Egypt. J. Chem. **Vol. 64,** No. 2 pp. 693 - 701 (2021)

Egyptian Journal of Chemistry

http://ejchem.journals.ekb.eg/

Principal Component Analysis Based Solvent Map for Optimisation of Rate and Yield of Curcumin Synthesis

Vinh Tien Nguyen[a], Hao Minh Hoang[a]*

[1]*Department of Chemical Technology, Faculty of Chemical and Food Technology, Ho Chi Minh City University of Technology and Education, 01 Vo Van Ngan Street, Ho Chi Minh City, Vietnam*.

Abstract

Curcumin, a diarylheptanoid normally isolated from *Curcuma longa* with multiple bioactivities, was synthesized from vanillin and pentane-1,4-dione as initial substrates, *n*-butylamine as a catalyst, boron oxide as a protecting agent, and tri-*n*-butyl borate as a water scavenger. The solvents play crucial roles in the reaction rate and the yield of curcumin synthesis. Applying principal component analysis (PCA) method on 89 molecular descriptors calculated for a set of 272 organic solvents, a two-dimensional solvent map was established. Two first principal components, which were found to characterize polarity and polarizability of solvents, accounted for 60% of the data variation and were used as the ordinates of the solvent map. Five solvents, including *n*-hexane, *N*,*N*-dimethylacetamide, *n*-butyl acetate, chloroform and isopropanol, were selected from different areas of the solvent map to investigate the solvent effects on the rate and the yield of curcumin synthesis. The experimental results revealed that the highest formation rate and yield of curcumin were obtained in *N*,*N*-dimethylacetamide. The results offered direction for further explorations of the solvent map, in which *N*,*N*-dimethylacetamide should be used as a reference to find the optimum solvent for the synthetic procedure of curcumin.

Keywords: PCA; solvent map; molecular descriptors; curcumin synthesis.

1. Introduction

Curcumin (diferuloylmethane)-(1,7-bis (4-hydroxy3-methoxyphenyl)-1,6-heptadiene-3,5-dione) is a well-known compound isolated from turmeric and possesses a wide range of pharmacological activities including anti-inflammatory, anti-cancer, anti-oxidant, wound healing, anti-microbial and anti-Parkinson effects [1-5]. The isolation of pure curcumin from turmeric rhizome is difficult due to the complexity of the turmeric extract that required large amounts of solvents for multiple extraction and recrystallization steps [6]. The inefficiencies and high costs in curcumin isolation and purification led to attempts to synthesize this compound. By using an aldol condensation reaction, curcumin was prepared from vanillin and pentane-1,4-dione as starting materials [7, 8]. The reaction rate and the yield of curcumin synthesis were dependent on several factors including catalyst, solvent, protecting agent, temperature and water scavenger. The solvent screening were carried out ussually by an experimental process of trial-and-error. This work required much time due to many experiments, and therefore has missed some actually optimum solvents among possible solvents [9, 10].

Principal component analysis (PCA) is a statistical procedure to reduce the number of dimensions of data set, thus enabling visualization the data in a 2- or 3-dimensional space [7]. PCA converts a set of solvents possessing a large number of solvent properties into a much smaller set of main values, or principal components. These components then can be used as ordinates to build a solvent map, in which solvents with similar properties are grouped together [7, 11, 12]. The bulk physicochemical properties (descriptors) of solvents for establishing a PCA

*Corresponding author e-mail: haohm@hcmute.edu.vn; (Hao Minh Hoang).
Receive Date: 23 June 2020, Revise Date: 22 July 2020, Accept Date: 06 October 2020
DOI: 10.21608/EJCHEM.2020.33197.2700
©2021 National Information and Documentation Center (NIDOC)

solvent map can be obtained by experiments and/or theoretical calculations [12, 13]. Based on the PCA results, the optimisation of solvent via design of experiments (DoE) can be carried out by selecting solvents from different areas (e.g., one solvent from each corner/vertex with a suitable centre point) of the solvent map. The DoE enables modelling the effect of each principle component on the reaction outcome. Consequently, a solvent area for a specific reaction can be screened and finally, the most suitable solvent can be obtained by a more focused study of solvents within that area [7, 14].

In our present work, a set 272 organic solvents and their molecular descriptors were used to establish a PCA solvent map and five solvents were then selected from this solvent map to investigate their effects on kinetics and reaction yield of the synthesis of curcumin. The best solvent among them can then be used as a starting point for further exploring better solvents in its proximity in the solvent map. This approach can be applied for solvent screening for new synthetic reactions, where the reaction mechanism was unknown, thus reducing the costs of solvent optimization by trial-and-error approach.

2. Experimental

2.1. Materials

Vanillin (99%, Acros Organics), 2,4-pentanedione (99.5%, Acros Organics), Boron oxide (99%, Acros Organics) and n-butylamine (99.5%, Acros Organics) were used as received. N,N-dimethylacetamide (DMAC, 99.5%, Fisher), chloroform (CT, 99.8%, Fisher), isopropanol (iPr, 99.5%, Acros Organics), n-hexane (HEX, 99.8%, Fisher) and n-butyl acetate (BuAc, 99.8%, Fisher) used as solvents and stored in 2.5 L glass bottles stored under nitrogen atmosphere.

