Systematic discovery about NIR spectral assignment from chemical structural property to natural chemical compounds

Lijuan Ma1,2,3, Yanfang Peng4, Yanling Pei1,2,3, Jingqi Zeng5, Haoran Shen1,2,3, Junjie Cao1,2,3, Yanjiang Qiao1,2,3 & Zhisheng Wu1,2,3

Spectra-structure interrelationship is still the weakness of NIR spectral assignment. In this paper, a comprehensive investigation from chemical structural property to natural chemical compounds was carried out for NIR spectral assignment. Surprisingly, we discovered that NIR absorption frequency of the skeleton structure with sp2 hybridization is higher than one with sp3 hybridization. Specifically, substituent was another vital factor to be explored, the first theory discovery demonstrated that the absorption intensity of methyl substituted benzene at 2330 nm has a linear relationship with the number of substituted methyl C-H. The greater the number of electrons given to the substituents, the larger the displacement distance of absorption bands is. In addition, the steric hindrance caused by the substituent could regularly reduce the intensity of NIR absorption bands. Furthermore, the characteristic bands and group attribution of 29 natural chemical compounds from 4 types have been systematic assigned. These meaningful discoveries provide guidance for NIR spectral assignment from chemical structural property to natural chemical compounds.

Near-infrared (NIR) spectroscopy is fast in process, intact to sample and friendly to environment1. As a process analytical technology (PAT), it is widely applied to the qualitative and quantitative analysis in agriculture and food industry2–5. At present, its application in the field of medicine is still a hot topic. Li et al. applied NIR spectroscopy to two traditional Chinese medicine (TCM) technical processes control6, and proposed a non-destructive method to analyze Compound E Jiao oral liquid7. However, the components of Chinese materia medica (CMM) are complex. The accuracy and robustness of NIR model are critical for its industrial application. For more robust model, Zhou et al. applied the boosting partial least square (PLS) regression for a better model performance8. Filgueiras et al. used synergy interval support vector regression (siSVR) to select effective variables and obtained a more robust model9. Pan et al. found that backward interval PLS was the appropriate method for the particle size model, while synergy interval PLS was the optimal model for the lobetiolin model10. Preview works demonstrated that variable selection is indeed necessary for the NIR application in CMM11,12.

However, the variables in these studies, which are selected through chemometric method, lacked practical significance and interpretability for the material structure. Many teams13–15 have found that using different chemometric methods to process the same data set, the selected variables vary greatly. The absorbance of NIR is mainly controlled by the stretching vibration overtones and combination modes of hydrogen-containing groups (X-H) including O-H, N-H, C-H, S-H, etc.16. Hence, to make the predictive models more robust and interpretive, it is essential to assign the spectra based on the interrelation between spectra and structure17. Recently, it is delighted to see that there are certain sporadic reports about these. The C-H and O=C hydrogen bonding were studied to explore the isothermal crystallization kinetics of poly (3-hydroxybutyrate)18. Czarnecki et al. reported a great deal of researches on the relationship between spectra and structure19–21. They summarized a large number of NIR absorption bands and group attribution of 29 natural chemical compounds from 4 types have been systematic assigned. These meaningful discoveries provide guidance for NIR spectral assignment from chemical structural property to natural chemical compounds.

1Beijing University of Chinese Medicine, Beijing, 102488, China. 2Key Laboratory of TCM-information Engineering of State Administration of TCM, Beijing, 102488, China. 3Beijing Key Laboratory for Basic and Development Research on Chinese Medicine, Beijing, 102488, China. 4Hubei University of Chinese Medicine, Hubei, 430065, China. 5Fujian University of Traditional Chinese Medicine, Fujian, 350122, China. Lijuan Ma and Yanfang Peng contributed equally. Correspondence and requests for materials should be addressed to Y.Q. (email: yjqiao@263.net) or Z.W. (email: wzs@bucm.edu.cn)
literature and emphasized the significance of interpretation of spectra-structure interrelationship for spectral assignment. Our previous researches also illustrated that spectral assignment was of great significance for NIR modeling. The NIR spectral assignments of Yunkang Oral Liquid and *Lonicera japonica* have been performed and made the models more accurate.

Spectral assignment methods include chemometric, quantum mechanical calculation, two-dimensional correlation spectroscopy, and so on. The second derivative (2nd) spectra and difference spectra (DS) can overcome the deficiency of the original spectrum to some extent. They are mainly used to eliminate the noise of the original spectrum and enhance the spectral resolution. Mathian *et al.* analyzed the detection limit of the major types of hypogene phyllosilicates by the NTR 2nd spectra. Gezici *et al.* pointed out the model performance of the characteristic band selected by the DS method was significantly better than that of the full spectrum.

**Figure 1.** The NIR raw spectra and 2nd spectra of benzene and cyclohexane with same molarity. There were obvious absorption differences in overtone region between benzene and cyclohexane have been presented. In addition, benzene has a series of absorption peaks at 2130 nm and 2460 nm.

**Figure 2.** NIR raw spectra of carbon tetrachloride with 1 mol/L toluene, xylene, ethylbenzene, and mesitylene. There was a strong absorption peak centered at 2330 nm.

| Compounds    | Absorbance |
|--------------|------------|
|              | 1670 nm | 2130 nm | 2330 nm | 2460 nm |
| Benzene      | 0.4410  | 0.5746  | —       | 2.0124  |
| Ortho-xylene | 0.2581  | 0.3598  | 1.0287  | 0.6920  |
| Meta-xylene  | 0.2926  | 0.3456  | 1.0546  | 1.0973  |
| Para-xylene  | 0.3156  | 0.4980  | 1.0237  | 1.1597  |
| Toluene      | 0.3122  | 0.4169  | 0.4775  | 1.1594  |
| Ethylbenzene | 0.3082  | 0.3963  | 0.6867  | 1.2853  |
| Mesitylene   | 0.4017  | 0.3846  | 1.6218  | 1.0621  |

**Table 1.** The absorption intensity corresponding to the absorption peaks of benzene and its methyl substitutes at 1670 nm, 2130 nm, 2330 nm, and 2460 nm.
Comparatively speaking, quantum mechanical method can quickly obtain the attribution of the characteristic group by combining the experimental NIR spectra and the theoretical frequency\textsuperscript{29}. DMol3 is the fastest quantum mechanical calculation method for molecular density functional theory (DFT) due to its unique advantages in handling electrostatics\textsuperscript{30}. The DFT calculation under the Materials Studio Dmol3 library have been used to investigate the structure, thermodynamics and chemical properties of the C-120-nanostructure modified with Ti atoms\textsuperscript{31}. Verissimo et al.\textsuperscript{32} also applied the DFT calculations to conduct a study on the structure of ethambutol.

