Sedyó Inez

Gyűrűrendszerek felismerése 2D NMR mérésekből származó kémiai gráfokban

Témavezető:
Dr. Rohonczy János
Szervetlen Kémiai Tanszék

——— Eötvös Loránd Tudományegyetem ———
—— Természettudományi Kar ——
— Budapest, 2012 —

Köszönetnyilvánítás

Köszönettel tartozom témavezetőmnek, Rohonczy Jánosnak. Köszönöm segítségét és útmutatását, és köszönöm, hogy mindig kész volt segíteni, ha problémába ütköztem.

Köszönöm Novák Zoltán csoportjának, hogy a rendelkezésemre bocsátottak néhány vegyületet, és ezzel elősegítették a tesztelést.

A projekt az Európai Unió támogatásával és az Európai Szociális Alap társfinanszírozásával valósul meg, a támogatási szerzõdés száma TÁMOP 4.2.1./B-09/KMR-2010-0003.

Tartalomjegyzék

1. Bevezetés

Az NMR-spektroszkópia (Nuclear Magnetic Resonance, azaz mágneses magrezonancia spektroszkópia) jelentősége a korszerű szerkezetkutatásban elvitathatatlan. Egy szerves vegyület szerkezetének igazolása, vagy éppen egy újonnan előállított anyag struktúrájának meghatározása ma már elképzelhetetlen NMR vizsgálatok nélkül, ugyanakkor a technika egyre nagyobb teret nyer a fehérjék és más biomolekulák tanulmányozásában is.

Az NMR-készülékek gyártói közül az egyik legkiemelkedőbb vállalat a Bruker BioSpin Corporation, mely az 1960-as megalakulása óta készít spektrométereket, és folyamatos fejlesztésekkel áll elő nemcsak az NMR-spektroszkópia technológiai oldalán, hanem az adatfeldolgozás területén is. Ilyen fejlesztés a spektrométerekhez használt TopSpin mérőszoftver 3.1-es verziójával együtt megjelent *Small Molecule Structure Elucidation* (CMC-se, kismolekulás szerkezetértelmezés) is. Ez az alkalmazás többféle - javarészt kétdimenziós korrelációs - NMR-spektrum alapján képes előállítani olyan lehetséges kétdimenziós molekulaszerkezeteket, melyek összhangban vannak a spektrumokban található információkkal. Ezt nem ismert vegyületek spektrumaiból álló adatbázis alapján teszi, hanem a magok közötti korrelációk elemzésével, ami az előbbinél sokkal univerzálisabb megközelítés [1; 2].

Azonban ami ennek a szerkezetmeghatározó módszernek előnye, egyben a hátránya is: mivel arra törekszik, hogy a tényleges molekulaszerkezetet biztosan megtalálja, ezért a rendelkezésre álló információknak megfelelő valamennyi szerkezetet legenerálja, ami sok esetben nagyszámú találatot eredményez. Emellett a tapasztalat azt mutatja, hogy a találatok között sokszor kevéssé „reális” szerkezetek is szerepelnek, mivel a spektrumokból csak a skaláris csatolásra (és ennek a csatolásnak a kötéseken átívelő hosszára) vonatkozó információk kerülnek felhasználásra, a program a spektrumból egyéb következtetéseket nem képes levonni.

A fentiek fényében adódott az ötlet, hogy a *Small Molecule Structure Elucidation* által generált molekulaszerkezeteket érdemes lenne valamilyen rendszerbe foglalni, azokról valamilyen előzetes információt adni, hogy a felhasználó a kapott nagy mennyiségű szerkezetet könnyebben átláthassa és a legvalószínűbbet kiválaszthassa. Mivel ez a program szerves kismolekulákra lett tervezve, ezen molekulák között pedig nagy számban fordulnak elő az igen változatos összetételű gyűrűs szerkezetek, a rendszerezés szempontjának a találatokban megtalálható gyűrűk típusa szinte magától adódik.

