

Работа с библиотека от ИЧ спектри

1. Спектрални сбирки.

Съществуват много на брой и вид *спектрални сбирки* (или *спектрални колекции*, *spectral data collections*) – сбирки от ИЧ и Раман спектри, масспектри, УВ/Вид (електронни) спектри, ^1H и ^{13}C -спектри. Обикновено колекцията от спектри се пази в един или повече текстови файлове, в които е представена спектралната информация, както и някои други данни за веществата, чийто спектри са включени в колекцията: химичното име е задължително включено в колекцията, за да може съединенията да се идентифицират, но напоследък съществено изискване е представянето и съхранението на 2D химичната структура на съединенията (таблиците на свързаност).

Едни от първите спектрални колекции са тези от ИЧ спектри, тъй като тяхното заснемане е било сравнително евтино и бързо, а информацията, която се съдържа в тях отразява в голяма степен строежа на съединенията. В следната таблица са представени някои от създадените спектрални сбирки в областта на ИЧ спектроскопия. Данните са взети от една обзорна работа на Warr [1] и представят само тези спектрални сбирки, за които има въведени структурите на химичните съединения.

Таблица 1. Създадени сбирки (библиотеки) от ИЧ спектри [1].

производител ^a	брой спектри	коментар
<i>Sadtler</i>	160,000	някои не са <i>FT-IR</i> спектри
<i>Sadtler</i> (в газова фаза)	9,200	<i>FT-IR</i> спектри
<i>Aldrich-Nicolet</i>	17,000	<i>FT-IR</i> спектри
<i>Sigma-Nicolet</i>	10,600	<i>FT-IR</i> спектри
<i>Aldrich</i> (в газова фаза)	5,000	<i>FT-IR</i> спектри
<i>NIST/EPA</i> (в газова фаза)	5,244	<i>FT-IR</i> спектри
<i>NIMCR</i> Japan	46,400	<i>FT-IR</i> спектри
<i>SpecInfo</i>	17,000	<i>FT-IR</i> спектри
<i>SpecInfo</i>	5,600	само пикови таблици
<i>Coblentz Society</i>	10,500	само 4400 пълни спектри

^a*NIST* - National Institute of Standards and Technologies; *EPA* - Environmental Protection Agency; *NIMCR* - National Chemical Laboratory, Research Institute for Polymers and Textiles.

Когато информацията в спектралната сбирка бъде структурирана и подредена, и допълнително свързана с програма за нейното поддържане, то имаме така наречената библиотека от спектри (*спектрална библиотека*, *spectral library*). В тази глава ще бъде описана структурата и работата с библиотеки от ИЧ спектри, които се поддържат от програмата **IRSS**.

2. Исторически обзор на методите за търсене в библиотеки от ИЧ спектри

Първият опит за автоматизиране на търсенето на ИЧ спектри датира от 1951 г., когато Kuentzel [2] описва приложението на IBM сортировач на перфокарти за тази цел. Типичната скорост на четене на перфокарти е била 500 карти на минута [3], което е изисквало 3.5 ч. за прочитането на 96,000 спектри, съдържащи се в

системата *SpecFinder* на фирмата *Sadtler*. Съответният брой перфокарти заема 48 кутии. Въпреки използването на т.н. *обърнато кодиране (inverted file)* [3] този брой се намалява едва на 14,300 карти (7 кутии), а времето за търсене по пълните спектрални характеристики - на 30 мин.

Много бързо спектроскопистите разбират ограничението на механичното сравнение, и затова следващите системи са ориентирани към използването на компютри. Първите програмни системи са написани на асемблер, базирани са на т.н. *mainframe* компютри и са достъпни за ограничен брой потребители. Спектралните файлове съдържат битово кодиране (0/1) на местоположението на ивиците в ИЧ спектър като напр. присъствие/отсъствие на ивици в 0.1 мкм интервали [4] или в интервали, с различна големина [5]; последното представяне е предложено от Rapp и съдържа само 10 числа от 0 до 9. Интензитетът на ивиците в ИЧ спектър съдържа важна информация и включването му при кодирането и сравнението повишава селективността на търсенето. Schrader и сътр. [6] разширяват кодирането на Rapp с включването на нови десет числа от 0 до 3, показващи интензитета на ивиците. В системата *MIRET* [7] всяка ивица се описва с битова маска, която освен интензитета отчита и невъзпроизводимостта на местоположението на ивицата. Zupan и сътрудници [8] предлагат кодиране с 240 бита на ИЧ спектри с големина на интервалите 33.3 cm^{-1} в областта $4000 - 2000 \text{ cm}^{-1}$ и 10 cm^{-1} в $2000 - 200 \text{ cm}^{-1}$. Разработената от тях система за библиотечно търсене използва също побитово кодирани ^{13}C -ЯМР и мас-спектри. Допълнително са кодирани интензитетите на ивиците с три бита. Тестовите изпитания от 300 търсения (по сто за всеки вид спектри) показват следните идентификационни способности: 77%, 95% и 91%, съответно за ИЧ-, ^{13}C -ЯМР и мас-спектрите.