Simultaneously, for complex compounds, the one-dimensional spectra mentioned above have limitation for further spectra-structure analysis\textsuperscript{33}. The two-dimensional correlation spectrum (2D-COS) was extended to “Generalized Two-Dimensional Correlation Analysis” by Noda\textsuperscript{34}. 2D-COS can assign more small and weak bands, which are masked by the autocorrelation peaks on the two-dimensional scale\textsuperscript{35}. The molecular structures (especially the hydrogen bond) of methanol and ethanol have been analyzed by NIR and 2D-COS\textsuperscript{36}. Two-dimensional correlation analysis was carried out to investigate the variation of sugars and water involved in the osmo-dehydration process\textsuperscript{37}. The experimental 2D-COS patterns have been successfully reproduced to monitor the effects of conformational isomerism to the shape of NIR spectra\textsuperscript{38}. 2D-COS coupled with Fourier transform infrared the changes of different organic constituents in landfill leachate were tracked in Fenton oxidation.

**Figure 3.** The partially enlarged NIR 2nd spectra of benzene and toluene, xylene, ethylbenzene, and mesitylene. Figure (a) showed two strong absorption peaks at 1138 nm and 1190 nm, respectively. In Figure (b), the absorption peak of the benzene ring skeleton at 1670 nm moved to the long wavelength direction by 30 nm. Figure (c) showed four sets of absorption peaks centered at 2150 nm. Figure (d) showed the absorption peak of the methyl group in the combined frequency region, in which, there is no regularity. Figure (e) showed a set of methyl absorption peaks centered at 2460 nm.

**Figure 4.** NIR DS between methyl substituted benzene and benzene. Compared with benzene, xylene, toluene, ethylbenzene and mesitylene had lower absorption intensities at 2460 nm, 2150 nm, 1670 nm and 1138 nm, and conversely higher at 1190 nm, 1400 nm, 1760 nm, 2330 nm and 2480 nm.
processes\textsuperscript{39}. Recently, Noda has further proposed the two-trace two-dimensional (2T2D) correlation and discussed its potential application\textsuperscript{40}.

Unfortunately, these spectra-structure studies usually focus on hydrogen-bonding\textsuperscript{41} and the interaction of intermolecular\textsuperscript{42}, although some about micro-heterogeneity\textsuperscript{43}. Chemical structural properties, such as the hybridization type of X (C, N, O, and S), the presence or absence of compound substituents, and the steric hindrance, have not been explored. In addition, the influence of solvent is a tough challenge in application of spectral assignment methods and needs to be discussed. Moreover, the spectral assignment of complex natural chemical compounds is necessary but insufficient, such as mosses, phenylpropanoids, alkaloids and steroids\textsuperscript{44}.

Accordingly, deuterated reagent was used to explore the effects of key factors of specific compounds with chemical structural properties, including atomic hybridization mode, quantity and position of substituents, and steric hindrance. Besides, the identification of characteristic variables was conducted by experiments and theoretical calculations. In addition, characteristic variables of mosses, phenylpropanoids, alkaloids and steroids have been explored. Furthermore, in order to get more detailed spectra-structure information, 2D-COS has been proposed for group attribution with concentration as the interference term. From chemical structural property to natural chemical compounds, this paper provides a systematic and valuable discovery for NIR spectral assignment.

Results

Near infrared (NIR) spectral assignment of hybridization type. There are differences in the NIR characteristic bands of substances with different structures, and the hybridization type is a critical factor that affects the diversity in absorption. Supplementary Fig. S1 shows the NIR raw spectrum of benzene and cyclohexane, from which, the third overtone absorption of the methylene in cyclohexane at 890 nm, the second overtone absorption at 1210 nm, the second combination mode at 1400 nm, and the first overtone absorption peak at 1760 nm has been assigned.

Another discovery is that the C-H of backbone structure of benzene has the third overtone, the second overtone, and the first overtone absorption respectively at 880 nm, 1140 nm, and 1660 nm. These results indicate that the absorption frequency of C with sp\textsuperscript{2} hybridization (benzene) is higher than one with sp\textsuperscript{3} hybridization (cyclohexane).

In order to exclude the difference in absorption intensity caused by the uneven number of molecules, further studies were carried out using the same molar concentration of benzene and cyclohexane. The NIR raw spectra and 2nd spectra of benzene and cyclohexane with same molarity were shown in Fig. 1, in which, obvious absorption differences in overtone region between benzene and cyclohexane have been presented.

We also found that benzene has a series of absorption peaks related to C-H and C-C stretching vibrations in the combination mode 2130 nm and single-strong absorption peak caused by C-H stretching and bending vibration at 2460 nm. In terms of cyclohexane, there were four strong absorption peaks centering 2400 nm, which was the result of combination mode of C-H stretching and bending vibration. The results demonstrated that the C-H absorption frequency of benzene is higher than one of cyclohexane with a smaller bond force constant, proving again that the carbon atom with sp\textsuperscript{2} hybridization type had a larger absorption frequency. The assignment of hybridization type is significantly meaningful theory for a wider range of application of NIR spectroscopy both in production and clinical diagnosis.

Owing to the same hybridization type and skeleton structure, the identification of homologues is another difficulty in spectra-structural analysis. The NIR raw spectra of carbon tetra chloride solutions of different methyl-substituted benzene were displayed in Fig. 2. It is noteworthy that the absorption peak of the benzene ring skeleton C-H is not covered by the methyl C-H absorption, but its absorption intensity ratio changes. Additionally, a strong absorption peak centered at 2330 nm appears in the NIR spectra of methyl-substituted benzene.
benzene, at where the absorption peak is the combined frequency absorption of methyl stretching vibration. This explanation provided another effective variable for identification of methyl substitution.

**NIR spectral assignment about the effects of the substitutes on core structure.** For further quantitative research, the absorption intensity corresponding to the absorption peaks of benzene and its methyl

