
MÉTODO DE CODIFICAÇÃO DE ESTRUTURAS DE FLAVONÓIDES, PARA USO EM MICROCOMPUTADORES

SHEILA MARINO SIMÃO
Dra., Prof. Adjunto, DCA - IF - UFRRJ
RENATO A. M. NASCIMENTO
Mestre, Prof. Adjunto, DS - IA - UFRRJ

ABSTRACT

A CODIFICATION METHOD OF FLAVONOIDS STRUCTURES, FOR USE IN MICROCOMPUTERS

An alphanumerical code system for flavonoids is described that makes possible to process in BASIC the evolutionary advancement parameters and to identify synonym. In the code, each digit informs to the microcomputer how to give proper values to each carbon of the micromolecule. The code allows, yet, storage of these structures in common Data Base.

INTRODUÇÃO

Nas muitas situações em que é necessário armazenar e comparar grande número de estruturas químicas, surgem dificuldades de diversos tipos, como por exemplo as decorrentes de sinonímia. Na literatura, é demasiado freqüente a menção de uma substância sob nomes diversos, às vezes

totalmente diferentes (nome vulgar versus nome IUPAC, por exemplo), outras vezes com algumas poucas variações de grafia (por mau entendimento da nomenclatura IUPAC ou qualquer outras razão). Por exemplo, os três grupos de nomes que se seguem referem-se, cada um, a uma mesma substância:

- (a) 12-O-methylcoumestrol;
 - 4' - methoxycoumestrol;
 - 4' - O-methylcoumestrol; e
 - 9 - O- methylcoumestrol.
- (b) 3,7,4' - trihydroxyflavone;
 - 4' , 7 - dihydroxyflavonol;
 - 7, 4'- di-hydroxyflavonol;
 - 5 -deoxykaempferol; e
 - deoxy -5-kaempferol.
- (c) afzelin;
 - kaempferol;
 - populnin; e
 - swartziol.

Hoje, tal dificuldade parece superada, ou minimizada, pelo uso de computadores. Entretanto, ao se processarem as informações relativas a uma classe química, da qual uma das substâncias exemplificadas faça parte (ou relativas a um taxon vegetal onde ela ocorra), utilizando-se dos nomes como parâmetros de comparação, tal substância aparecerá tantas vezes quanto forem os sinônimos, qualquer que seja o tipo de banco de dados, em que somente o nome da substância está armazenada e não sua estrutura. Sendo pequeno o número de observações com que se trabalha e dispendo-se da estrutura de cada uma delas, é possível detectar-se a ocorrência de sinonímia e eliminá-la por mera observação visual. Quando porém, é muito grande o número de informações a analisar, as possibilidades de erro nesse confronto crescem assintoticamente, fora o tempo que se perde e a característica tediosa do trabalho. Os programas sofisticados, que permitem a representação gráfica nos periféricos (vídeo ou impressora), são pouco mais que meros auxiliares do desenho e, também neles, o tempo unitário de representação de cada estrutura, embora aparentemente pequeno, atinge a um somatório exaustivo quando se trata de uma grande massa de informações. Além disso, esses programas gráficos ocupam muito espaço de memória, o que constitui uma limitante ao emprego de microcomputadores mais simples.

Adicionalmente, determinados tipos de exame de estruturas, como os necessários aos estudos quimiossistemáticos, além de não prescindir dessa eliminação da sinonímia, exigem cálculos e contagens que, se feitos manualmente e aplicados a um número muito grande de informações, conduzem facilmente a percentuais excessivos de falhas humanas.