2. Irodalmi áttekintés

Az NMR-spektroszkópiában mágnesesen aktív magok (például a ^{1}H, ^{13}C, ^{15}N, stb. izotópok) vizsgálatára van lehetőség, ugyanis a mérés során a magok mágneses állapotai között hozunk létre átmenetet. A mágneses aktivitás az impulzusmomentumra (L) vezethető vissza, ami kvantált mennyiség:

$$L = \hbar\sqrt{I(I+1)}, \tag{1}$$

ahol I a spinkvantumszám, $\hbar$ pedig a redukált Planck-állandó ($\hbar = h/2\pi \approx 1{,}05457 \cdot 10^{-34}$ Js). Az impulzusmomentumról ismert, hogy egyszerre csak a hossza és az egyik irányba eső vetülete mérhető (legyen ez az irány most z):

$$L_z = \hbar m, \tag{2}$$

ahol m a $-I$, $(-I+1)$, $(-I+2)$….$(+I-1)$, $+I$ értékeket veheti fel. Ezek a tulajdonságok azért fontosak, mert a mágneses momentum arányos az impulzusmomentummal, az arányossági tényező pedig a giromágneses állandó (γ), ami az egyes magfajtákra jellemző (^{1}H magra az értéke $26{,}7 \cdot 10^{7}$ T^{-1}s^{-1}, ^{13}C-ra $6{,}73 \cdot 10^{7}$ T^{-1}s^{-1}, ^{15}N-re pedig $-2{,}71 \cdot 10^{7}$ T^{-1}s^{-1}):

$$\mu = \gamma\hbar\sqrt{I(I+1)}, \tag{3}$$

$$\mu_z = \gamma\hbar m. \tag{4}$$

A (3) egyenletből látszik, hogy ha a mag spinkvantumszáma zérus, akkor nincs mágneses momentuma, azaz NMR spektroszkópiával nem vizsgálható. A (4) egyenlet alapján a mágneses momentumnak $(2I+1)$-féle állapota lehetséges: ezek az állapotok degeneráltak, az egyes magok mágneses momentumai bármilyen térállásban lehetnek, így kiátlagolódnak és nincs makroszkopikus mágnesezettség.

Egy külső konstans B_0 mágneses tér azonban a mágneses állapotok energiáinak felhasadását okozza (Zeeman-felhasadás): mivel a magok paramágnesesek, a B_0 térrel párhuzamos állás lenne a legkedvezőbb energetikailag. A mágneses momentum tulajdonságaiból adódóan azonban ez nem valósulhat meg, így az a B_0 tér tengelyével (ami konvenció szerint z-tengely irányú) csak olyan szöget zárhat be, hogy a z-tengelyre eső vetületek a (4) egyenlet által definiált diszkrét értékeket vegyék fel. Ekkor, mivel a mágneses momentumok már nem lehetnek tetszőleges térállásban, makroszkopikus mágnesezettség is megjelenik a z-tengellyel

párhuzamosan, az egyes mágneses momentumok összegeként. A B_0 térnek egy másik fontos hatása, hogy a mikroszkopikus mágneses momentumok a tér tengelye körül forognak, amit Larmor-precessziónak nevezünk. A forgás frekvenciája (ν_0):

$$\nu_0 = \gamma B_0 . \tag{5}$$

Kvantummechanikai megfontolások alapján két állapot között csak akkor jöhet létre átmenet, ha azok szomszédosak. Két szomszédos állapot közti energiakülönbség a

$$\Delta E = \gamma \hbar H_0 \tag{6}$$

egyenlettel adható meg, vagyis az átmenet frekvenciája (ν):

$$\nu = \gamma B_0 . \tag{7}$$

Az (5) és a (7) egyenletek összehasonlításából kitűnik, hogy a Larmor-frekvenciával azonos frekvenciájú sugárzással lehet a mágneses állapotokat gerjeszteni. Ezt egy B_1 rádiófrekvenciás pulzus segítségével valósítjuk meg, ami a B_0 térre merőleges, és mivel frekvenciája a magokéval azonos, azok önmagukhoz képest „állónak” érzékelik és precesszálni kezdenek a tengelye mentén. Ez a precesszálás megfelel a két kvantumállapot közötti átmenetnek.