Sadtler въвеждат търсенето по пикова таблица в своите компютърни системи за библиотечно търсене [9]. В пиковата таблица обикновено се включва информация за местоположението и интензитета [10], а при някои системи и за полуширината на ивиците [11], които имат интензитет по-висок от зададен от потребителя *npag (threshold)*.

Представянето на цялата спектрална крива в библиотеките и търсенето по нея не се възприема от спектроскопистите до средата на 80^{те} години, когато *Nicolet Instrument Corporation* въвежда сравняващ алгоритъм с изчисляване на сума от абсолютните разлики между спектрални криви, редуцирани до 225 точки [9,12]. Вместо това са проведени редица изследвания за начина на редуциране на спектралната крива.

Azaraga и съавтори [13] прилагат Фурие трансформация (*FT*) на спектралната крива, взета през 8 cm^{-1} в областта $2000 - 800 \text{ cm}^{-1}$. Сравнението между 60-мерните интерферограми се извършва с изчисляването на скаларно произведение. Техните резултати показват, че търсенето по интерферограми е по-слабо зависимо от наличието на шум в ИЧ спектри. Owens и Isenhour [14] прилагат същия метод, но в областта $3704 - 740 \text{ cm}^{-1}$. Авторите предлагат 100-кратно намаление на обема на спектралната информация, която се съхранява чрез побитово кодиране на интерферограмите, получени при *FT*. Kawata и сътрудници [15] използват само фазовата част на Фурие преобразуването за сравнение на спектри. Това подобрява резултатите при сравнението на спектри, тъй като фазовата компонента съдържа информация за относителното разположение на ивиците по абсисата и за отношението между техните височини, докато амплитудната част - за сумарния интензитет на ивиците, тяхната форма и абсолютното им местоположение. Предложеният алгоритъм за сравнение на спектрите чрез корелация между фазовите части показва по-голяма селективност от корелацията между спектралните криви. Isenhour и Wang [16] прилагат анализ по

главните компоненти (*principal component analysis*) на спектрите в библиотеката, които предварително са Фурие-трансформирани. Авторите проверяват възможността с така преобразуваните спектри да се идентифицира непознат спектър. В друго свое изследване Isenhour и сътрудници [17] прилагат анализ по главните компоненти едновременно върху ИЧ и мас-спектри. Обработката и на двата вида спектри повишава значително селективността при библиотечното търсене в сравнение с търсенето поотделно в две сбирки от ИЧ и мас-спектри. Във всички разгледани работи се докладва за значително намаляване на дисковото пространство, необходимо за съхраняване на спектрите, както и на времето за търсене при минимално намаление на точността на получаваните резултати.

Освен чрез съкращаване на броя на спектралните признаци скоростта на търсене може да се увеличи чрез структуриране и подредба в спектралното пространство. Zupan и Munk [18] предлагат кластеризиране на спектралната сбирка в двоично йерархично дърво. Те използват итеративен алгоритъм, наречен *three-distance clustering*, който осигурява почти 100% успеваемост при търсене по дървото. Освен възможността за съкращаване на времето на търсене максимално $N/\log_2 N$ пъти, където N е броят на спектрите в библиотеката, “преминаването” по дървото при търсене на спектъра дава информация за структурните елементи в съединението, чийто спектър се изследва [19,20].