---

| Compounds       | Chemical structure | Characteristic bands (nm) | Fundamental frequency | First double-frequency | Second double-frequency | Third double-frequency | Attribution of characteristic bands |
|-----------------|--------------------|---------------------------|-----------------------|------------------------|------------------------|------------------------|---------------------------------------|
| Methanol        | ![Methanol structure](image) |                           | 3199–3338             | 1649–1964              | 1111–1391              | 842–1113               | C-H of methyl                        |
|                 |                    |                           | 2593                  | 1336–1525              | 900–1080               | 682–864                | O-H                                  |
| Ethanol         | ![Ethanol structure](image) |                           | 3348                  | 1726–1970              | 1163–1395              | 881–1116               | C-H of methylene                   |
|                 |                    |                           | 3310                  | 1706–1947              | 1149–1379              | 871–1103               | C-H of methyl                       |
|                 |                    |                           | 3249                  | 1675–1911              | 1128–1354              | 855–1083               | C-H of methyl and methylene         |
|                 |                    |                           | 3228                  | 1644–1899              | 1121–1345              | 850–1076               | C-H of methyl and methylene         |
|                 |                    |                           | 3217                  | 1658–1892              | 1117–1340              | 847–1072               | C-H of methyl                       |
|                 |                    |                           | 2604                  | 1342–1531              | 904–1085               | 685–868                | O-H                                  |
| Benzene         | ![Benzene structure](image) |                           | 3122–3163             | 1609–1860              | 1084–1318              | 822–1054               | C-H of benzene ring                 |
| Toluene         | ![Toluene structure](image) |                           | 3165–3305             | 1661–1944              | 1119–1377              | 848–1102               | C-H of methyl                       |
|                 |                    |                           | 3120–3165             | 1608–1862              | 1083–1319              | 821–1055               | C-H of benzene ring                 |
| Ethyl-benzene   | ![Ethyl-benzene structure](image) |                          | 3307                  | 1705–1945              | 1148–1378              | 870–1102               | C-H of methylene                   |
|                 |                    |                           | 3293                  | 1697–1937              | 1143–1372              | 867–1098               | C-H of methylene                   |
|                 |                    |                           | 3268                  | 1685–1922              | 1135–1362              | 860–1089               | C-H of methylene                   |
|                 |                    |                           | 3223                  | 1660–1896              | 1118–1343              | 847–1074               | C-H of methyl                       |
|                 |                    |                           | 3128–3172             | 1612–1866              | 1086–1322              | 823–1057               | C-H of benzene ring                 |
| Para-xylene     | ![Para-xylene structure](image) |                        | 3223–3306             | 1661–1944              | 1119–1377              | 848–1102               | C-H of methyl                       |
|                 |                    |                           | 3146–3168             | 1622–1864              | 1093–1320              | 828–1056               | C-H of benzene ring                 |
| Meta-xylene     | ![Meta-xylene structure](image) |                        | 3227–3253             | 1677–1957              | 1130–1386              | 856–1109               | C-H of methyl                       |
|                 |                    |                           | 3144–3171             | 1620–1865              | 1092–1321              | 827–1057               | C-H of benzene ring                 |
| Ortho-xylene    | ![Ortho-xylene structure](image) |                        | 3233–3322             | 1712–1955              | 1153–1385              | 874–1108               | C-H of methyl                       |
|                 |                    |                           | 3120–3166             | 1608–1863              | 1083–1319              | 821–1055               | C-H of benzene ring                 |
| Mesitylene      | ![Mesitylene structure](image) |                        | 3215–3222             | 1657–1954              | 1116–1384              | 846–1107               | C-H of methyl                       |
|                 |                    |                           | 3159–3198             | 1628–1881              | 1097–1332              | 831–1066               | C-H of benzene ring                 |
| Phenol          | ![Phenol structure](image) |                           | 3122–3164             | 1609–1861              | 1084–1318              | 822–1055               | C-H of benzene ring                 |
|                 |                    |                           | 2607                  | 1533                   | 1086                   | 869                    | O-H                                  |
| Benzy alcohol   | ![Benzy alcohol structure](image) |                          | 3355                  | 1867–1973              | 1136–1398              | 861–1118               | C-H of methylene                   |
|                 |                    |                           | 3123–3174             | 1610–1867              | 1085–1322              | 835–1058               | C-H of benzene ring                 |
|                 |                    |                           | 2612                  | 1546–1536              | 907–1088               | 687–871                | O-H                                  |
| Benzy alcohol   | ![Benzy alcohol structure](image) |                          | 3479                  | 1793–2046              | 1208–1450              | 916–1160               | C-H of aldehyde group               |
| Methyl (CH₃)    | ![Methyl structure](image) |                           | 3109–3479             | 1603–1853              | 1079–1313              | 818–1050               | C-H of benzene ring                 |
| Methylene (CH₂) | ![Methylene structure](image) |                           | 1111–1340             | 1649–1892              | 1894–2079              | 2214–2479              |                                       |
| Benzene ring (C₆H₆) | ![Benzene ring structure](image) |             | 1117–1176             | 1608–1863              | 2119–2192              |                                       |                                       |
| Aldehyde (CHO)  | ![Aldehyde structure](image) |                           | 1229–1294             |                       |                       |                                       |                                       |
| Hydroxyl (OH)   | ![Hydroxyl structure](image) |                           | 1358–1456             |                       |                       |                                       |                                       |

**Table 2.** The theoretical fundamental frequency, overtones, combination modes, and vibration group of the ramifications of benzene and the characteristic groups.
substitutes at 1670 nm, 2130 nm, 2330 nm and 2460 nm were calculated and shown in Table 1. Accidentally, we spied out that at 2330 nm, the absorption intensity ratio of the toluene, xylenes, and mesitylene is about 1:2:3.5. For ethylbenzene, the absorption intensity at 2330 nm is between toluene and xylene. These indicate that the absorption intensity of methyl substituted benzene at 2330 nm is linear with the number of substituted methyl C-H. Briefly, as the number of carbon-hydrogen bonds increases, the NIR absorption intensity increases proportionally. This seminal discovery provides the theoretical basis of NIR quantitative for the first time.

Furthermore, in-depth research of the differences in homologues have been carried out for more accurate quantitative analysis. Figure 3 displayed the partially enlarged NIR 2nd spectra of benzene, toluene, xylene, ethylbenzene, and mesitylene. The absorption peaks of the benzene ring skeleton C-H stretching vibration at 1138 nm and 1670 nm shifted to the long-wave direction due to the substitution of the methyl group, and the largest displacement was caused by mesitylene with 10 nm and 30 nm, respectively. Interestingly, the order of displacement of these two absorption peaks has a certain regularity, which is toluene ≈ ethylbenzene < ortho-xylene < meta-xylene < para-xylene < mesitylene. This major finding demonstrated that the inductive effect of the electron donating substituent made the benzene ring positively charged, inducing the displacement. In brief, the greater the number of electrons donating substituent, the larger the displacement distance of absorption peak is.

Compared to the 2nd method, the difference spectrum can react more quickly and directly to the effects of substituents on the basic structure. Figure 4 shows NIR DS between methyl substituted benzene and benzene. Clearly, for benzene ring C-H of toluene, xylene, ethylbenzene, and mesitylene, the absorption values of combination mode at 2460 nm and 2150 nm, the first overtone at 1670 nm and the second overtone at 1138 nm were negative.

Figure 6. Partial enlargement of the 2nd spectra of anthraquinones. There were significant differences of anthraquinones at 1371–1470, 1620–1838, 1976–2020, and 2094–2200 nm.

| Compounds         | Characteristic bands (nm)                                                                 |
|-------------------|-------------------------------------------------------------------------------------------|
| Anthraquinones    | 1371–1470, 1620–1838, 1976–2020, 2094–2200                                                |
| Lignins           | 1160–1280, 1325–1472, 1700–1840, 1970–2200                                                |
| Coumarins         | 1370–1480, 1550–1840, 2040–2200, 2301–2478                                               |
| Simple Phenylalanines | 1370–1490, 1117–1285, 1560–1830, 1975–220, 2314–2470                                     |
| Terpenes          | 940–1000, 1120–1300, 1380–1480, 1650–1840, 1880–1925, 1970–2180, 2280–2460               |
| Alkaloids         | 1110–1260, 1370–1495, 1620–1820, 2040–2195, 2300–2460                                    |
| Flavones          | 1120–1280, 1350–1550, 1620–1830, 1990–2200, 2300–2460                                    |

Table 3. NIR absorption characteristic bands of different natural chemical compounds in CMM.
In contrast, for the methyl C-H, the absorption values of combination mode at 2330 nm and 2480 nm, the second combination mode at 1400 nm, the first and second overtone at 1760 nm and 1190 nm were all positive. All these wavelengths reflect the difference between the toluene, xylene, ethylbenzene, and mesitylene. The partially enlarged DS of methyl substituted benzenes and benzene was shown in Supplementary Fig. S2. Strikingly, the orders of the overtone intensity of methyl C-H at 2330 nm and benzene C-H at 1138 nm, 1670 nm and 2130 nm, especially in the range of 780–2040 nm, all were ethylbenzene $\approx$ methylbenzene $<$ xylene $<$ mesitylene. The NIR DS between methyl substituted benzenes and toluene (see Supplementary Fig. S3) also matched this result. This was consistent with the result 2nd spectra, proving again the influence regularity of substituted methyl.