Tendo de analisar, comparar, contar e classificar cerca de 11.000 registros bibliográficos de ocorrências de flavonóides. com fins quimiossistemáticos, deparamo-nos, em alto grau, com esse problema de sinonímia e perda de tempo decorrente da sua eliminação. Em razão disto, desenvolvemos um código alfanumérico seqüencial, linear - como se uma palavra fosse - para cada estrutura, de fácil atribuição pela simples observação visual da estrutura química desenhada, que ocupa pouco espaço no disquete e na memória, e cuja entrada no sistema computacional (digitação) pode ser confiada a auxiliares sem especial preparo químico. Dadas as suas características, a esse código puderam ser aplicados programas simples em BASIC, linguagem de computação disponível em todos os microcomputadores de uso pessoal, de modo a transferir para a máquina a tarefa tediosa e demorada de detectar e eliminar a sinonímia e outras, como a de fazer cálculos numéricos, no caso índices e parâmetros evolutivos.

Pode-se fazer uma idéia da fonte de erros decorrente desse procedimento, examinando-se a forma de cálculo de um desses índices, o de oxidação. Para cada substância, atribuem-se valores numéricos conforme os substituintes ligados à cadeia carbônica (valor - 1 a cada ligação C-H, valor +1 a cada ligação C-X - onde X= heteroátomo - e assim por diante). Em seguida, faz-se o somatório dessa pontuação, o qual é então dividido pelo número de carbonos provenientes do precursor biossintético. Tomam-se evidentes os problemas desse tipo de contagem, se manual, quando aplicado para 11.000 ocorrências, de todas as ligações G-H e C-X de cada molécula de uma classe química em que o número de carbonos de cada estrutura varia entre 14 e 15 (ou seja, cerca de 150 mil pontos de contagem).

Também outros procedimentos semelhantes ressentem-se muito do grande aumento do número de observações. Estão neste caso os parâmetros evolutivos de O-metilação e O-glicosilação, que tem como base a divisão, por taxon vegetal, do número de substâncias que possuem estes substituintes pelo número total de substâncias. Identificar, caso a caso e de forma visual, quais são as estruturas assim substituídas é um trabalho monótono e entediante que, invariavelmente, é feito mais de uma vez, por margem de segurança. pois já se reconhece a grande possibilidade de erros.

Dado que esse código, criado exclusivamente para uso próprio, resolveu esses problemas de forma rápida e - principalmente - segura, julgamos oportuno dele dar conhecimento à comunidade científica, não tanto o método em si (que só é válido para essa classe de substâncias), mas a idéia central que o preside, a de valer-se de um critério posicional, geométrico, para a representação alfanumérica da estrutura e - sobretudo - para que esse tipo de representação possibilite a transferência ao computador de tais tarefas. Cabe ainda acrescentar que ele se torna mais útil mediante a aplicação de programas em BASIC, que permitam a entrada dos dados e a sua leitura, e demais operações. Programas desse tipo são fáceis de elaborar, mesmo para quem conheça pouco mais que o nível I dessa linguagem. Ademais, a feição alfanumérica permite a estocagem como arquivo-texto seqüencial, formato este que pode a qualquer momento ser lido pela maioria dos aplicativos (planilhas e banco de dados), dos quais se pode então aproveitar as facilidades usuais (ordenação, seleção, busca, etc.).

METODOLOGIA

Os flavonóides são de estrutura relativamente rígida, isto é, não sofrem rearranjos complicados que venham a descaracterizar a estrutura do precursor. A relativa uniformidade do padrão estrutural dessa classe permite entender-lhes as estruturas como sendo geométricamente (gráfica não quimicamente) variações de um esquema básico.

Na idealização do código, aproveitou-se a faculdade que tem o computador de identificar posições em uma dada seqüência de dígitos alfabéticos (strings ou cadeias). Para isso, adotou-se, como base, uma estrutura hipotética fundamental (1) e duas variantes (2 e 3), que dela divergem apenas pela posição em que o anel B se liga ao C. Para a atribuição do código, as estruturas reais dos flavonóides são entendida como diferindo (do ponto de vista gráfico, não químico) apenas quanto à forma com que estão ocupadas as posições desse padrão fundamental ou de uma de suas variantes. Nessa estrutura fundamental, que se compõe dos três anéis flavonoidicos (A, B e C) e deve ser sempre desenhada e numerada numa mesma posição (pré-estabelecida), cada posição recebe, conforme o anel, um ou dois dígitos que expressam os substituintes e/ou insaturações daquela posição.