A modern NMR-spektroszkópia Fourier-transzformációs technikát alkalmaz. A pulzusgerjesztés hatására a makroszkopikus mágnesezettség elmozdul a *z*-tengelyről, attól függő mértékben, hogy a B_1 tér mennyi időn keresztül hatott rá. Ha például populáció inverzió jött létre, akkor a makroszkopikus mágnesezettség a *z*-tengely ellentétes oldalára kerül. Az ilyen pulzust 180°-os pulzusnak nevezzük. 90°-os pulzusnál a makroszkopikus mágnesezettség *z* komponense megszűnik, mivel a gerjesztett és az alapállapotban ugyanannyi molekula található, ellenben *x* vagy *y* komponensre szert tesz. A pulzus kiadása után a gerjesztett állapotok relaxálni kezdenek, visszakerülnek az alapállapotba, ennek következtében a makroszkopikus mágnesezettség *z* komponense egyre nő, *x* és *y* komponense pedig idővel nullára csökken. Az *x* és az *y* komponens nagyságát az időben egy szinusz (vagy koszinusz) görbe és egy exponenciálisan lecsengő görbe szorzata írja le. Ezt az exponenciálisan lecsengő jelet, a FID-et (free induction decay) detektáljuk az időben. A FID időfüggvény, belőle frekvenciafüggő spektrumot Fourier-transzformációval nyerhetünk.

Az egydimenziós NMR-spektrumokból nyerhető egyik fő információ az úgynevezett kémiai eltolódás. Az NMR-spektroszkópia szerkezetkutatásban való felhasználhatóságának az alapja

ugyanis, hogy egy molekulában az azonos kémiai minőségű magoknak nem azonos a frekvenciája, mert a rájuk ható B_0 teret a magok kémiai környezete befolyásolja:

$$\nu_{helyi} = \gamma B_{helyi} = \gamma B_0 (1 - \sigma), \tag{8}$$

ahol σ az árnyékolási tényező. A frekvenciák számértékét a B_0 tér nagysága befolyásolja, így különböző készülékeken különböző spektrumokat lehetne felvenni. Az egységesítés céljából nem a konkrét frekvenciaértékeket ábrázolják a spektrumon, hanem a kémiai eltolódásokat (δ), melyek definíciója:

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \cdot 10^6 \text{ ppm}, \tag{9}$$

ahol ν_{ref} valamilyen referenciavegyület (szén és proton esetében tetrametil-szilán, TMS) rezonanciafrekvenciája.

Az egyes magok a tőlük pár, jellemzően 1-3 kötésre levő mágnesesen aktív magokkal skaláris csatolásban lehetnek, de ismert jelenség az egymástól ennél több kötéssel elválasztott magok közti távolható csatolás is (például allilcsatolás). Ennek a csatolásnak a hatására az adott mag jele felhasad, attól függően, hogy mennyi és milyen spinű maggal, továbbá milyen erősségű csatolásban van: így jönnek létre a spektrum finomszerkezetét adó multiplettek. Egy feles spinű csatoló mag hatására dublett, két (mágnesesen ekvivalens) mag hatására triplett, stb. jön létre. Ha több, mágnesesen nem ekvivalens maggal áll fenn csatolás, bonyolultabb multiplettek is létrejöhetnek (pl. dublett triplettje). A csatolás erősségét jellemző állandó a *J* csatolási állandó, melyet elsőrendű csatolás esetén a megfelelő multiplett vonalak távolsága adja meg [3].

2.1. A Bruker vállalat *Small Molecule Structure Elucidation* alkalmazása

A *Small Molecule Structure Elucidation* egy, az atommagok közötti korrelációk jelenlétén alapuló, félig automatikus NMR-spektrumértékelő szoftveres eszköz, és mint ilyen, nem függ semmilyen adatbázistól. A segítségével előállítható az összes olyan szerkezet, ami megfelel a felhasznált NMR-spektrumok által támasztott feltételeknek.

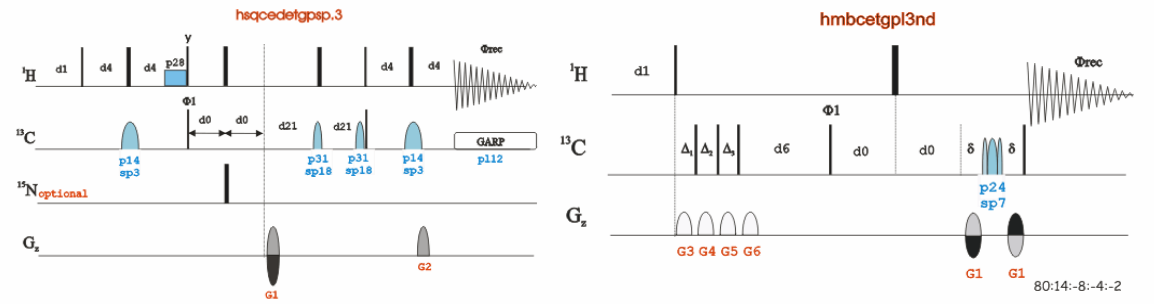
Jelen formájában a program szerves kismolekulákra alkalmazható, komplexeket és makromolekulás szerkezeteket nem támogat. A vizsgált molekula szén- és hidrogénatomokon kívül tartalmazhat oxigén-, nitrogén-, kén-, jód-, bróm-, klór-, fluor-, foszfor-, szilícium- és