The symmetry of chemical structure is a pivotal factor affecting the NIR absorption intensity of substances. For further research, the NIRS DS of xylenes - benzene, and xylenes - toluene have been displayed in Fig. 5, and the enlarged views were shown in Supplementary Figs S4 and S5 for details. It is worth noting that the absorption intensities caused by benzene ring skeleton of xylenes at 1138 nm, 1670 nm and 2130 nm were negatively correlated with their symmetry, with the order as ortho-xylene $<$ meta-xylene $<$ para-xylene. It illustrates that the higher the chemical structure symmetry, the lower the NIR absorption intensity is.

Moreover, the characteristic bands of some simple group substitutes centered on the benzene were assigned, including ortho-xylene, meta-xylene, para-xylene, phenol, benzyl alcohol, and benzaldehyde. From the results shown in Supplementary Figs S6–S9, it could be found that 1700 nm and 2100–2500 nm could be used for O-H for NIR modeling. These results were meaningful for qualitative and quantitative analysis of isomers by NIR spectroscopy.

**Rapid NIR spectral assignments based on the theory of quantum mechanics.** Quantum mechanics theory can construct the molecular model to calculate the theoretical absorption frequency of the molecule for spectral assignment. In order to avoid the complicated experimental process brought by the 2nd spectroscopy and the difference spectrum method, combining density functional theory and the raw NIR spectra shown in Supplementary Fig. S9, we quickly obtained the theoretical fundamental frequency, overtones, combination modes, and vibration groups of the methanol, ethanol, benzene, toluene, ethylbenzene, ortho-xylene, meta-xylene, para-xylene, phenol, benzyl alcohol, and benzaldehyde shown in Table 2.

Combined with Supplementary Fig. S9 and Table 2, the attributions of the NIR absorption were easily analyzed and the results were summarized and also listed in Table 2. Amazingly, the NIR absorption peak width of the substance is independent of the number of methyl groups but related to steric hindrance. Moreover, as the increase of the steric hindrance, the absorption peak width of C-H on benzene ring became broader, with the peak width order as ortho-xylene $>$ meta-xylene $>$ para-xylene $>$ benzene. Conversely, in terms of the methyl C-H, the absorption peak shifts to a high wave number as the steric hindrance increases, and the larger the steric hindrance, the narrower the absorption peak is, which is the opposite of the benzene ring. This method provides a guideline for the application of Quantum Mechanics calculation in spectra-structure analysis.

**NIR spectral assignments of natural chemical compounds in CMM for variable selection.** As a process analytical technology, NIR is often applied for quality control of CMM production processes.
due to the complex chemical composition and similar structure of natural chemical compounds in CMM, such as phenylpropanoids, alkaloids, steroids and so on, the original spectral overlap is too serious to analysis. To overcome this problem in some extent, the 2nd spectra of chrysophanol, cryptotanshinone, tanshinone IIA, tanshinone I have been investigated. From Fig. 6, it is clear to observe the significant differences at 1371–1470, 1620–1838, 1976–2020, and 2094–2200 nm. Therefore, 1371–2200 nm could be proposed for NIR modeling as the characteristic band of anthraquinones.

Figure 8. The 2D-COS of 13 kinds of phenylalanines compounds including Magnolol, and honokiol, psoralen, isopsoralen, imperatorin, isoimperatorin, osthole, cinnamic acid, cinnamaldehyde, eugenol, schisandra A, schisandra B, and schisandra ester A.
Similarly, the other essential compound groups of natural chemical compounds including phenylpropanoids, alkaloids, moss and so on have been studied. Supplementary Figs S10 to S14 respectively show the full and enlarged view NIR 2nd spectra of simple phenylpropanoids (schizandrin A, schizandrin B, schisandracester A); Coumarins (psoralen, imperatorin, isopsoralen, isoimperatorin) lignin (honokiol); mosses (cantharidin, menthol, borneol, oleanolic acid, artemisinin, dehydroandrographolide, evodilactone, curcumin dione, curcumol, licorice) acid); alkaloids (fritigenin B, matrine, reserpine, sinomenine). The characteristic bands of each component category were summarized in Table 3, which were significant for more robust and interpretative models.

NIR spectral assignments for more details of natural chemical compounds in CMM by 2D-COS.

The 2nd can assign the characteristic band of a substance, but it is powerless for a more detailed structure analysis. 2D-COS can identify small and weak peaks that are masked in one-dimensional spectrum by subtle spectral changes caused by external disturbances. To determinate the parameter of the two-dimensional correlation analysis, firstly, the 2D-COS of six magnolol solution with gradually increasing concentrations were displayed in Fig. 7. Accidentally, the intensity of autocorrelation peaks is stronger synchronized with the increasing range of sample concentration. Accordingly, the influence of sample concentration was concluded, that is tiny sample concentrations were not suitable for 2D-COS analysis because of unrecognized autocorrelation peaks, and large sample concentrations were not recommended either for increased costs. Secondly, the 2D-COS by three types of concentration interval have been analyzed, and accidentally, we found all had 12 identical autocorrelation peaks (see Supplementary Fig. S15). Interestingly, the number of samples had the same effect on the 2D-COS as the concentration interval (see Supplementary Fig. S16). These provided a meaningful guideline for the 2D-COS analysis of the chemical compound.

Owing to the advance of the significant resolution enhancement of heavily overlapped NIR bands, the 2D-COS was proposed to the NIR spectral assignment of CMM. Through analyzing the 2D-COS of 13 phenylpropanoids shown in Fig. 8, the characteristic absorption bands of simple phenylpropanoids (eugenol, cinnamaldehyde, and...
cinnamic) have been concluded as 1620–1740 nm, 1863–1963 nm, and 2200–2500 nm, and those of Lignins as 1387–1468 nm, 1619–1714 nm, and 2200–2500 nm. By the same method, the 2D-COS of 7 alkaloid compounds and 9 terpenoid compounds were displayed in Figs 9 and 10. Obviously, the small and weak peaks hidden in one-dimensional spectra can be observed easily.