Използваните мерки за близост (сравнение) в спектралното пространство са пряко зависими от начина на представяне на спектралните данни. Тъй като при първите системи спектралните данни са представени като набор от битове ($u_1, u_2 \dots u_M$), то сравнението на спектрите е функция на битовите логически операции - *OR*, *AND* или *XOR* [2-7]. Delaney и сътрудници [21] предлагат нова метрика за подобие между спектрите: $\Sigma[(u_k \text{ XOR } r_k) - \mu (u_k \text{ AND } r_k)]$. Нейното оптимизиране по μ при спектралните търсения дава стойност $\mu = 2.0$, което улеснява програмното приложение на метриката чрез побитовите операции на отместване (*shift*). Използвайки мощен компютър, свързан към *FT-IR* апарат, Lowry и Huppler от фирмата *Nicolet Instruments* предлагат използването на таблиците от пикове в библиотечните ИЧ спектри. Разработеният софтуер прилага поредица от интерактивни (взаимодействащи с потребителя) логически проверки за присъствието или отсъствието на дадени пикове от спектъра на непознатото вещество в пиковите таблици на библиотечните спектри [22]. Полученият хит-лист е резултат от логически операции (*boolean operations*) между резултатите от отделните проверки. Потребителят спира въвеждането на търсените пикове при достигане на обозрим набор от решения (спектри), които могат да се сравнят с този на непознатото вещество.

Първите приложени мерки за сравнение на целите спектрални криви очевидно са възможно най-простите - средно абсолютно отклонение и средно квадратично отклонение между спектрите. Първото системно изследване на тези мерки за спектрално подобие е извършено от Lowry и сътрудници [12]. Те допълнително изследват същите мерки, но приложени върху първите производни на спектрите. Изследването показва, че първата производна на спектралната крива осигурява по-добро търсене за *идентификация* на непознатото съединение, защото тя премахва постоянната част от базовата линия, но при търсене *по подобие* резултатите са по-лоши от тези получени с непроменени спектрални криви. Следващите предложени мерки са коефициентът на корелация между спектралните криви [23,24], тяхното скаларно произведение [25], както и нечетните моменти на крос-корелационната функция (*cross-correlation function*) [26]. Допълнително в разработките са проверени и сравнени редица производни метрики и начини за представяне на спектралните криви.

Успоредно с разработването на методите за библиотечно търсене се предлагат и подходи за оценка на тяхната ефективност (производителност). Delaney и сътрудници въвеждат критерия *FOM* (наречен от тях *figure of merit*), който отчита доколко резултатите от търсенето по даден метод са сходни с тези, получени при прилагането на вече утвърден метод за търсене в библиотеки от спектри [27]. Така формулиран, *FOM* е относителен критерий и може да се използва само за отсяване на поредица методи или спектрални представяния, които боравят с непълна спектрална информация. Isenhour и сътрудници [28] дефинират критерия *QRM* (*quantitative reliability metric*), който позволява оценка на проведеното търсене на спектър на непознатото вещество в библиотеката, като получените резултати се сравняват с една поредица от вътрешно-библиотечни (*intralibrary*) търсения. По този начин потребителят може да оцени с критерий, независим от визуалното сравнение на спектрите, доколко първият спектър в получения хит-лист отговаря на непознатото вещество. Clerc и сътрудници разработват цялостна методика за оценка на ефективността на една система за библиотечно търсене [29], или за сравнение на няколко системи [30]. Проведените от тях изследвания са обобщени в няколко работи, засягащи различни аспекти на библиотечното търсене [31-33].

3. Описание на програмната система

IRSS е потребителски ориентирана програма (*user friendly*), която работи в среда на *Windows*. Програмата поддържа библиотеки от ИЧ спектри, които са представени с пълната спектрална крива в интервала 3700-500 cm^{-1} . Спектралната крива е съставена от стойности на абсорбцията с големина байт през 4 cm^{-1} .

Функционалните възможности на програмата са следните:

- зареждане спектър на непознато вещество в един от трите работни буфера (екрана), както и разглеждане на резултатите от библиотечното търсене в четвъртия екран;
- показване на спектрите в даден спектрален интервал, в единици абсорбция или пропускливост, както и на съответната таблица на пиковете и структурата на химичното съединение;
- намиране на пиковите таблици на спектрите от библиотеката и на спектъра в даден буфер с праг (*threshold*), определен от потребителя;
- преобразуване на спектрален файл от формат *JCAMP-DX* в библиотечен спектрален файл;
- търсене по пикови таблици, спектрална крива и химично име;
- интерактивен регресионен анализ на спектъра на смес;
- записване на хит-лист във файл, както и в такъв за обработка от програмата *ToSiM* с цел намирането на набор от характеристични подструктури;
- класифициране на ИЧ спектри по набор от подструктури с методите на *ИНМ* и *ЛДА*, както и с използване на характеристичните интервали на подструктурите.