2.2. Solvent descriptor calculation and PCA map building

A set of 272 solvents including different classes of organic solvents was selected to establish a solvent space (see Supporting Information). Their chemical structures were drawn by using MOE 2008.10 software. A database containing 164 2D-physicochemical descriptors of all solvents was obtained by using the ribbon (Computer\Descriptors) in MOE 2008.10. All descriptors are numerical values (see Supporting Information). Descriptors having the same values were screen out by Microsoft Excel. The numerical values of 89 remaining descriptors were normalized and imported into Minitab 19.2.0 software to conduct PCA analysis. The first two principal components (PC1 and PC2) from a multidimensional space constituting new variables (principal components) were used to plot a 2D-map of the 272 solvents.

2.3. Curcumin synthesis procedure and kinetic calculations

A total synthetic procedure of curcumin was carried out as follows [8]: A 100 mL three-neck round-bottom flask was charged with 1.218 g (17.49 mmol) of boron oxide and 1.74 g of 2,4-pentanedione (17.38 mmol). The mixture was stirred at 50 ^{o}C for 15 min and then 4.0 g of tributylborate (17.38 mmol) was added. After 5 min of stirring, a certain amount of solvent (0.319 mol) and 5.20 g (34.17 mmol) of vanillin were added. The reaction temperature was raised to 70 ^{o}C under stirring until vanillin completely dissolved. The catalyst of *n*-butylamine of 0.30 g (4.11 mmol) was subsequently added to the reaction mixture. The same speed of stirring (200 rpm) was maintained during the synthetic steps. The reaction time was counted after adding the catalyst.

At each duration time, a reaction aliquot (20 µL) was taken out and diluted with ethanol (95%). The absorbance of the solution was scanned from 250 nm to 600 nm by using a UV-vis spectrophotometer (Hitachi – Japan). The absorbance value at 425 nm was used to determine the concentration of curcumin with a standard curve (R^2 = 0.9969) built on pure curcumin (y = 0.1384x + 0.0216, where x, y are curcumin concentration (mg/L) and its absorbance at 425 nm in ethanolic solution, respectively). The initial reaction rate (*r*) was calculated from the amount of curcumin formed after 15 min: $r = X \cdot C/15$ mg·L^{-1}·min^{-1}, where C (mg/L) is the concentration of curcumin diluted in ethanol, and X is a dilution factor when the reaction mixture was added to ethanol.

The pure curcumin used to build the standard curve above was obtained according to [7] and subsequent thrice recrystallization in absolute ethanol. The identity of curcumin synthesized was confirmed by the following NMR (Fourier 80 Benchtop NMR, Bruker, Germany) and MS (Orbitrap Eclipse Tribrid MS, ThermoFisher Scientific) analyses results.

3. Results and discussion

Identity and purity of the synthesized curcumin were confirmed based on the following data on appearance, MS, melting point, 1H-NMR, and 13C-NMR.

Curcumin: Red-orange powder, $C_{21}H_{20}O_6$ (M = 368.38 g/mole), $[M+H]^+$ = 369.0, Mp: 181-183 °C. **^{1}H-NMR** (500 MHz, $CDCl_3$): δ = 6,415 ppm (s, 1H, C_{sp}^2-H), 7,098 ppm (d, 1H, C_{sp}^2-H, *J* = 16,5 Hz), 7,281 ppm (d, 1H, C_{sp}^2-H, *J* = 16,5 Hz), 6,810 ppm (d, 1H, C_{sp}^2-H, *J* = 16,0 Hz), 6,978 ppm (d, 1H, C_{sp}^2-H, *J* = 16,5 Hz), 7,023 ppm (dd, 1H, H^{Ar}, *J* = 2,0 Hz, J = 8,0 Hz), 7,027 ppm (dd, 1H, H^{Ar}, *J* = 2,0 Hz, J = 8,0 Hz) 7,069 ppm (dd, 1H, H^{Ar}, *J* = 2,0 Hz, J = 8,0 Hz) 7,071 ppm (dd, 1H, H^{Ar}, *J* = 2,0 Hz, J = 8,0 Hz), 6,934 ppm (d, 1H, H^{Ar}, *J* = 8,0 Hz), 6,923 ppm (d, 1H, H^{Ar}, *J* = 8,0 Hz), 3,957 ppm (s, 6H, -OCH_3). **^{13}C-NMR** (125 MHz, $CDCl_3$): δ = 97,62 ppm (1C, -**C**$_{sp}^2$-H); 162,16 ppm (1C, >**C**=N); 168,52 ppm (1C, **C**-O); 108,24; 108,81; 128,23; 128,56 ppm (4C, -**C**$_{sp}^2$-H); 121,50; 121,63; 110,91; 113,88; 146,72; 146.86; 146,97; 146,86; 134,89; 135,61 ppm (12C, -**C**$_{sp}^2$-, C^{Ar}); 55,96; 55,99 ppm (2C, -O**C**H_3).

3.1. Building of the PCA solvent map and its interpretation

The results of our PCA treatment on the original area variables are given in Table 1.
The first PC explained 45.9% of the variation of the initial solvent data, and the second PC explained 14.3%. Because the next PCs explained a minor proportion of data variation, and for simplicity, we chose a PCA model based on the first two PC to describe all the solvents. This model was expressed visually by a two-dimensional plot (map) with PC1 and PC2 as axes (Figure 1).