The NIR absorbance is dominated by the functional groups containing hydrogen atoms (e.g. OH, CH, NH). Accordingly, further spectra-structure research of these 29 compounds were performed, and the details information of which were concluded in Tables 4, 5 and 6, respectively. These indicated that 2D-COS could be seen as an excellent method for the NIR spectral assignment of complex natural chemical compounds.

Supplementary Table S2 showed a comparison of the analytical results of the second derivative and 2D-COS. It can be seen that the characteristic bands of several kinds of compounds assigned by these two methods were very close. For example, the characteristic bands of terpenoids assigned by the second derivative were “940–1000, 1120–1300, 1380–1480, 1650–1840, 1880–1925, 1970–2180, 2280–2460 (nm)”, while one assigned by 2D-COS were “388–1442, 1676–1761, 1864–1920, 1990–2145, 2320–2500 (nm)”. Almost all of the characteristic variables obtained by 2D-COS were within the range of the bands assigned by the second derivative spectroscopy method, which verified that the 2D-COS was more accurate.

Discussion
According to the spectra-structure interrelationship, a serial of critical results of NIR absorption bands have been revealed progressively. Firstly, for skeleton structure, sp² hybridization can induce a higher absorption frequency than sp³ hybridization. As to the influence of substituent, with the increase of the number of electrons donating substituent, absorption intensity of methyl substituted benzene increases proportionally, and the displacement distance of absorption peak became larger. In addition, as another key factor, the temporal steric could decrease the NIR absorption intensity regularly. Furthermore, not only chemical structure with specific property, but also complex natural chemical compounds, both can be assigned for more interpretative NIR model. 2D-COS was more accurate for characteristic bands and detail information. These laid a solid theoretical foundation of qualitative and quantitative analysis for more interpretative model based on spectral assignment.

Figure 10. 2D-COS of 9 kind of phenylpropanoids compounds, including menthol, curcumol, borneol, gossypol, limonin, oleanolic acid, dehydroandrographolide, artemisinin and norcantharidin.
| Category          | Chemical compound | Chemical structure | NIR absorption groups | Characteristic absorption bands (nm) | Attribution of characteristic bands | Common characteristic absorption bands (nm) |
|-------------------|-------------------|--------------------|-----------------------|--------------------------------------|-------------------------------------|---------------------------------------------|
| Simple phenylalanine | Eugenol          | ![Eugenol structure](image) | C-H in benzene ring, OCH₃, CH₃, and C=C; O-H in benzene ring | 1419–1472 1617–1782 1920–1959 1992–2500 2415–2500 | The combination absorption bands of C-H in CH₂ and O-H in benzene ring 1643–1725 2095–2373 The absorption bands of C-H in CHO 2415–2500 The absorption bands of C-H in C=C | 1620–1740 1863–1963 2000–2500 |
|                   | Cinnamaldehyde   | ![Cinnamaldehyde structure](image) | C-H in CHO, C=C and benzene ring | 1643–1725 2095–2373 2415–2500 | The absorption bands of C-H in benzene ring The absorption bands of C-H in CHO The absorption bands of C-H in C=C | 1649–19703 1865–1922 The absorption bands of C-O and O-H in carboxylic group 2073–2500 The combination absorption bands of deuterium chloroform solvent and C-H and C-C in benzene ring |
|                   | Cinnamic acid     | ![Cinnamic acid structure](image) | O-H in carboxylic group; C-H in C=C and benzene ring | 1649–1972 2095–2373 2415–2500 | The combination absorption bands of C-H in benzene ring The absorption bands of C-H in CHO The absorption bands of C-H in C=C and O-C | 1641–1782 1863–1963 2000–2500 |
|                   | Magnolol and Honokiol | ![Magnolol and Honokiol structure](image) | C-H in benzene ring, CH₃, and C=C; O-H in benzene ring | 1383–1472 1615–1739 1864–1969 2018–2500 | The combination absorption bands of C-H in benzene ring and C=C | |
| Lignin            | Schizandrin A     | ![Schizandrin A structure](image) | C-H in benzene ring, OCH₃, CH₃, and CH₂ | 1385–1452 1640–1800 1870–1915 1920–2500 | The combination absorption bands of C-H in CH₂ and CH₃ The absorption bands of C-H in OCH₃ The combination absorption bands of deuterium chloroform solvent and C-H and C-C in benzene ring | 1387–1468 1619–1714 1873–1912 2018–2180 2215–2500 |
|                   | Schizandrin B     | ![Schizandrin B structure](image) | C-H in benzene ring, OCH₃, CH₃, OCH₂, and CH₂ | 1391–1421 1640–1800 1870–1915 1920–2500 | The combination absorption bands of C-H in CH₂ and CH₃ The absorption bands of C-H in OCH₂ The combination absorption bands of deuterium chloroform solvent and C-H and C-C in benzene ring | 1387–1468 1619–1714 1873–1912 2018–2180 2215–2500 |
|                   | Schisandra ester A | ![Schisandra ester A structure](image) | C-H; O-H; and C=O in ester groups | 1376–1452 1644–1768 1874–1926 2035–2079 2115–2177 2235–2500 | The combination absorption bands of C-H in benzene ring and OCH₃ The combination absorption bands of deuterium chloroform solvent and C-H and C-C in benzene ring | |
| Continued         |                   |                    |                       |                                      |                                     |                                             |
Methods

Experimental instruments and materials. Carbon tetrachloride, cyclohexane, benzene, toluene, o-xylene, m-xylene, and p-xylene, mesitylene, and phenol were purchased from Beijing Chemical Reagent Factory. Ethylbenzene, benzyl alcohol and benzaldehyde were purchased from Tianjin Chemical Reagent Factory. Deuterated chloroform was purchased from Cambridge Isotope Laboratories, Inc. (Massachusetts, USA). All reference standards were supplied by the Chinese Food and Drug Inspection Institute (Beijing, China).

Sample preparation of NIR raw spectra. 2.2 mL of benzene, 2.6 mL of toluene, 3.0 mL of ethylbenzene, 3.1 mL of o-xylene, 3.1 mL of m-xylene, 3.1 mL of p-xylene, 3.5 mL of mesitylene, 2.5 mL of benzaldehyde, 2.6 mL of benzyl alcohol, and 2.4 g of phenol were measured precisely, and diluted with carbon tetrachloride to obtain 1 mL/L stock solution.

Furthermore, 4, 8, 12, 192, 196, 200 μL stock solution of benzene, toluene, xylenes and mesitylene were diluted respectively with carbon tetrachloride to obtain a series of solution with volume fraction from 0.2% to 10%. Similarly, 12–200 μL stock solution of methanol and ethanol were diluted respectively with carbon tetrachloride to obtain a series of solution with volume fraction from 0.6% to 10%.

The collection of NIR raw spectra, 2nd spectra and difference spectra collection. The spectra were collected by FOSS RLA holographic grating NIR spectrometer (Metrohm China, China). Spectra collection was set as transmission mode. The background was set as air inside the instrument. The average spectrum was gained by three parallel spectra collected with the resolution as 0.5 nm, scanning range as 400–2500 nm, and scanning times as 32.