A codificação começa pela numeração das posições dessa estrutura fundamental, feita no sentido horário e entendida como posição número 1, em cada um dos anéis, a que ocorre na posição correspondente as 12 horas (1, 2 e 3).

No caso de auronas (4) e de benzofuranos (5 e 6), a seqüência de numeração para o anel C é invertida e numera-se por último o carbono externo ao anel.

O código compõem-se de duas partes. A primeira é formada pelos primeiros vinte e três dígitos, nos quais são representados os substituintes e a forma de ligação do anel B ao anel C. A segunda parte é formada pelo vigésimo quarto dígito e subsequentes, em número variável; os dígitos desta segunda parte fornecem explicações sobre os substituintes, quando necessárias. Na primeira parte do código, cada uma das posições dos anéis A e B é representada, no código, por um dígito. Já no anel C, as posições 3 e 4 o são por dois dígitos, cada uma delas, enquanto que as posições 1 e 2 o são por um dígito apenas. Em consequência, essa primeira parte do código tem o seguinte formato:

```
123456 123456 123344
000000 . 000000 . 000000/00
anel A  anel B  anel C
```

Os dois dígitos situados após a barra (21° e 22° dígitos) representam a ligação do anel B ao anel C, sendo que o primeiro deles (21°) indica a posição do anel C à qual o B está ligado e o segundo (22°), a estereoquímica desta ligação. Este segundo dígito será igual a "A" quando a ligação for alfa, "B" quando for beta, "O" quando o carbono não for assimétrico e "?" quando a estereoquímica não for conhecida, o código que se segue, por exemplo, refere-se a uma estrutura em que o anel B prende-se a posição 2 do anel C e a ligação é beta:



Para cada posição, não se gasta dígito para representar o átomo de carbono pertencente à estrutura fundamental. Nela, o dígito indica apenas: 1) o substituinte que a ele esteja ligado; 2) o átomo de oxigênio do anel C, comum a maioria dos flavonóides; ou 3) a inexistência de qualquer átomo, o que ocorre em uma ou duas posições do anel C, para determinados flavonóides (chalconas (8), estilbenos (9) e diarilpropenos (10)). Assim, esta primeira parte do código tem sempre o mesmo número de dígitos, e o computador pode ser programado para ler o que há em cada posição do código. Por exemplo, a presença de um 0 (zero) como

15° dígito indica que na posição 1 do anel C inexistente o oxigênio, tratando-se então de um flavonóide de uma das três classes mencionadas acima. Essa rigidez posicional é que permite elaborar programas em BASIC simples, mediante os quais é o computador quem: a) indica sinonímia e/ou elimina sinônimos; b) calcula índices para cada substância; c) procura e separa estruturas pelo padrão de oxigenação de cada anel; d) estabelece a classe de flavonóide à qual a estrutura pertence (flavonol, isoflavona, rotenóide, etc). Por exemplo, num conjunto de flavonóides, os que possuem substituintes oxigenados na posição 5 do anel A, terão como 4° dígito um de determinados símbolos (um O ou um M, por exemplo) e não outros (um H, por exemplo).

Os dígitos são letras (Tabela 1). Os pontos (7° e 14° dígitos) indicam a separação entre anéis; são dispensáveis, mas visam facilitar digitação e conferência.

Quando da atribuição de códigos às posições 3 e 4 do anel C, o zero (O) tem precedência sobre qualquer substituinte e o hidrogênio (H) tem precedência sobre qualquer substituinte oxigenado (11):

TABELA 1. Códigos empregados na representação dos substituintes do esqueleto flavonoídico.