There is considerably more variation in solvent properties in terms of PC1 with a wide distribution across solvent map (-18 <PC1 < +26), compared with PC2 (-9 < PC2 < +5). Water is a notable outlier with the lowest PC1 (highly polar), while decamethylcyclopentasiloxane $[(CH_3)_2SiO]_5$ with the highest PC1 (nonpolar). (Figure 2)

The loading plot of descriptors (Figure 2) shows that many molecular descriptors cluster on the right side of the map, meaning that they are positively correlated with PC1, which explains the high eigenvalue and high contribution (45.9 %) of PC1 to the whole variation of the data.

Analysis of the loading plot also indicates that the PC1 correlates negatively to the solvent polarity (nonpolar solvents with high PC1 values), and PC2 approximately correlates negatively with polarisability (easily polarized solvents with low PC2 value). This can be seen from the example of the homologous series of primary alcohols (Figure 3). It is well-known in organic chemistry that when the number of carbon in a hydrocarbon derivative increases, the compound becomes less polar and tends to be more polarizable [15]. This basic rule is clearly shown in the increasing trend of PC1 and decreasing trend of PC2 of these primary alcohols. (figure 3)

A basic screening of the effect of solvent on a reaction can be carried out effectively using this PCA map, and the result can indicate the areas with high reaction rate or reaction yield. In this study, we chose five solvents (red squares in Figure 4), so that they cover a relatively large area of the solvent map [14]. The choice was also made based on the low toxicity, high availability and low price of these solvents, which are important criteria when considering scaling up the synthetic process.

The chosen solvents were: n-hexane (HEX), isopropanol (iPr), carbon tetrachloride (CT), *N*,*N*-dimethylacetamide (DMAC) and butyl acetate (BA). However, when the reaction was conducted in n-hexane, no product was obtained at all investigated reaction condition. We conclude that the reaction is not possible in nonpolar solvents with low polarizability (such as hydrocarbons with low number of carbon atoms) (**Figure 4)**.

Figure 5 shows the scheme of the synthetic reaction of rosocyanine and its subsequent hydrolysis in 95% ethanol to produce curcumin. The intermediate rosocyanine is a product of Claisen-Schmidt condensation between vanillin and 2,4-pentanedione [16]. To avoid Knoevenagel condensation at C-3 of 2,4-pentanedione**,** boron oxide is used to form a complex with it in the first step. The condensation began when n-butylamine as the catalyst was added [16]. The reaction progress was monitored by recording the UV-vis spectra of the reaction mixture after diluting with 95% ethanol.(Figure 5)

Figure 6 demonstrates typical kinetics of UV-vis absorbance spectra of the reaction mixture diluted in 95% ethanol. There are two peaks in the spectrum: one at 305 nm corresponding to the excess vanillin remaining, and one at 425 nm corresponding absorption peak of curcumin [17, 18]. The peak at 305 nm decreases over time indicating consumption of vanillin over 240 min. The peak at 425 nm, in contrast, reaches the highest value after 60 min, and decreased after that. This change of 425 nm peak indicates that the amount of curcumin produced during the reaction reached a maximum value, and then was consumed due to unknown subsequent reactions [8]. (Figure 6)

3.2. Solvent effect on the reaction rate of rosocyanine formation

Figure 7 and the fact that the reaction did not proceed in n-hexane demonstrated that the reaction rate decreased in the order of solvents: DMAC > iPr ≈ CCl_4 > BuAc>> hexane. This means that the synthesis of curcumin is preferred in aprotic polar solvents, which is in accordance with the literature about Claisen-Schmidt reactions [19]. In the mechanism of this reaction type with a base catalyst, an enolate intermediate is formed and needs to be stabilized. Therefore, it is reasonable to see polar solvents (DMAC, iPr and BuAc) or nonpolar solvents with high polarizability (CT) promote the reaction of curcumin synthesis. (Figure 7)

Figure 7 also demonstrates that increasing the amount of the catalyst accelerated the reaction, but with different extents. For DMAC and BuAc as the solvents, 0.3 g of catalyst seems to be effective and economical, while for *i*Pr and CCl_4, further increase of catalyst amount should be explored for higher effectiveness.

3.3. Solvent effect on the maximum yield of curcumin

For each solvent, the highest concentration of curcumin in the UV-vis spectra was used to calculate the maximum yield of curcumin, which is considered one of the most important criteria when choosing any synthetic process. (Figure 8)

Figure 8 shows that DMAC not only favors the reaction kinetics, but also the highest yield of curcumin. This result is in accordance with the reference [8], in which DMAC was among the preferred solvents in curcumin synthesis by this method. Using the PCA map in the section above, we can further explore the other solvents around DMAC to look for an even better solvent for this reaction. This systematic approach can be applied to any other synthetic processes to optimize the solvent without using the trial-and-error approach.