The NIR difference spectra of toluene, ethylbenzene, o-xylene, m-xylene, p-xylene, mesitylene, phenol, benzyl alcohol, and benzaldehyde were respectively obtained by respective mutual subtraction of the raw spectra.

NIR 2D-COS collection. The samples were prepared according to Supplementary Table S1. The NIR raw spectra were collected by the same method above. Two-dimensional correlation analysis was carried out on the

| Category | Chemical compound | Chemical structure | NIR absorption groups | Characteristic absorption bands (nm) | Attribution of characteristic bands | Common characteristic absorption bands (nm) |
|----------|------------------|--------------------|----------------------|--------------------------------------|-------------------------------------|------------------------------------------|
|          | Psoralen and Isopsoralen | ![Psoralen structure](image) | C-H in benzene ring and C=C | 1381–1422 | The absorption bands of C-H in benzene ring | 1379–1415, 1590–1695, 1862–1920, 2046–2500 |
|          | Imperatorin and Isoperforatorin | ![Imperatorin structure](image) | C-H in benzene ring, CH3, CH2, and C=C | 1388–1413 | The absorption bands of C-H in benzene ring | 1379–1415, 1590–1695, 1862–1920, 2046–2500 |
|          | Osthole | ![Osthole structure](image) | C-H in benzene ring, OCH3, CH3, CH2, and C=C | 1328–1429 | The combination absorption bands of C-H in CH3 and CH2 | 1379–1415, 1590–1695, 1862–1920, 2046–2500 |

Table 4. The characteristic absorption band and structure of 13 phenylpropanoids compounds.
| Chemical compound | Chemical structure | NIR absorption groups | Characteristic absorption bands (nm) | Attribution of characteristic bands |
|-------------------|--------------------|-----------------------|-------------------------------------|-------------------------------------|
| Sophoridine and Matrine | ![Chemical structure](image1) | C-H and N-H | 1386–1451 | The absorption bands of C-H in CH₂ |
| | | | 1688–1800 | The combination absorption bands by symmetric and antisymmetric oscillations of C-H in CH₂ |
| | | | 1819–1913 | |
| | | | 2165–2500 | The characteristic absorption bands of organic nitrogen compound |
| Reserpine | ![Chemical structure](image2) | C-H | 1385–1416 | The absorption bands of C-H in CH₂ |
| | | | 1863–1919 | |
| | | | 2113–2500 | The combination absorption bands of organic nitrogen compound and C-H in benzene ring |
| Sinomenine | ![Chemical structure](image3) | C-H, O-H, and N-H | 1386–1480 | The combination absorption bands of C-H in CH₂ |
| | | | 1632–1803 | The combination absorption bands of C-H in CH₂ and CH₃ |
| | | | 1867–1915 | |
| | | | 1936–2150 | The absorption bands of N-H |
| | | | 2219–2500 | The combination absorption bands of organic nitrogen compound and C-H in benzene ring |
| Colchicine | ![Chemical structure](image4) | C-H | 1339–1506 | The combination absorption bands of C-H in CH₂ and N-H in CONH₂ |
| | | | 1611–1800 | The combination absorption bands of C-H in CH₂ and CH₃ |
| | | | 1874–1915 | |
| | | | 2213–2500 | The combination absorption bands of organic nitrogen compound and C-H in benzene ring |
| Oxysophoridine and Oxymatrin | ![Chemical structure](image5) | C-H and N-H | 1380–1421 | The absorption bands of C-H in CH₂ |
| | | | 1656–1759 | The combination absorption bands of C-H in CH₂ and C=C |
| | | | 1862–1930 | |
| | | | 2149–2500 | The characteristic absorption bands of organic nitrogen compound |

**Table 5.** The characteristic absorption bands and structure assignment of 7 alkaloid compounds.
| Chemical compound | Chemical structure | NIR absorption groups | Characteristic absorption bands (nm) | Attribution of characteristic bands |
|-------------------|--------------------|-----------------------|--------------------------------------|-------------------------------------|
| Menthol           | ![Menthol](image1.png) | O-H, C-H of CH₃, CH₂ and RCH(CH₃)₂ | 1395–1435 1683–1781 1870–1909 1984–2081 | The combination absorption bands of C-H in RCH(CH₃)₂ and O-H in alcoholic hydroxyl The combination absorption bands by symmetric and antisymmetric oscillations of C-H in CH₂ The combination absorption bands of O-H in alcoholic hydroxyl The combination absorption bands of C-H in CH₁ and CH and O-H in alcoholic hydroxyl |
| Borneol           | ![Borneol](image2.png) | O-H, C-H of CH₃ and CH₂ | 1386–1432 1671–1782 1869–1919 1978–2009 2030–2089 2223–2500 | The combination absorption bands of C-H in RCH(CH₃)₂ and O-H in alcoholic hydroxyl The combination absorption bands of C-H in CH₁ and CH₂ The combination absorption bands by stretching vibration and bending vibration of O-H The absorption bands by deformation vibration of O-H The combination absorption bands of C-H in CH₁ and O-H in alcoholic hydroxyl |
| Curcumol          | ![Curcumol](image3.png) | O-H, C-H of CH₃, C=O, CH₂ and CH and benzene ring | 1386–1440 1626–1649 1674–1778 1867–1925 1993–2059 2102–2132 2206–2500 | The absorption bands of C-H in RCH(CH₃)₂ The absorption bands of C-H in C=O The combination absorption bands of C-H in CH₁ and CH₂ The combination absorption bands by stretching vibration and bending vibration of O-H The combination absorption bands of C-H in CH₂ and CH₃ The combination absorption bands by stretching vibration and bending vibration of O-H The combination absorption bands of C-H in CH₁ and O-H in alcoholic hydroxyl |
| Gossypol          | ![Gossypol](image4.png) | C=O, O-H, C-H of CH₃, CH₂ and benzene ring | 1381–1482 1866–1921 2059–2500 | The combination absorption bands of C-H in CH₂ The combination absorption bands of O-H in benzene ring |
| Oleanolic acid    | ![Oleanolic acid](image5.png) | C=O, O-H, C-H of CH₃, CH₂ and C=C | 1395–1426 1677–1780 1860–1919 2222–2500 | The combination absorption bands of C-H in CH₁ The combination absorption bands of C-H in CH₂ The combination absorption bands by stretching vibration of uncombined O-H The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₁ and CH₂ |
| Limonin           | ![Limonin](image6.png) | C=O, O-H, C-H of CH₃, CH₂ and C=C | 1387–1414 1869–1919 2217–2500 1395–1426 1677–1780 1860–1919 2222–2500 | The combination absorption bands of C-H in CH₁ and CH₂ The combination absorption bands of C=O and C-H in OCH₃ The overlapping absorption of C-H in CH₂ The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₁ The combination absorption bands of O-H in alcoholic hydroxyl and C=O The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₁ and CH₂ |

Continued
raw spectra of each substance disturbed by concentration. According to the peak position and number of auto-correlation peaks in the synchronization spectra, the NIR absorption characteristics of these substances were attributed.