Substituinte	Código	Substituinte	Código
COOH	Q	Oxigênio	O
Hidrogênio	H	OH	X
2 Hidrogênio	U	-O-substituinte diferente de H e Me	J
3 Hidrogênio	I	C=O	D
OMe	M	-CHO	Z
GMe	E	-O-C=O	S
O-Glicosídeo	G	-O-CH-OMe	K
C-Glicosídeo	L	-O-CH	W
O-Prenila	P	-O-CH ₂	T
C-Prenila	R	-CH ₂	F
C-Prenila Ligada pela cauda	V	-O-CH-OH	Y

HOHHM.HXHHOH.OOOHOD/20 (certo)
 HOHHM.HXHHOH.OOHODO/20 (errado)

Esta precedência é necessária porque para o computador estes dois códigos representam estruturas diferentes.

A segunda parte do código é usada todas as vezes em que a primeira for insuficiente para representar corretamente a estrutura. Acrescenta-se no final do código um ponto (.), isto é, o vigésimo quarto dígito, ao qual se segue a explicação necessária:

HOHXH.HXHHOH.OOOJOD/20.p-coumaroil (12)

Deve ser esclarecido que estas explicações complementares não são levadas em conta no cálculo dos parâmetros de Gottlieb. de modo que, em não havendo necessidade de estudar-se esses detalhes, a programação automaticamente poderá eliminá-los. até o ponto (24º dígito) para fins do índice de oxidação.

Ao representar uma chalcona ou estilbeno. deve-se informar como última explicação qual é o isômero espacial em questão (cis ou trans). Se este dado não for conhecido. não se escreve nada:

HMOHHX.HTOHH.OHOHOD/20.trans (13)

Quando o anel B estiver ligado ao C por mais de uma ligação (por ex. rotenóides). indica-se depois da barra primeiro a ligação carbono-carbono dos anéis e posteriormente a que é feita por intermédio de oxigênio:

HOHXH.HXHHOH.OUOHOD/3?4? (14)

Os flavonóides poliméricos são codificados da mesma forma que os monoméricos, sendo que entre os códigos das unidades acrescentam-se os números que representam os pontos de ligação entre elas. Estes números acompanham a nomenclatura da IUPAC para flavonóides. Apesar de, aparentemente; tomar confusa a codificação. ao misturar dois tipos diferentes de numeração do esqueleto (IUPAC e a aqui criada), tal foi necessário. pois a indicação dos pontos de ligação entre os dois esqueletos. pela nossa numeração, teria de ser acompanhada pela indicação dos anéis aos quais pertencem os dois carbonos, aumentando ainda mais o tamanho do código.

Ano 1/1994

Os sinais de adição (+) são empregados para separar as informações sobre o tipo de ligação. entre as unidades flavonoidicas. dos códigos representativos destas unidades.

OOOXHE.HMHGH.OOOHOD/20+8->8+000XHE.HMHGH.H.OOOHOD/20 (15)

HOHXD.OHMXMO.OOOOH/30+2->3.2'->2+HMOHHM.HXMHH.OOOHH/20 (16)

OOHHO.HMHGH.OHOOD/30+8->3.7->4+XHOHXH.XXHOHH.OHOHO/3? (17)

Observe-se que na estrutura 15 a numeração da segunda unidade monomérica continua a seguir o padrão aqui estabelecido, apesar de graficamente estar desenhada de cabeça para baixo.