Table 1. Eigenvalues, proportion of data variation explained by the first 8 PC

PC order	1	2	3	4	5	6	7	8
Eigenvalue	38.578	12.05	8.071	7.168	3.391	3.163	1.961	1.478
Proportion	0.459	0.143	0.096	0.085	0.040	0.038	0.023	0.018
Cumulative	0.459	0.603	0.699	0.784	0.824	0.862	0.885	0.903

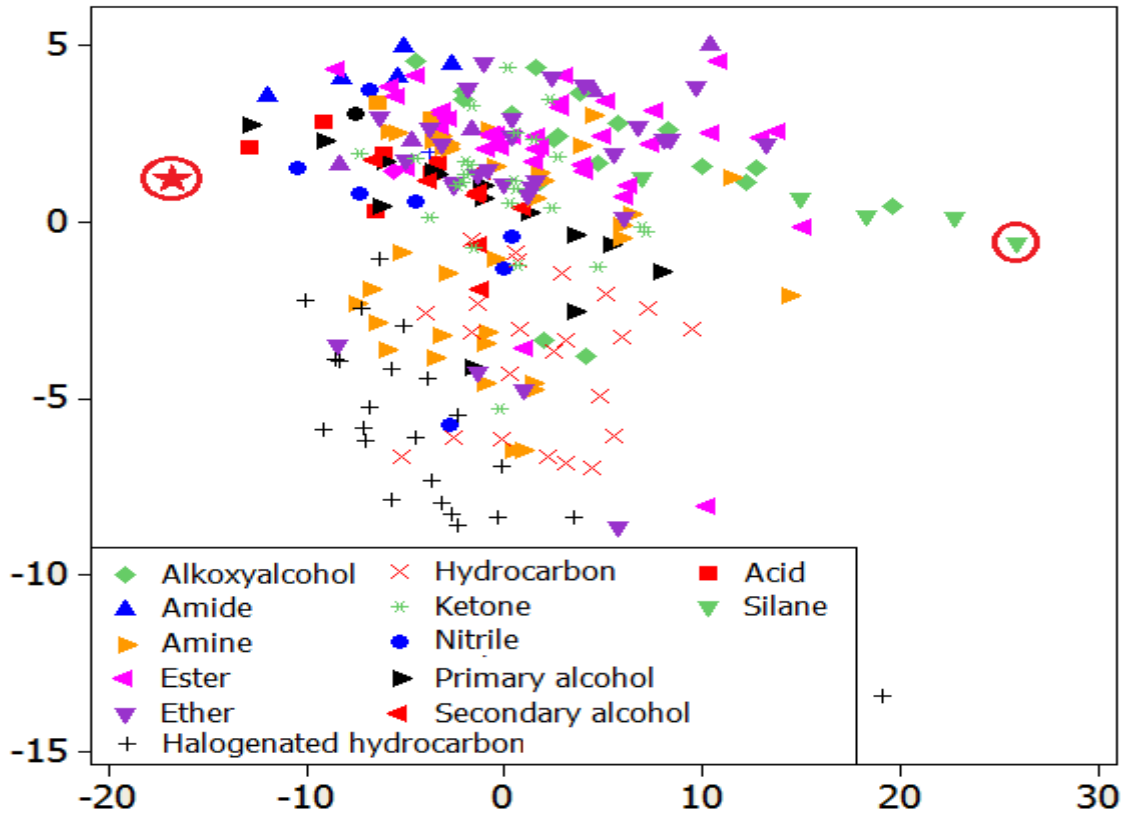


Figure 1. PCA solvent map of the 272 solvents.

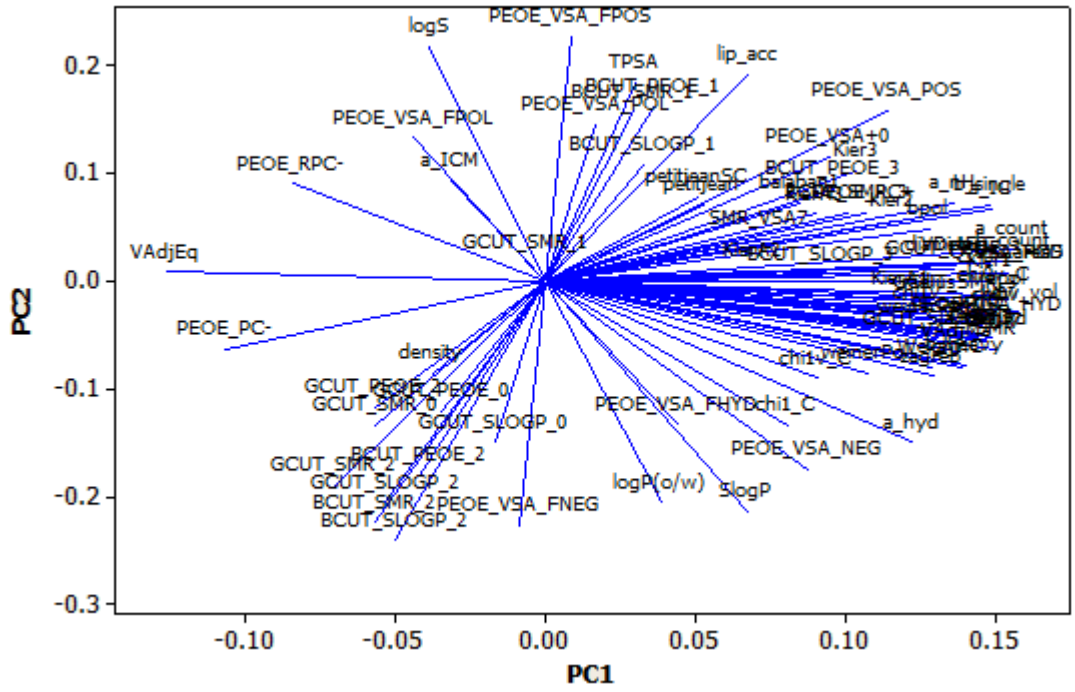


Figure 2. Loading plot of the original descriptors on PC1 and PC2.