bóratomot, bár a fluor- és foszfor-tartalmú vegyületek felhasználói közreműködést igényelnek a távolható csatolások miatt. Ezeken kívül más elemet tartalmazó vegyület nem vizsgálható. A funkciós csoportokat tekintve szintén van némi korlátozás: a kéntartalmú funkciós csoportokat és az esetleges nitrocsoportot ismerni kell és meg is kell adni az elemzés során; az izonitril, tioizonitril, diazo, azid és amin-oxid funkciós csoportot tartalmazó vegyületek nem vizsgálhatók.

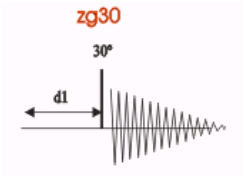
A programnak három féle spektrumra feltétlenül szüksége van a vizsgálni kívánt anyagról: ^{1}H, ^{1}H/^{13}C HSQC (Heteronuclear Single-Quantum Correlation spectroscopy) és ^{1}H/^{13}C HMBC (Heteronuclear Multiple-Bond Correlation spectroscopy) spektrumokra. Ezen kívül ismerni kell a vegyület pontos összetételét, ami például egy tömegspektrometriás mérésből ismert pontos tömegből kiszámítható. Fontos tudni, hogy kevés hidrogénatomot tartalmazó vegyületek esetében megtörténhet, hogy nem kapunk lehetséges szerkezetet, mivel az analízis a hidrogénekkel való korrelációk meglétén alapul. A már említett spektrumokon felül a program kiegészítő információkat képes nyerni a ^{13}C, ^{1}H/^{1}H COSY (Correlation Spectroscopy), ^{1}H/^{15}N HSQC és ^{1}H/^{15}N HMBC spektrumokból, ezért ezek felvétele is ajánlott. A spektrumok jó felbontása és artifaktum-mentessége kritikus a jó eredményekhez, ezért a spektrumok felvételéhez ajánlott pulzusprogramok neve mellett erre a célra optimált paraméterkészletek is rendelkezésre állnak (1. táblázat). Természetesen az ajánlott felbontásnál kisebb is elegendő, ha nem kell a jelek átfedésére számítani. Az ajánlott pulzusprogramok az 1. ábrán láthatóak.

1. táblázat. A szükséges spektrumokhoz ajánlott pulzusprogramok és paraméterkészletek.

	^{1}H spektrum	^{1}H/^{13}C HSQC spektrum	^{1}H/^{13}C HMBC spektrum
ajánlott pulzusprogram neve	zg30	hsqcedetgpsp.3	hmbcetgpl3nd
optimalizált paraméterkészlet	CMCse_1H	CMCse_HSQC	CMCse_HMBC
ajánlott felbontás	-	^{1}H dimenzió:2048 ^{13}C dimenzió: 400	^{1}H dimenzió:4096 ^{13}C dimenzió: 512



1/1.ábra.HSQC (balra) és HMBC (jobbra) spektrum felvételéhez használt pulzusprogram.



1/2. ábra. Protonspektrum felvételéhez használt pulzusprogram.

Az általános protonspektrumok felvételéhez használt pulzusprogram a lehető legegyszerűbb: egy *d1* késleltetési időt követő 90°-os rádiófrekvenciás pulzus hozza létre a két mágneses állapot közti átmenetet, amit a FID detektálása (akvizíciós idő) követ. A *d1* késleltetési idő beiktatására azért van szükség két pulzus között, hogy a rendszer ezalatt az idő alatt tudjon relaxálni, azaz a mágneses állapotok betöltöttsége visszaálljon a gerjesztés előtti állapotba. Ez garantálja egyrészt, hogy a rendszer nem kerül telítésbe, másrészt, hogy a mérés kvantitatív. A használt zg30 pulzusprogram ezzel szemben nem 90°-os, hanem 30°-os gerjesztő pulzust használ, aminek előnye, hogy mivel a 30°-os pulzusnak kisebb mértékű a gerjesztő hatása, a rendszernek kevesebb időre van szüksége a relaxációhoz, így kisebb *d1* értékkel is kvantitatív spektrum nyerhető, azaz a mérés gyorsabb [2].