Seguemse exemplos da codificação de algumas estruturas:

HOHXH.HHHHO.OOOHD/30 (18)

HOHXH.HHHOH.OHOHHU/2B.3B (19)

HOHXH.XXHOH.OOXOH/20 (20)

HOHHM.OHOHH.OUOHOD/4A (21)

A ocorrência de glicosídeo é representada escrevendo-se após o ponto de explicação o nome do açúcar (Tabela 2) diretamente ligado ao esqueleto precedido do código "G-" (quando for O-glicosilação) ou do código "L-" (quando for C-glicosilação). Quando houver necessidade de duas ou mais explicações diferentes. isto é, quando estas estiverem ligadas a pontos diferentes da estrutura, elas serão separadas por um ponto e descritas na ordem em que aparecem: primeiro as do anel A, depois as do anel B e por último as do anel C. Havendo, num mesmo anel, mais de uma posição ocupada por açúcares, as explicações correspondentes serão feitas da mesma maneira como anteriormente explicado, mantendo-se a seqüência da numeração do anel:

LOOXHG.HGHGH.OOOGOD/20.L-Glu..G-Glu.G-Rha.G-Xy1 (22)

Se o objetivo da codificação for a representação fidedigna das estruturas com vistas, por exemplo, a eliminação de casos de sinonímias, pode-se acrescentar todo tipo de informação que se fizer necessária. como o modo ou o tipo de ligação dos açúcares:

HOHHM.HHGHO.OHOOD/30.G-Glu (beta-D) Glu (beta-D) (6->1) (23)

O mesmo pode ocorrer Para os substituintes prenitados. Estes podem necessitar de explicações: a) quanto à numeração; b) quanto à ocorrência e posição de ligações duplas; c) quanto à formação de novos anéis pela cadeia prenítica; e d) quanto à presença de substituintes também na cadeia prenítica.

A notação para prenitações tem como base a presença de grupo 3-metilbutil ligado ao esqueleto flavonoídico. A O-prenitação é representada por "P" e a C-prenitação por "R" (Tabela 1). O carbono 1 da prenita é aquele que se encontra ligado à estrutura flavonoídica. Esta regra foi estabelecida de modo contrário às determinações da IUPAC por ser visual e geometricamente mais cômoda, principalmente quando a digitação é feita por pessoa pouco familiarizada com este tipo de nomenclatura. A primeira explicação que se segue é a presença de ligação dupla, na prenita. Faz-se isto pela repetição do código R ou P, conforme o caso, e pela adição de um algarismo correspondente à posição da insaturação. após o ponto (.) para explicações:

ROOMHX.HMHHOH.OODOH/40.R2 (24)

Ciclizações do grupo prenita são representadas primeiro pela posição de insaturação seguida de "O" e posteriormente

pela indicação de qual carbono se liga ao oxigênio no fechamento do anel:

ROOHHO.HHHHHO.OOOOHU/20.R10R3 (25)

Quando no grupo prenita ciclizado estiverem ausentes as metilas geminadas do carbono 3, adiciona-se um zero (O) antes da notação explicativa:

ROOHHO.THMMHO.OHOHOD/3A.2A.OR10R2 (27)

Quando a prenita ciclizada não possui insaturação, a notação começa por "O" e inclui-se, após a indicação da posição de fechamento, notação que explique a estereoquímica das ligações:

ROOXHO.YHMMHO.OHOHOD/3A.2A.OR2B (26)

RHOHHO.HHHHHO.OOOHOH/30.0R27.trans (28)

Explicações outras, que não sejam referentes a glicosilações e a prenitações, devem ser as últimas a aparecer no código, conforme está demonstrado no exemplo da estrutura 28, para a ligação dupla entre os anéis flavonoídicos A e B.

A ocorrência de prenita ciclizada sem insaturação e sem dimetil geminado é representada por "O" seguido por "O" e, finalmente, a posição de fechamento do anel:

ROOMHO.HXHHOH.OHHXHH/2B.3A.OOR3 (29)

TABELA 2. Códigos empregados na representação de açúcares ligados ao esqueleto flavonoídico.