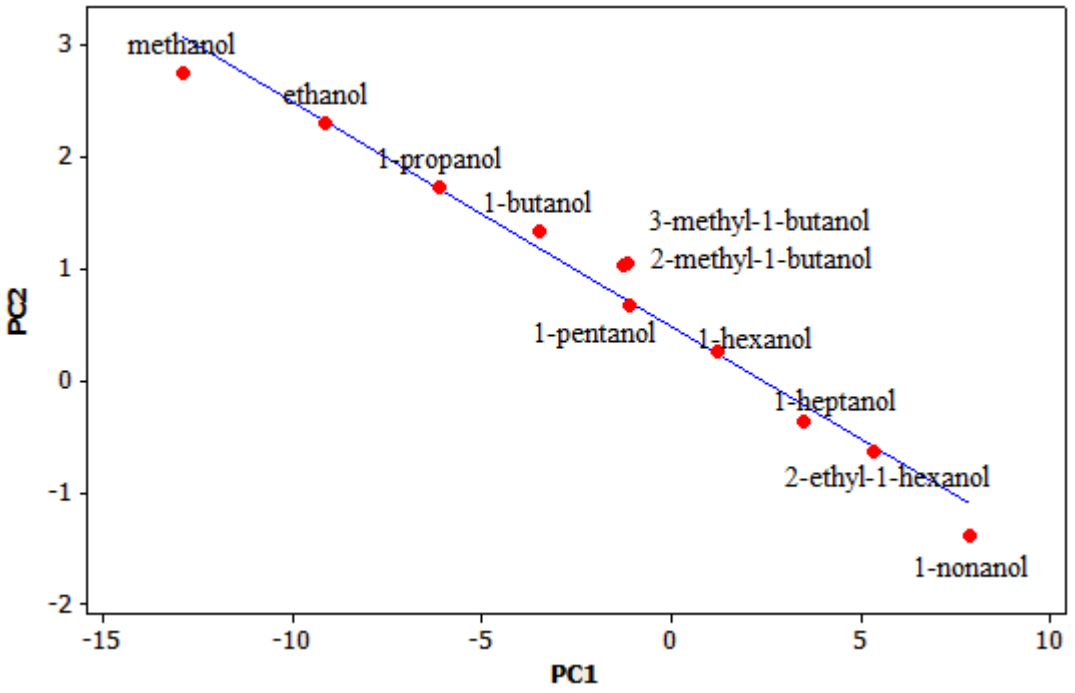


Figure 3. Homologous series of primary alcohols on the solvent map.

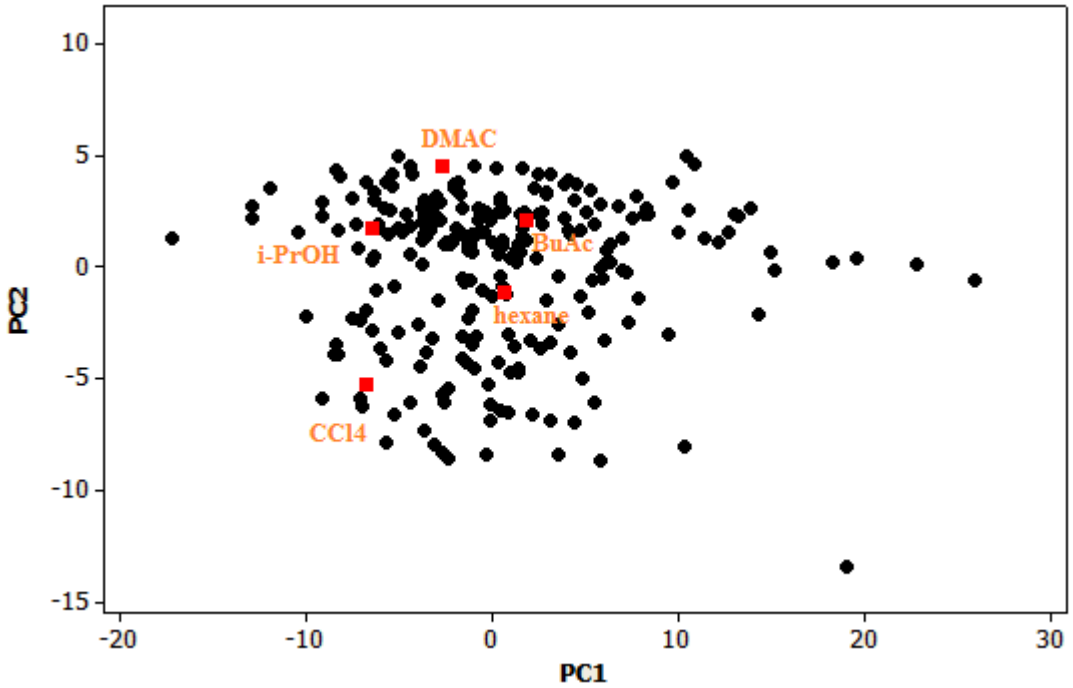


Figure 4. Five chosen solvents for the study of rosocyanine synthesis.