Схемата приложение на методите за идентификация и/или интерпретация на ИЧ спектри е дадена на фигура 1. За класификацията на ИЧ спектри от програмния модул **IRIS** са създадени класификатори, които дават информация за присъствието или отсъствието на 20 химични подструктури с точност на класификацията над праг, определен от потребителя. Тези, общо 40 класификатора прилагат модел на изкуствена невронна мрежа или метода на линейния дискриминантен анализ. Програмата работи също и с класификатори за 20 структурни елемента, които използват техните характеристични интервали. В следната таблица е даден класификатор за трет. бутилова група.

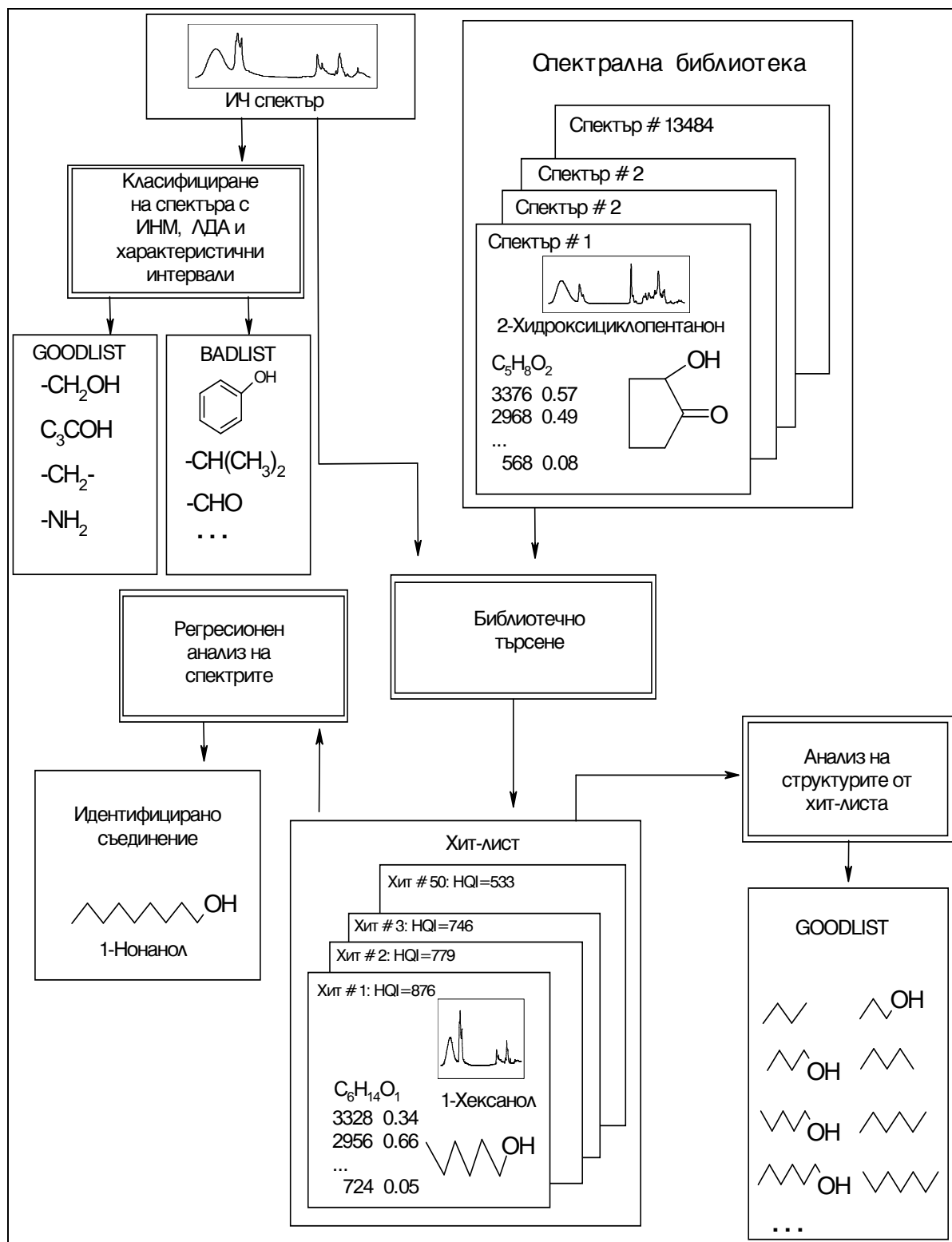
```

## CLASSNAME = -C(CH3)3
## COMMENTS = tertiary butyl, EXS
## MOLFORM = C 4 H 9
## CLASSTYPE = EXS
## CLASSINIT = PL: 2007-12-16 22:18
## NOINPNEUR = 6
## FEATURES =
CHI 2999 2800 100 4
CHI 1490 1417 100 8
CHI 1395 1360 60 9
CHI 1365 1355 60 10
CHI 1268 1235 60 10
CHI 1210 1160 60 10
## CLASSEND

