
Hyperparameter tuning in machine learning: a comprehensive review
J. Eng. Res. Rep., 26 (6) (2024), pp. 388-395.