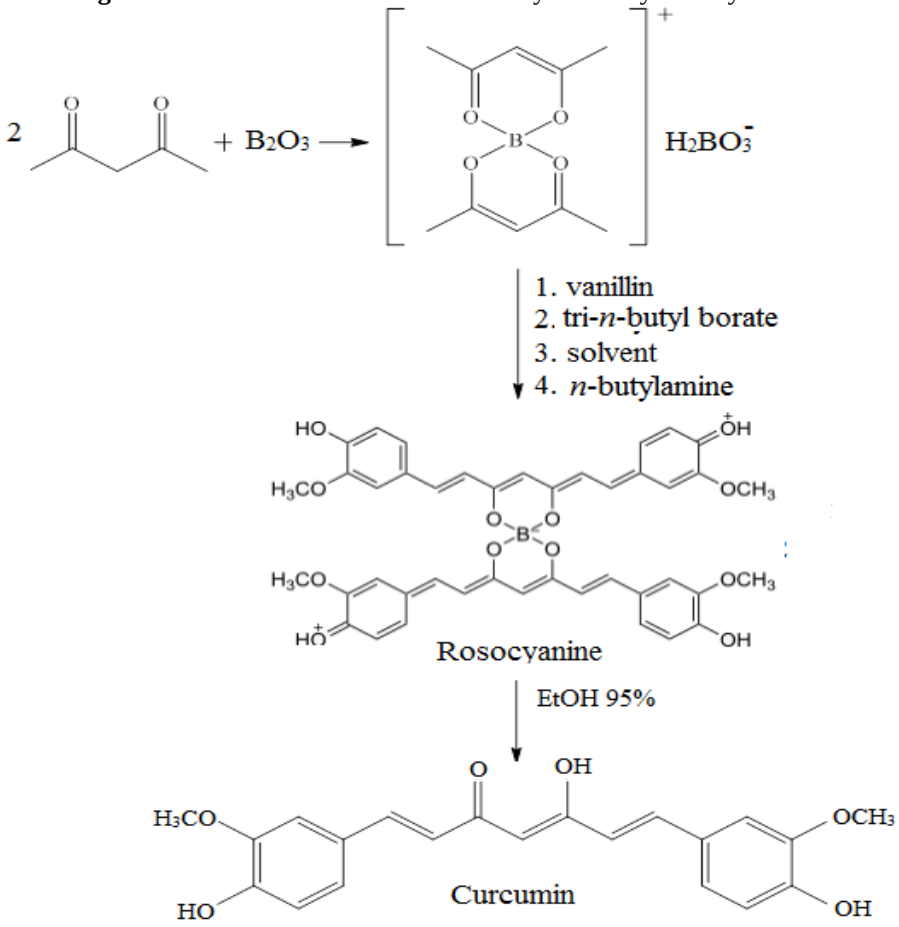


Figure 5. The reaction scheme of curcumin synthesis

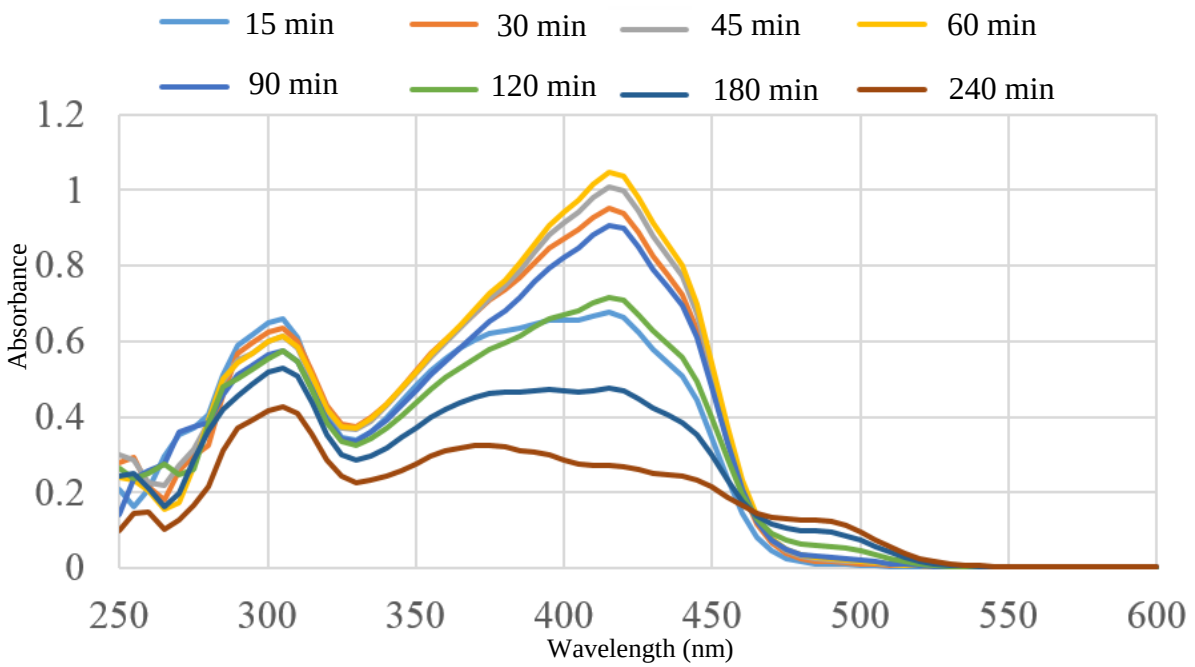


Figure 6. UV-vis spectra of the diluted reaction mixture after different time periods.