Código	Açúcar	Código	Açúcar
Rut	Rutinose	Sam	Sambubiose
Gly	Glicosídeo	Api	Apiose
Glu	Glucose	Get	Gentiobiose
Rha	Rhamnose	Sop	Sphorose
Gal	Galactose	Ron	Robinobiose
Xyl	Xilose	Glc	Ac. Glucurônico
Ara	Arabinose	Lat	Latirose
Neo	Neohesperidose	Bio	Biose
Vic	Vicianose	Fru	Frutose
Rob	Robinose	Cel	Celubiose
Gbi	Glucobiose		

Qualquer substituição que ocorra na prenila segue a notação usada na Tabela 1, precedida pela notação de insaturação e ciclização. As indicações são unidas por um hífen:

ROOHHO.THTOHO.OHOHOD/3B.2B.OR2B-G3.G-Glu	(30)
ROOMHO.HHHHOH.OOOHOD/20.F1B-J2A-J3.acetil	(31)
.acetil	(31)
ROOMHX.HHHHOH.OOOHOD/20.R1-S3->1	(32)
ROOMHX.HHHHOH.OOOHOD/20.D2-W3->1	(33)
ROOMHT.HHHHOH.OOOHOD/20.J1B-D2-W3->1(cis)	(34)
.acetil	(34)
ROOMHX.HHHHOH.OOOHOD/20.02?->3	(35)

A estereoquímica dos substituintes prenílicos é indicada imediatamente após a explicação do substituinte, sem separação, conforme os exemplos 30, 31, 34 e 35.

A notação "cis" e "trans" adotada no exemplo 34 segue a normalmente empregada em rotenóides, isto é, os substituintes ligados aos carbonos assimétricos vicinais estão para o mesmo lado do plano (cis) ou de lados opostos (trans).

As setas empregadas nos exemplos 32, 33, 34 e 35 tem a finalidade de apontar o segundo local de ligação do substituinte à cadeia prenílica.

Se, ao invés de uma prenila, o substituinte for um grupo geranila, este também é indicado por R ou P, só que as posições de insaturação são representadas como se fossem duas prenilas independentes unidas pelo modo de ligação das duas unidades entre parênteses; o número do carbono de ligação da primeira prenila seguido por uma seta e pelo número do carbono de ligação da segunda prenila:

RHOHHX.HXHHOH.OHOHOD/20 R1(5->1)R2.trans	(36)
--	------

A existência de mais de um grupo prenilado ligado ao esqueleto é descrita na ordem em que estes grupos aparecem na codificação, de forma análoga as glicosilações e as ciclizações em grupos geranila são codificadas de modo semelhante ao das prenilas:

RHOHRO.HPROOH.OHOHHH/20.R2.R3OR2.P2.R1OR3.trans	(37)
ROOHHO.HORHOH.OHHHHH/2?.R1OR3(5->1)R2.R1OR3	(38)

RESULTADOS

Muitos resultados foram obtidos pelo uso deste tipo de codificação, principalmente por facilitar a elaboração de programas em BASIC para muitas formas de depuração e organização de dados estruturais (eliminação de sinonímia, cálculos de índices, classificação de estruturas e outros). Foi também útil ao usar bancos DataBase comerciais. No caso do Apple, que permite converter arquivos-texto, tanto comuns como do tipo DIF, seja em DOS 3.3, seja em PRODOS, em arquivos DataBase, a vantagem do código alfanumérico foi particularmente grande. Como exemplo da utilidade deste método de codificação, segue-se a listagem de um programa em BASIC, para cálculo do índice de oxidação.

```

100 HOME
110 PRINT "DIGITE O CODIGO DA SUBST": PRINT
    PRINT "ATE A BARRA,EXCLUSIVE "; INPUT
    " "; S$
120 FOR J = 1 TO 6: GOSUB 2000: NEXT
130 S = S / 6: S1 = S: S = 0
140 FOR J = 8 TO 13: GOSUB 2000: NEXT
150 S = S / 6: S2 = S: S = 0
160 FOR J = 15 TO 20: GOSUB 2000: NEXT
170 S3 = S / 3: S = 0
180 ST = S1 + S2 + S3
190 PRINT: PRINT "COD. DA SUBST "; S$:
    PRINT: PRINT "NO.OX. "; ST
200 END
2000 REM SUBROT CALC OXI
2010 X$ = MID$(S$,J,1)
2020 IF X$ = "E" OR X$ = "F" OR X$ = "L" OR
    X$ = "R" OR X$ = "V" OR X$ = "Z" OR X$ =
    "O" THEN S = S: GOTO 2080
2030 IF X$ = "H" THEN S = S-1: GOTO 2080
2040 IF X$ = "D" OR X$ = "Q" THEN S = S + 2:
    GOTO 2080
2050 IF X$ = "U" THEN S = S-2: GOTO 2080
2060 IF X$ = "I" THEN S = S-3: GOTO 2080
2070 S = S + 1
2080 RETURN