```

Класификаторът дава *присъствие* на трет. бутил в съединението, чийто спектър се класифицира, *ако има пикове и в 6^{те} характеристични интервала*. Съответното предсказание се приема, че е с точност, равна на 100% - класификация от вида “да”/“не”. Проведените тестови класификации на около 200 спектъра от библиотеката *Pl-Uni* показаха редица слабости на този тип класификация, които се изразяват най-вече в т.н. свръхпредсказания. Последните се дължат на определянето на характеристичните интервали, при което те стават прекалено големи. Ето защо, класификацията с характеристичните интервали се разглежда като помощно (ориентировачно) средство и се приема за почти сигурна само при отрицателните предсказания.

При липса на спектър на изследваното съединение в спектралната библиотека резултатите от класификацията с модула **IRIS** могат да бъдат въведени автоматично в програмата *MolGen* за генериране на всички възможни химични структури, които отговарят на структурните ограничения, получени при класификацията. Наборът от характеристични подструктури, получен при анализа с концепцията на максимална обща подструктура с помощта на програмата *ToSiM* също може да бъде използван при генерацията на структурите, но процесът не може да се автоматизира напълно, поради това, че характеристичните структури не са предварително дефинирани. Проведените тестови изпитания показаха, че е възможно еднозначно определяне на структурата само на съединения с ниска молекулна маса. За съединенията, притежаващи сложна структура - повтарящи се и/или големи подструктури - интерпретирането единствено на ИЧ спектър не позволява получаването на обзрим набор от възможни структури. Според нас, това се дължи на недостатъчната информация в ИЧ спектри и може да бъде преодоляно само с допълнителната класификация/интерпретация на ЯМР и мас-спектри.



Фигура 1. Схема на приложение на методите за идентификация и/или интерпретация на ИЧ спектри.

4. Алгоритми за библиотечно търсене, програмирани в системата

В системата бяха реализирани повечето от описаните в литературата алгоритми за изчисляване на подобие между библиотечните спектри и спектъра на съединението, което се идентифицира.

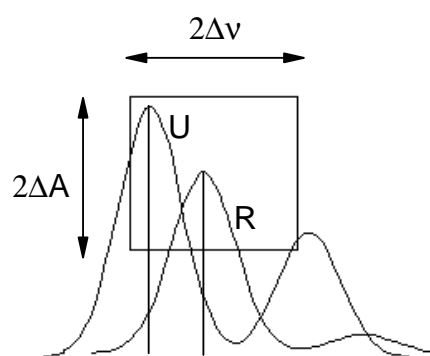
4. 1. Алгоритми за търсене по пикове

Алгоритмите за търсене по пикове, описани в литературата, се разделят на (a) *прави (forward)* - за идентификация на чисти вещества, и (b) *обратни (reverse)* - за разкриване на съставките на анализираната смес. В контекста на своето приложение първите използват мярка за подобие, която отчита колко добре спектърът на непознатото вещество съвпада с библиотечния спектър, а при вторите спектралното подобие отразява степента на присъствие на спектралните признаци на библиотечния спектър в набора от спектрални признаци на изследваната смес. Реализациите на тези алгоритми сме взимали от системата Sadtler [9], но сме извършили редица промени с цел по-голяма ефективност и точност при прилагането им.

Ако непознатият спектър се разглежда като множество U , съдържащо M пика, а всеки библиотечен - като множество R , съдържащо N пика сечението между U и R е множество I , съдържащо K пика. Идеалното съвпадение между непознатия спектър и референтния ще е изпълнено при

$$U = R = I, \text{ т.е. } N = M = K.$$

В приложната спектроскопия точното съвпадение на пиковете на два спектъра е по-скоро изключение, отколкото правило, ето защо алгоритмите трябва да боравят с една неопределеност в положението на ивиците на референтния спектър, както по абсцисата (вълновото число), така и по ординатата (абсорбцията). Съвпадането на два пика от U и R в този случай означава попадането на върха на пика от U в един правоъгълник с център върха на пика в R . Този правоъгълник е с ширина $2\Delta\nu$ и височина $2\Delta A$, които се определят от потребителя. Величините $\Delta\nu$ и ΔA са т.н. *неопределености при съвпадение* на пиковете (*tolerances* в англоезичната литература).