### Software
Data was collected by VISION spectra collection and processed by analysis software (Metrohm China, China). Data pre-processing was conducted by Unscrambler 9.7 software (CAMO software co, Norway). Data of 2D-COS was pre-processed by self-programming of MATLAB (The MathWorks Co., American).

### References
1. Wu, Z. S. et al. NIR spectroscopy as a process analytical technology (PAT) tool for monitoring and understanding of a hydrolysis process. *Bioresource Technol* **137**, 394–399, https://doi.org/10.1016/j.biortech.2013.03.008 (2013).
2. Brouckaert, D. et al. Potential of Near-Infrared Chemical Imaging as Process Analytical Technology Tool for Continuous Freeze-Drying. *Anal Chem* **90**, 4354–4362, https://doi.org/10.1021/acs.analchem.7b03647 (2018).
3. Porep, J. U., Kammerer, D. R. & Carle, R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci Tech* **46**, 211–230, https://doi.org/10.1016/j.tifs.2015.10.002 (2015).
4. Vargas, J. M. et al. Process analytical technology in continuous manufacturing of a commercial pharmaceutical product. *International Journal of Pharmaceutics* **538**, 167–178, https://doi.org/10.1016/j.ijpharm.2018.01.003 (2018).
5. Workman, J., Lavine, B., Chrisman, R. & Koch, M. Process Analytical Chemistry. *Anal Chem* **83**, 4557–4578, https://doi.org/10.1021/ac200974w (2011).
6. Li, W. L., Han, H. F., Zhang, L., Zhang, Y. & Qu, H. B. A feasibility study on the non-invasive analysis of bottled Compound E Jiao oral liquid using near infrared spectroscopy. *Sensors and Actuators B-Chemical* **211**, 131–137, https://doi.org/10.1016/j.snb.2015.01.073 (2015).
7. Li, W. L. et al. A feasibility research on the monitoring of traditional Chinese medicine production process using NIR-based multivariate process trajectories. *Sensors and Actuators B-Chemical* **231**, 313–323, https://doi.org/10.1016/j.snb.2016.03.023 (2016).
8. Zhou, Z. et al. Comparison of Ensemble Strategies in Online NIR for Monitoring the Extraction Process of Pericarpium Citri Reticulatae Based on Different Variable Selections. *Planta medica* **82**, 154–162, https://doi.org/10.1055/s-0035-1558085 (2016).
9. Figueras, P. R., Alves, J. C. L. & Poppi, R. J. Quantification of animal fat biodiesel in soybean biodiesel and B20 diesel blends using near infrared spectroscopy and synergy interval support vector regression. *Talanta* **119**, 582–589, https://doi.org/10.1016/j.talanta.2013.11.056 (2014).
10. Pan, X. N. et al. Near infrared spectroscopy model development and variable importance in projection assignment of particle size and lobetyolin content of Codonopsis radix. *J Near Infrared Spec* **23**, 327–335, https://doi.org/10.1255/jnirs.1175 (2015).
11. Zou, X. B., Zhao, J. W., Povey, M. J. W., Holmes, M. & Mao, H. P. Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* **667**, 14–32, https://doi.org/10.1016/j.aca.2010.03.048 (2010).