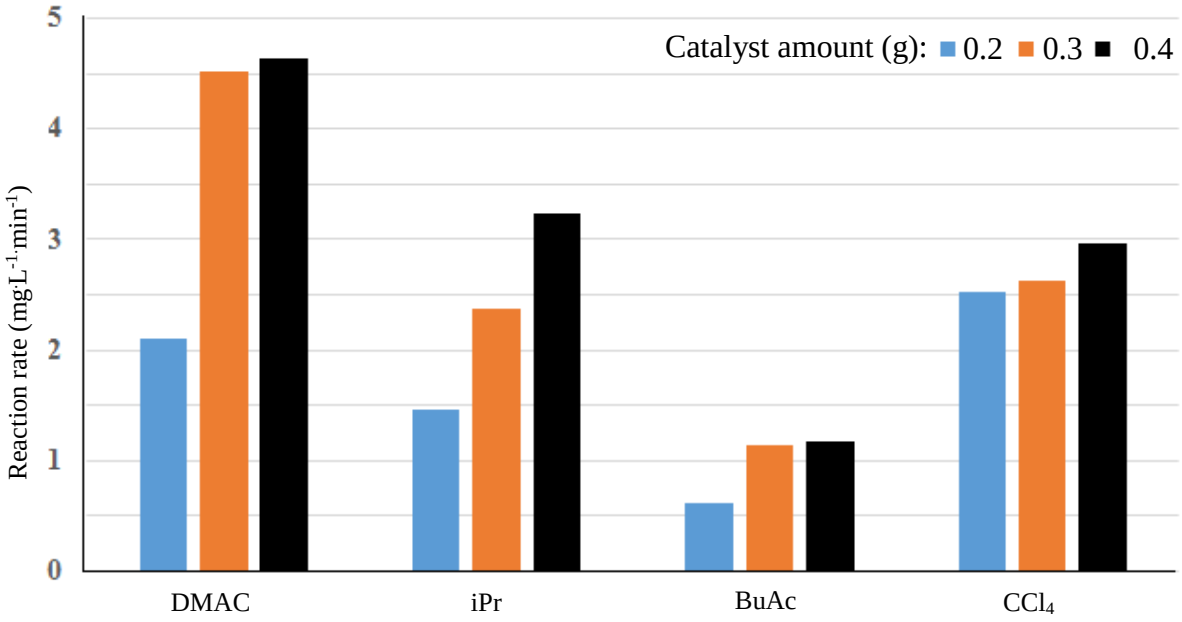
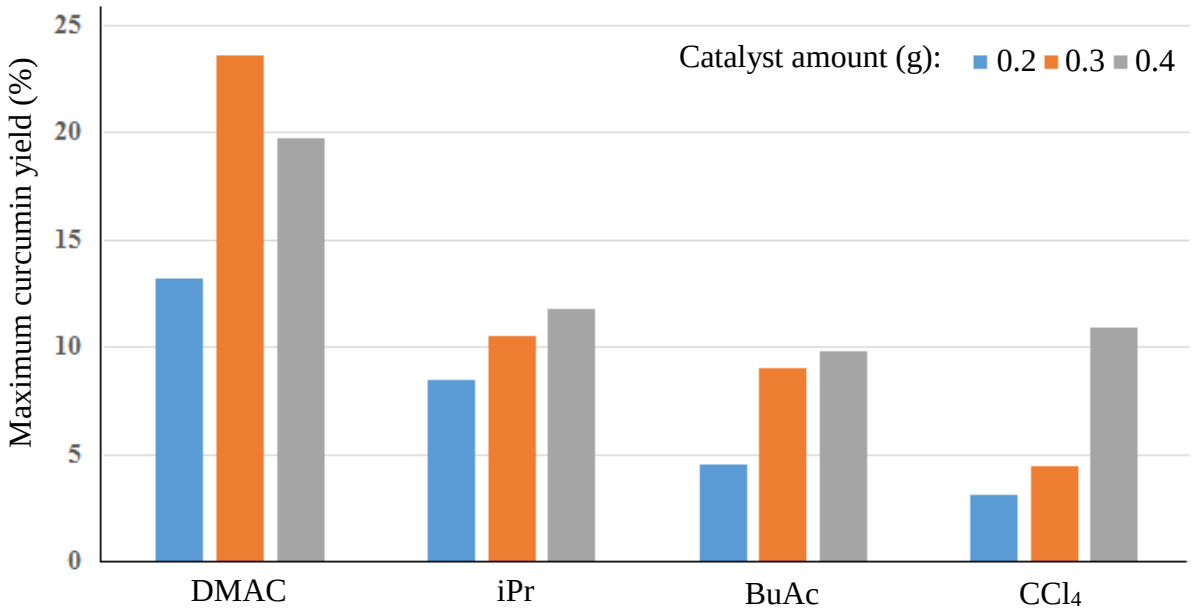


Figure 7. Effects of the solvent and the catalyst amount on the rate of rosocyanine formation.