Фигура 1. Съвпадение на два пика с използване на неопределености (*tolerances*) по абсцисата и ординатата. С U е означен пик от непознатия спектър (*unknown spectrum*), а с R – пик от библиотечния спектър (*reference spectrum*).

Мярката за спектрално подобие, реализирана в **IRSS** представлява трицифрено число ABC , всяка цифра от което се изчислява по независим начин. Първите две числа определят доколко сечението между двете пикови таблици е еднакво с пиковата таблица на непознатия спектър (A) или с тази на библиотечния (B):

$$A = 9 K / M \quad \text{и} \quad B = 9 K / N,$$

закръглени до цели числа.

Третото число определя доколко добре пиковете съвпадат по вълново число:

$$C = 9 - \text{const} \sum |v_U^K - v_R^K|,$$

където v_U^K и v_R^K са положенията по абсцисата на k -тумте пикове в непознатия и библиотечния спектър (вълновите числа). *Const* зависи от неопределеностите $\Delta\nu$ и ΔA и се определя така, че ако всички пикове на непознатия спектър са по върховете на съответните правоъгълници с ширина $2\Delta\nu$ и височина $2\Delta A$, C е равно на нула.

В програмната система по тази схема се изчисляват две мерки на подобие, наречени *хит-лист качествен индекс* (*hit-list quality index, HQI*) [9]:

- при правия алгоритъм $HQI_F = ABC$
- при обратния $HQI_R = BAC$

Така изчислявани тези мерки притежават редица недостатъци. Един от тях е еднаквата неопределеност за всички пикове (особено по ординатата), която е критикувана от Lowry и Huppler [12]. Те предлагат търсене последователно по всеки пик в спектъра, изпозвайки различна неопределеност с последващи операции на сечение между получените извадки от спектри до получаването на разумен брой решения, които могат да се прегледат визуално от изследователя. Освен това въвеждането на изискването за съвпадение на пиковете по височина (с някаква неопределеност) не отразява намаляването на интензитета на пиковете на отделните съставки в смесите им. Също така числото B силно зависи от броя ивици в библиотечния спектъра. Това води до появата на спектри с по-малък брой пикове N в началото на хит-листа.

За избягване на тези недостатъци беше дефиниран и приложен HQI , изчисляван по уравнение (1), наречен *скаларно произведение на пикови таблици*:

$$HQI_P = \frac{\sum_k A_k^U A_k^R}{\|A^U\| \cdot \|A^R\|} \quad (1)$$

където скаларното произведение в числителя се извършва само за съпадащи (с някаква неопределеност) пикове, а в знаменателя на HQI_P стоят съответните норми на пиковите таблици.

HQI_P е косинус* от ъгъла между двата вектора на пиковите таблици U и R . Ето защо стойността му не би трябвало значително да се променя (различава от 1.0) при намаляване на интензитетите на ивиците в единия от спектрите - U или R . При анализ на смеси предвид участието на пиковете на няколко вещества се очаква определено намаление на HQI_P , но в повечето случаи индексът има локални максимуми при съставките на сместа.

4. 2. Алгоритми за търсене по спектрална крива

В програмата са заложили четири различни алгоритми за търсене по спектрална крива в зависимост от използваните мерки за спектрално подобие между спектрите: (1) *средно квадратично отклонение*, (2) *средно абсолютно отклонение*, (3) *скаларно произведение*, и (4) *коэффициент на корелация*. Ако с A_k^U означим абсорбцията при k -тото вълново число в непознатия спектъра, а с A_k^R - тази в референтния, то съответните хит-лист качествени индекси са пропорционални на:

- средното квадратичното отклонение между спектрите:

$$S_1 = \sqrt{\sum_k (A_k^U - A_k^R)^2 / N} \quad (2)$$

- средното абсолютно отклонение между спектрите:

* в спектралното пространство, определено от спектралната област и неопределеността по абсцисата $\Delta\nu$.

$$S_2 = (1/N) \sum_k |A_k^U - A_k^R| \quad (3)$$

- скалярното произведение между спектрите:

$$S_3 = \frac{\sum_k A_k^U A_k^R}{|A^U| \cdot |A^R|} \quad (4)$$

- коефициента на корелация между спектрите:

$$S_4 = \frac{\sum_k (A_k^U - \overline{A^U})(A_k^R - \overline{A^R})}{\sqrt{\sum_k (A_k^U - \overline{A^U})^2 * \sum_k (A_k^R - \overline{A^R})^2}}, \quad (5)$$

където $|A_U|$ и $|A_R|$ са големините на векторите (спектрите), а N е броят на разглежданите абсорбционни стойности.