### Table 6.
The characteristic absorption band and structure assignment of 9 terpenoid compounds.

| Chemical compound | Chemical structure                              | NIR absorption groups                      | Characteristic absorption bands (nm)                      | Attribution of characteristic bands                                                                 |
|-------------------|------------------------------------------------|---------------------------------------------|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Artemisinin       | ![Artemisinin Structure](image)                 | O-H, C-H of CH₃ and CH₂                     | 1384–1422                                                | The combination absorption bands of C-H in CH₃ and CH₂                                              |
|                   |                                                 |                                              | 1665–1779                                                | The absorption band by stretching vibration of C-H in CH₂                                          |
|                   |                                                 |                                              | 1866–1936                                                 |                                                                                                      |
|                   |                                                 |                                              | 2222–2500                                                | The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₃ and CH₂                |
| Nor cantharidin   | ![Nor cantharidin Structure](image)             | C=O, C-H of CH₃, CH₂, and CH                | 1386–1418                                                | The absorption bands by stretching vibration of C-H in CH₂                                         |
|                   |                                                 |                                              | 1668–1734                                                | The absorption bands by stretching vibration of C-H in CH₂                                         |
|                   |                                                 |                                              | 1865–1919                                                 | The absorption bands by stretching vibration of C=O                                                |
|                   |                                                 |                                              | 2218–2500                                                | The combination absorption bands of C-H in CH₃, CH₂, and CH                                        |
| Dehydroandrographolide | ![Dehydroandrographolide Structure](image)       | C=O, O-H, C-H of CH₃, CH₂ and C=C       | 1382–1431                                                | The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₂                        |
|                   |                                                 |                                              | 1673–1761                                                | The absorption bands by stretching vibration of C-H in CH₂                                         |
|                   |                                                 |                                              | 1859–1912                                                | The absorption bands by stretching vibration of C=O                                                |
|                   |                                                 |                                              | 1986–2171                                                | The absorption bands by stretching vibration of uncombined O-H                                     |
|                   |                                                 |                                              | 2210–2500                                                | The combination absorption bands of O-H in alcoholic hydroxyl and C-H in CH₂, CH₃, and CH₂         |
12. Balabin, R. M. & Smirnov, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. Anal Chem Acta 692, 63–72, https://doi.org/10.1016/j.aca.2011.03.006 (2011).
13. Zhao, N. et al. Optimization of Parameter Selection for Partial Least Squares Model Development. Sci Rep 5, 11647, https://doi.org/10.1038/srep11647 (2015).
14. Bi, Y. M. et al. A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. Anal Chem Acta 909, 30–40, https://doi.org/10.1016/j.aca.2016.02.013 (2016).
15. Cheng, J. H., Sun, D. W. & Pu, H. B. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen-thawed fish muscle. Food Chem 197, 855–863, https://doi.org/10.1016/j.foodchem.2015.11.019 (2016).
16. Ozaki, Y. Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry. Anal Sci. 28, 545–563, https://doi.org/10.2116/analsci.28.545 (2012).
17. Schwanninger, M., Rodrigues, J. C. & Fackler, K. A review of band assignments in near infrared spectra of wood and wood components. J Near Infrared Spec 19, 287–308, https://doi.org/10.1255/nirs.955 (2011).
18. Hu, Y. et al. C-H center dot center dot center dot O=C hydrogen bonding and isothermal crystallization kinetics of poly(3-hydroxybutyrate) investigated by near-infrared spectroscopy. Macromolecules 39, 3841–3847, https://doi.org/10.1021/ma060208q (2006).
19. Czarniecki, M. A. Two-Dimensional Correlation Analysis of Hydrogen-Bonded Systems: Basic Molecules. Appl Spectrosc Rev 46, 67–103, https://doi.org/10.1080/05704920600845868 (2011).
20. Chen, Y. J., Ozaki, Y. & Czarniecki, M. A. Molecular structure and hydrogen bonding in pure liquid ethylene glycol and ethylene glycol-water mixtures studied using NIR spectroscopy. Phys Chem Chem Phys 15, 18694–18701, https://doi.org/10.1039/c3cp52146e (2013).
21. Bec, K. R., Grabksa, J. & Czarniecki, M. A. Spectra-structure correlations in NIR region: Spectroscopic and anahmonic DFT study of n-hexanol, cyclohexanol and phenol. Spectrochim Acta A 197, 176–184, https://doi.org/10.1016/j.saa.2018.01.041 (2018).
22. Czarniecki, M. A., Morisawa, Y., Futami, Y. & Ozaki, Y. Advances in Molecular Structure and Interaction Studies Using Near-Infrared Spectroscopy. Chem Rev 115, 9707–9744, https://doi.org/10.1021/acs.chemrev.5b00315 (2015).
23. Pei, Y. L. et al. NIR assignment of isopsonalen by 2D-COS technology and model application in Yunkang Oral Liquid. Journal of Innovative Optical Health Sciences 8, https://doi.org/10.1142/s179554815500226 (2015).
24. Wu, Z. et al. Multivariate detection limits of on-line NIR model for extraction process of chlorogenic acid from Loniceria japonica. J Pharmaceut Biomed 77, 16–20, https://doi.org/10.1016/j.jpba.2012.12.026 (2013).
25. Siedliska, A., Zubik, M., Baranowski, P. & Mazurek, W. Algoritki for detecting cytotoxicity pins on the basis of transmittance mode hyperspectral data. Int Agrophys 31, 539–549, https://doi.org/10.1515/intag-2016-0075 (2017).
26. Haines, C. & Davis, S. Forensic soil analysis by subtractive FTIR spectroscopy. Forensic Sci Int 265, 120–129, https://doi.org/10.1016/j.fsi.2016.03.004 (2016).
27. Mathian, M. et al. Multivariate detection limits of on-line NIR model for extraction process of chlorogenic acid from Loniceria japonica. J Pharmaceut Biomed 77, 16–20, https://doi.org/10.1016/j.jpba.2012.12.026 (2013).
28. Czarniecki, M. A. Two-Dimensional Correlation Analysis of Hydrogen-Bonded Systems: Basic Molecules. Appl Spectrosc Rev 46, 67–103, https://doi.org/10.1080/05704920600845868 (2011).
29. Chen, Y. J., Ozaki, Y. & Czarniecki, M. A. Molecular structure and hydrogen bonding in pure liquid ethylene glycol and ethylene glycol-water mixtures studied using NIR spectroscopy. Phys Chem Chem Phys 15, 18694–18701, https://doi.org/10.1039/c3cp52146e (2013).
30. Liu, H., Xu, J. P., Qu, L. B. & Xiang, B. R. Generalized two-dimensional correlation near-infrared spectroscopy and principal component analysis of the structures of methanol and ethanol. Sci China Chem 53, 1155–1160, https://doi.org/10.1007/s11426-010-0172-2 (2010).
31. Verissimo, L. M. & Delley, B. Half-metallic ferromagnetism in Cu-doped ZnO: Density functional calculations. Phys Rev B 73, https://doi.org/10.1103/PhysRevB.73.033203 (2006).
32. Lopez-Chavez, E. et al. Decorated C-120 nanotorus: A new molecular structure for hydrogen storage. Int J Hydrogen Energ 42, 30237–30241, https://doi.org/10.1016/j.ijhydene.2017.08.095 (2017).
33. Mathian, M. et al. Identifying the phylllosilicate minerals of hypogene ore deposits in lateritic saprolites using the near-IR spectroscopy second derivative methodology. Geochim Explor 186, 298–314, https://doi.org/10.1016/j.gexplto.2017.11.019 (2018).
34. Czarniecki, M. A. Two-Dimensional Correlation Analysis of Hydrogen-Bonded Systems: Basic Molecules. Appl Spectrosc Rev 46, 67–103, https://doi.org/10.1080/05704920600845868 (2011).
35. Gezici, O., Demir, I., Demircan, A., Unlu, N. & Karaarslan, M. Subtractive-FITR spectroscopy to characterize organic matter in lignite samples from different depths. Spectrochim Acta A 96, 63–69, https://doi.org/10.1016/j.saa.2012.05.004 (2012).
36. Cramer, C. J. & Trehlar, D. G. Density functional theory for transition metals and transition metal chemistry. Phys Chem Chem Phys 11, 10757–10816, https://doi.org/10.1039/b907148b (2009).
37. Ye, L. H., Freeman, A. J. & Delley, B. Half-metallic ferromagnetism in Cu-doped ZnO: Density functional calculations. Phys Rev B 73, https://doi.org/10.1103/PhysRevB.73.033203 (2006).
38. Lopez-Chavez, E. et al. Ti-decorated C-120 nanotorus: A new molecular structure for hydrogen storage. Int J Hydrogen Energ 42, 30237–30241, https://doi.org/10.1016/j.ijhydene.2017.08.095 (2017).
39. Liu, H., Xu, J. P., Qu, L. B. & Xiang, B. R. Generalized two-dimensional correlation near-infrared spectroscopy and principal component analysis of the structures of methanol and ethanol. Sci China Chem 53, 1155–1160, https://doi.org/10.1007/s11426-010-0172-2 (2010).
40. Sinelli, N., Casiraghi, E., Barzaghi, S., Brambilla, A. & Giovanelli, G. Near infrared (NIR) spectroscopy as a tool for monitoring the isothermal epoxy curing reaction. Spectrochim Acta A 197, 114–120, https://doi.org/10.1016/j.saa.2017.11.043 (2018).
41. Wrzeszcz, W. et al. Microheterogeneity in binary mixtures of methanol with aliphatic alcohols: ATR-IR/NIR spectroscopic chemometrics and DFT studies. Rsc Adv 6, 37195–37202, https://doi.org/10.1039/c6ra04595b (2016).
42. Ma, Q. J. & Jiang, J. G. Functional Components from Nature-Derived Drugs for the Treatment of Rheumatoid Arthritis. Curr Drug Targets 17, 1673–1686, https://doi.org/10.2174/1389450110666160527122233 (2016).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (81773914), the National Key R&D program of China (2018YFC1706900), Major new drug innovation project of the ministry of science and technology (2018ZX09201011), Young Elite Scientists Sponsorship Program by CAST (2018QNRC001), and Innovative team project of Beijing University of Chinese Medicine (2019-JYB-TD011).
Author Contributions
Zhisheng Wu and Yanjiang Qiao conceived and designed this research. Lijuan Ma wrote the main manuscript text with the help of Zhisheng Wu. Data was processed by Yanfang Peng, Lijuan Ma and Yanling Pei. All other authors were involved in data analysis and manuscript preparation.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-45945-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019