4. Conclusion

Using the mathematical technique of PCA, we can build a solvent map from calculated molecular descriptors. This solvent map can be based on the first two PC that can explain to 60% of the data variation. The first PC is correlated with polarity, while the second PC is correlated with polarizability of the solvents. The applicability of the built solvent map in synthetic optimization was tested on the synthesis of curcumin. The obtained results demonstrated that solvents around DMAC should be further explored to find the one with the highest reaction yield.

Supporting information

The list of 272 (names and SMILE code) solvents and 89 molecular descriptors used in this study can be downloaded from the following link: http://bit.ly/SI_solvents_descriptors

The physical meaning of these molecular descriptors can be found on this website of The Chemical Computing Group: http://www.cadaster.eu/sites/cadaster.eu/files/challenge/descr.htm.

Conflicts of interest

There are no conflicts to declare.

Funding

This study was conducted without any funding source.

Acknowledgments

The authors gratefully acknowledge Ho Chi Minh City University of Technology and Education for supplying the necessary facilities and equipment to complete this study.

References

[1] S.J. Hewlings, D.S. Kalman, Curcumin: a review of its' effects on human health, Foods 6(10) (2017) 92.

[2] A. Goel, A.B. Kunnumakkara, B.B. Aggarwal, Curcumin as "Curecumin": from kitchen to clinic, Biochem. Pharmacol. 75(4) (2008) 787-809.

[3] C. Mohanty, S.K. Sahoo, Curcumin and its topical formulations for wound healing applications, Drug Discov. Today 22(10) (2017) 1582-1592.

[4] A.C. da Silva, P.D. de Freitas Santos, J.T. do Prado Silva, F.V. Leimann, L. Bracht, O.H. Gonçalves, Impact of curcumin nanoformulation on its antimicrobial activity, Trends Food Sci. Technol 72 (2018) 74-82.

[5] X.-S. Wang, Z.-R. Zhang, M.-M. Zhang, M.-X. Sun, W.-W. Wang, C.-L. Xie, Neuroprotective properties of curcumin in toxin-base animal models of Parkinson's disease: a systematic experiment literatures review, BMC Complementary Altern. Med. 17(1) (2017) 412.

[6] K.I. Priyadarsini, The chemistry of curcumin: from extraction to therapeutic agent, Molecules 19(12) (2014) 20091-20112.
[7] R. Carlson, T. Lundstedt, C. Albano, Screening of suitable solvents in organic synthesis. Strategies for solvent selection, Acta Chem. Scand. B 39 (1985) 79-91.
[8] M.H. Krackov, H.E. Bellis, Process for the synthesis of curcumin-related compounds, Google Patents, 1997.
[9] M. Ahmed, M. Abdul Qadir, M. Imtiaz Shafiq, M. Muddassar, A. Hameed, M. Nadeem Arshad, A.M. Asiri, Curcumin: Synthesis optimization and in silico interaction with cyclin dependent kinase, Acta Pharm 67(3) (2017) 385-395.
[10] C.J. Clarke, W.C. Tu, O. Levers, A. Brohl, J.P. Hallett, Green and Sustainable Solvents in Chemical Processes, Chem. Rev. 118(2) (2018) 747-800.
[11] M. Chastrette, M. Rajzmann, M. Chanon, K.F. Purcell, Approach to a general classification of solvents using a multivariate statistical treatment of quantitative solvent parameters, J. Am. Chem. Soc. 107(1) (1985) 1-11.
[12] A.R. Katritzky, D.C. Fara, M. Kuanar, E. Hur, M. Karelson, The classification of solvents by combining classical QSPR methodology with principal component analysis, The Journal of Physical Chemistry A 109(45) (2005) 10323-10341.
[13] F. Eckert, A. Klamt, Fast solvent screening via quantum chemistry: COSMO-RS approach, AlChE J. 48(2) (2002) 369-385.
[14] P.M. Murray, F. Bellany, L. Benhamou, D.K. Bucar, A.B. Tabor, T.D. Sheppard, The application of design of experiments (DoE) reaction optimisation and solvent selection in the development of new synthetic chemistry, Org Biomol Chem 14(8) (2016) 2373-84.
[15] D.R. Klein, Organic Chemistry as a Second Language: First Semester Topics, John Wiley & Sons2019.
[16] U. Pedersen, P.B. Rasmussen, S.O. Lawesson, Synthesis of naturally occurring curcuminoids and related compounds, Liebigs Ann. Chem. 1985(8) (1985) 1557-1569.
[17] A.H. Sneharani, Curcumin–sunflower protein nanoparticles—A potential antiinflammatory agent, J. Food Biochem. 43(8) (2019) e12909.
[18] N.I.o. Standards, Technology, Chemistry WebBook, SRD 69, (2018).
[19] B. Choudary, K. Ranganath, J. Yadav, M.L. Kantam, Synthesis of flavanones using nanocrystalline MgO, Tetrahedron Lett. 46(8) (2005) 1369-1371.