Мерките S_1 и S_2 показват различието между спектрите. Тяхната максимална теоретична стойност е 1.0 при условие, че спектрите са нормирани в интервала 0.0 - 1.0 а.е. Те заемат минималната си стойност 0.0 при напълно идентични спектри. В програмата те са преобразувани в HQI по формулата:

$$HQI_k = 999 (1 - S_k); k = 1, 2$$

S_3 и S_4 показват подобие то между спектрите, като тяхната максимална стойност е 1.0 за напълно идентични спектри. Минималната стойност на S_3 е 0.0 (за напълно ортогонални спектри) поради това, че стойностите на абсорбцията са по-големи или равни на нула. Минималната стойност на S_4 е -1.0 за отрицателно корелирани спектри. Ето защо съответните HQI се изчисляват по следния начин:

$$HQI_3 = 999 S_3$$

$$HQI_4 = 999 (1 + S_4)/2$$

Математическият анализ на хит-качествените индекси показва редица техни недостатъци и предимства. Така например първият и вторият HQI са зависими от положението и вида на базовата линия, докато това не се отнася за третия и четвъртия. HQI_3 е донякъде (но слабо) зависим от нивото на базовата линия поради това, че тя представлява вектор с посока, различна от тази на спектъра. Поради втората степен в HQI_1 той е по-нечувствителен към малки разлики между спектралните криви в сравнение с HQI_2 . Коефициентът на корелация (и съответно HQI_4) е добра оценка за (не)зависимостта на две серии от данни само ако те са разпределени нормално, което очевидно не е изпълнено за стойностите на абсорбцията в спектралните криви. Поради липса на нормировка HQI_1 и HQI_2 са пригодни за използване само в целия обхват на спектъра, докато третият и четвъртият алгоритъм могат да се използват в произволен спектрален интервал.

ЛИТЕРАТУРА

1. W.A. Warr; Computer-Assisted Structure Elucidation. Part 1: Library Search and Spectral Data Collections. *Anal. Chem.*, 1993, **65**, A1087-A1095
2. L.E. Kuentzel; New Codes for Hollerith-Type Punched Cards to Sort Infrared Absorption and Chemical Structure Data. *Anal. Chem.*, 1951, **23**, 1413-1418.
3. F.E. Lytle; Computerized Searching of Inverted Files. *Anal. Chem.*, 1970, **42**, 355-357.
4. D.S. Erley; Fast Searching System for the ASTM Infrared Data File. *Anal. Chem.*, 1968, **40**, 894-898.