```

Este programa é a simplificação de parte de um dos programas que elaboramos para calcular o índice de oxidação para flavonóides.

O programa original lia arquivos-texto, mas foi aqui simplificado para que, por digitação no teclado, o leitor possa obter o índice de qualquer das estruturas aqui exemplificadas (no caso dos biflavonóides, obviamente cada uma das duas unidades estruturais deverá ser digitada em separado). Utiliza o BASIC do APPLE (Applesoft), mas serve para a maioria dos outros sistemas, com pequenas modificações (como por exemplo, o comando HOME que equivale ao CLS de outros sistemas).

Consiste do seguinte. Na linha 110, pede-se a digitação da estrutura, que é armazenada como S\$. Nas linhas 120-160, há três laços FOR... NEXT, que remetem o computador a subrotina 2000-2080, onde são somados sucessiva e algebricamente os valores correspondentes ao substituinte de cada posição. Nas linhas 180 e 190, os valores de cada anel (S1, S2 e S3) são somados e é exibido no vídeo o valor do índice de oxidação (ST).

Programas tão simples quanto este permitem atingir a outros propósitos, como o de estabelecer a subclasse de flavonóide a que pertence uma dada estrutura. Por exemplo, o programa seguinte separa chalconas, estilbenos e diarilpropenos, entre si e das demais subclasses.

```
100 HOME
110 PRINT "DIGITE O CÓDIGO DA SUBST": PRINT :
    PRINT "ATE O PRIMEIRO DIGITO APOS A
    BARRA.INCLUSIVE "; INPUT "";S$
120 IF MID$(S$,15,1) = "O" THEN 140
130 PRINT "OUTRA CLASSE" : END
140 IF MID$(S$,16,1) = "0" THEN PRINT :
    PRINT "ESTILBENO" : END
150 IF MID$(S$,22,1) = "4" THEN PRINT :
    PRINT "DIARILPROPENO" : END
160 PRINT "CHALCONA"
```

Para identificar outras subclasses, a linha 130 teria de incluir um comando GOTO para remeter a um subrotina apropriada.

Naturalmente, se as estruturas estivessem estocadas em arquivo texto, tais tipos de operação dispensariam a digitação individual das estruturas, por leitura direta do arquivo em disquete.

DISCUSSÃO

Diversos sistemas de codificação de estruturas químicas vem sendo elaborados e divulgados na literatura científica. No desenvolvimento desses sistemas, os autores preocupam-se sempre com a universalidade. Procurando aplicabilidade a qualquer estrutura química, independentemente da classe estrutural, são sistemas desenvolvidos para armazenar milhões de substancias, com objetivo de acessar a comunidade científica todos os dados da literatura sobre uma determinada estrutura ou sobre um grupo de substancias estruturalmente afins. Para tarefa tão ampla, é necessário que a codificação seja efetivamente muito abrangente, havendo por conseguinte a necessidade de que o usuário seja capaz de reconhecer notações químicas, como por exemplo o tautomerismo, ou de utilizar métodos não-instintivos, como a numeração de anéis no sentido anti-horário ou como a formação do mesmo código, ora no sentido esquerda-direita, ora no sentido direita-esquerda. Há casos em que para se construir o código de uma estrutura é necessário a consulta a cinco tabelas diferentes

Devido ao objetivo desses sistemas, faz-se necessária a utilização de equipamentos de alta capacidade de armazenamento e de manipulação de grandes quantidades de informação, equipamentos caros e pouco disponíveis.

Há ainda o fato de que, por causa desse objetivo, as codificações não têm um número fixo de dígitos, o que implica na criação de programas (software) de manipulação de dado complicados, que não podem ser criados por pessoas sem um bom conhecimento de programação em linguagens bem mais complexas do que o BASIC.

Nosso objetivo foi diferente. Não nos propusemos a criar um código universal. Ao contrário, como explicitado anteriormente, a codificação aqui exposta teve por objetivo simplificar o manuseio de substancias de uma única classe estrutural, somente com fins quimiossistemáticos. Procurou-se manter rigidez na atribuição de códigos, de modo a facilitar a programação na linguagem fácil e acessível, que é o BASIC. Tal não seria possível nos sistemas de codificação que não

obedecem a este critério. Por dispor de um computador de 8 bits e, portanto, de baixa capacidade de manipulação de dados (velocidade de manipulação e capacidade de armazenamento), buscou-se um sistema que tivesse o menor número possível de dígitos e para isto não se representam os átomos que compõem o esqueleto básico dos flavonóides, mas somente os substituintes das estruturas, planejando este que não pode ser seguido nos outros sistemas de codificação, uma vez que neste o objetivo é lidar com qualquer tipo de estrutura.

Outra meta que se buscou foi a de fazer com que a operação em que se atribui o código à estrutura fosse a mais simples possível e, além disso, destituída da exigência de conhecimentos químicos, para que essa tarefa pudesse ser transferida a pessoas sem esses conhecimentos, como um digitador comum, somente acostumado a lidar com números e letras e não com notações químicas.

CONCLUSÃO

Pode-se notar que estruturas difíceis de diferenciar visualmente, como as dos exemplos 32 e 33, são, porém, bastante distintas pelo código para o computador.

A segunda parte do código poderia ter sido mais simplificada (pela adoção de símbolos mais curtos, como por exemplo "ac" em lugar de "acetil"), mas para o nosso uso isso não foi indispensável. Cada usuário pode criar, livremente, qualquer tipo de codificação que melhor lhe convier para esta parte do código, uma vez que esta não tem valor para os cálculos de parâmetros evolutivos, importando somente para tratar casos de sinonímia.

Essa forma de codificação foi aplicada aos flavonóides e só para eles serve, mas provavelmente será possível estender à codificação de outras classes de micromoléculas o seu princípio geral de elaboração.

BIBLIOGRAFIA

- Gottlieb, O.R. ; "Micromolecular evolution, systematics and ecology", Springer-Verlag; Berlin (1982).
- Emerenciano, V. de P., Ferreira, F. S., Kaplan, M. A. C., Gottlieb, O. R. ; *Phytochemistry* (1987), 26, 3103.
- Morgan, H. L.; *J. Chem. Docum.* (1967), 5, 107 .
- Gastmans, J. P. ., Furlan, M. ., LoPes, M. N. ., Borges, J.H.G., Emerenciano, V. de P.; *Quimica Nova* (1990), 13, 10 .
- Skolnik, H.; *J. Heterocyclic Chem.* (1969), 6, 689.
- Skolnik, H.; *J. Chem. Docum.* (1970), 10, 216.
- Bawden, D.; *J. Chem. Inf. Comput. Sci.* (1983), 23, 14.
- B Dittmar, P.G., Farmer, N.A., Fisanick, W., Haines, R.C., Mockus, J.; *J. Chem. Inf. Comput. Sci.* (1983), 23, 93.
- Skolnik, H.; *J. Chem. Docum.* (1971), 11, 120.