5. C.S. Rann; Automatic Sorting of Infrared Spectra. *Anal. Chem.*, 1972, **44**, 1669-1672.
6. L. Wolf-Rudiger, C. Domingo, W. Niggemann, B. Schrader; Automatic Reduction and Evaluation of Infrared and Raman Spectra. *Frasenius Z. Anal. Chem.*, 1980, **303**, 337-348.
7. R.W. Sebesta, G.G. Johson; New Computerized Infrared Substance Identification System. *Anal. Chem.*, 1972, **44**, 260-265.
8. J. Zupan, M. Penca, D. Hadzi, J. Marsel; Combined Retrieval System for Infrared, Mass, and Carbon-13 Nuclear Magnetic Resonance Spectra. *Anal. Chem.*, 1977, **49**, 2141-2145.
9. H. Somberg; Infrared Databases - Their Use, Structure and Implementation on a Microcomputer System, pp. 64-91 in: J. Zupan (Ed.); *Computer-supported Spectroscopic Data Bases*. Ellis Horwood, Chichester, UK, 1986.
10. R.C. Fox; Computer Searching of Infrared Spectra Using Peak Location and Intensity Data. *Anal. Chem.*, 1976, **48**, 717-721.
11. E.C. Penski, D.A. Padowski, J.B. Bouck; Computer Storage and Search System for Infrared Spectra Including Peak Width and Intensity. *Anal. Chem.*, 1974, **46**, 955-957.
12. S.R. Lowry, D.A. Huppler, C.R. Anderson; Data Base Development and Search Algorithms for Automated Infrared Spectral Identification *J. Chem. Inf. Comput. Sci.*, 1985, **25**, 235-241.
13. L.V. Azarraga, R.R. Williams, J.A. de Haseth; Fourier Encoded Data Searching of Infrared Spectra (FEDS/IRS). *Appl. Spectrosc.*, 1981, **35**, 466-469.
14. P.M. Owens, T.L. Isenhour; Infrared Spectral Compression Procedure for Resolution Independent Search Systems. *Anal. Chem.*, 1983, **55**, 1548-1553.
15. S. Kawata, T. Noda, S. Minami; Spectral Searching by Fourier-Phase Correlation. *Appl. Spectrosc.*, 1987, **41**, 1176-1182.
16. C.P. Wang, T.L. Isenhour; Infrared Library Search on Principal-Component-Analyzed Fourier-Transformed Absorption Spectra. *Appl. Spectrosc.*, 1987, **41**, 185-194.
17. S.S. Williams, R.B. Lam, T.L. Isenhour; Search System for Infrared and Mass Spectra by Factor Analysis and Eigenvector Projection. *Anal. Chem.*, 1983, **55**, 1117-1121.
18. J. Zupan, M.E. Munk; Hierarchical Tree Based Storage, Retrieval, and Interpretation of Infrared Spectra. *Anal. Chem.*, 1985, **57**, 1609-1616.
19. J. Zupan, M. Novic; Hierarchical ordering of spectral databases. pp. 42-63, in: J. Zupan (Ed.); *Computer-supported Spectroscopic Data Bases*. Ellis Horwood, Chichester, UK, 1986.
20. J. Zupan, M.E. Munk; Feedback Search of Hierarchical Trees. *Anal. Chem.*, 1986, **58**, 3219-3225.
21. M.F. Delaney, J.R. Hollowell, F.V. Warren, Jr.; Optimization of a Similarity Metric for Library Searching of Highly Compressed Vapor-Phase Infrared Spectra. *J. Chem. Inf. Comput. Sci.*, 1985, **25**, 27-30.
22. S.R. Lowry and D.A. Huppler; Boolean Logic System for Infrared Spectral Retrieval. *Anal. Chem.*, 1983, **55**, 1288-1291.
23. L.A. Powell, G.M. Hieftje; Computer Identification of Infrared Spectra by Correlation-Based File Searching. *Anal. Chim. Acta*, 1978, **100**, 313-327.
24. S. Saeki, K. Tanabe; Full Automation of Infrared Qualitative Analysis of Binary Mixtures by Use of a Spectral Curve Compilation. *Appl. Spectrosc.*, 1984, **38**, 693-697.
25. J.W. Sheriman, J.A. de Haseth, D.G. Cameron; A Window Fourier-Domain Infrared Search System. *Appl. Spectrosc.*, 1989, **43**, 1311-1316.
26. J.-P. Yu, H.B. Friedrich; Odd Moments of the Cross-Correlation Function for Library Searching of Infrared Spectra. *Appl. Spectrosc.*, 1987, **41**, 869-874.
27. M.F. Delaney, F.V. Warren, Jr., J.R. Hollowell; Quantitative Evaluation of Library Searching Performance. *Anal. Chem.*, 1983, **55**, 1925-1929.
28. P.B. Harrington, T.L. Isenhour; A Quantitative Measure of the Reliability of Searches of Spectral Libraries. *Anal. Chim. Acta*, 1987, **197**, 105-119.
29. M. Ruprecht, J.T. Clerc; Performance Analysis of a Simple Infrared Library Search System. *J. Chem. Inf. Comput. Sci.*, 1985, **25**, 241-244.
30. J.T. Clerc, E. Pretsch, M. Zuercher; Performance Analysis of Infrared Library Search Systems. *Microchim. Acta [Wien]*, 1986, **II**, 217-242.
31. J.T. Clerc; Automated Spectra Interpretation and Library Search Systems, pp. 145-162 in: H.L.C. Meuzelaar, T.L. Isenhour (Eds.); *Computer-Enhanced Analytical Spectroscopy*. Plenum Press, New York, 1987.
32. J.T. Clerc; Techniques for Searching Data for Qualitative Analysis, pp. 13-24. in W.O. George, H.A. Willis (Eds.); *Computer Methods in UV, Visible and IR Spectroscopy*. Royal Society of Chemistry, London, 1990.
33. M. Zurcher, J.T. Clerc; General Theory of Similarity Measures for Library Search Systems. *Anal. Chim. Acta*, 1988, **206**, 161-172.