A HSQC kísérletet először G. Bodenhauser és J. Ruben írta le 1980-ban [4]. A kétdimenziós spektrum két tengelye a két kémiai eltolódás skála, a benne lévő jelek alapján a kötésben lévő hidrogén- és szénatomok rendelhetők egymáshoz (az $^{1}J_{CH}$ csatolási állandóval rendelkező magpárok adnak jelet). A pulzus tartalmaz két INEPT (Insensitive Nuclei Enhancement by Polarization Transfer) szekvenciát is. Ez a szekvencia az érzéketlen magok mérésére nyújt hatásos megoldást: mivel egy magról mérhető spektrum intenzitása arányos a mag giromágneses állandójával, és ez az állandó a protonok esetében a legnagyobb, az INEPT a protonok mágnesezettségét „viszi át” a másik, kevésbé érzékeny magra, ezáltal növelve az intenzitást. Az evolúciós időt követő második INEPT szekvencia hatására a mágnesezettség visszakerül a protonra, és ezen a magon detektálódik, a szénmagok lecsatolása mellett (ami ebben az esetben a GARP pulzus segítségével történik).

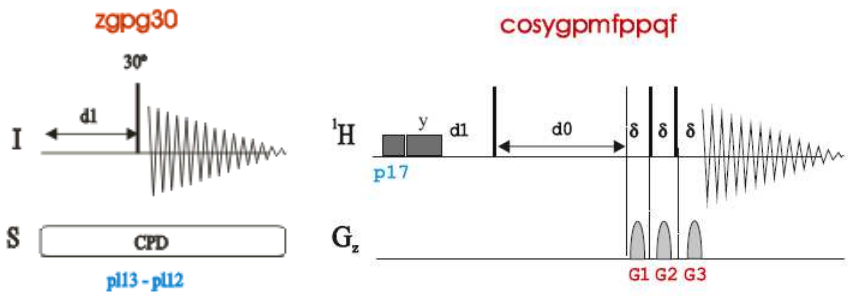
Az itt használt pulzusprogram multiplicitás szerint szerkesztett HSQC. Ez azt jelenti, hogy a spektrum a keresztcsúcsok multiplicitásáról is tartalmaz információt: a CH- és CH_3-csoportok jelei ellentétes fázisúak lesznek a CH_2-csoport jeleivel, vagyis fáziskorrekció után előbbiek pozitív, míg utóbbiak negatív előjelűek lesznek. A program a CH- és CH_3-csoportok között a HSQC csúcsainak integrálértékei alapján tesz különbséget (manual). A multiplicitás szerint szerkesztett HSQC pulzusszekvencia W. Willker *et al.* [5] munkája alapján készült, felhasználva R. Boyer *et al.* [6] intenzitást javító, valamint C. Zwahlen *et al.* [7] a csatolási állandók nagysága alapján történő szűrést célzó kiegészítéseit.

A HMBC kísérlet A. Bax és M. Summers nevéhez fűződik [8]. A kapott spektrum hasonló a HSQC spektrumhoz, csak itt jellemzően a 2-3, ritkán 4 kötésen keresztül érvényesülő csatolásban ($^2J_{CH}$, $^3J_{CH}$, $^4J_{CH}$) lévő magok adnak jelet. A pulzusprogramban szereplő késleltetési idők a detektálható csatolási állandó értékkel szorosan összefüggnek, és mivel a 2-4 kötéses csatolási állandók széles tartományt lefednek (szemben az egykötéses csatolási állandókkal, amik jellemzően egy érték körül helyezkednek el), nem detektálható minden esetben az összes ilyen csatolás. További problémát jelent a direkt (egykötéses) csatolások esetleges feltűnése a proton kémiai eltolódás tengellyel párhuzamosan, a (meg nem jelenő) keresztcsúcstól körülbelül 130-160 Hz-nyire elhelyezkedő dublettként. Ezeket ki kell szűrni a spektrumból. A HMBC spektrum méréséhez használt pulzusprogram a D. Cicero *et al*. által javasolt módosításokat tartalmazza [9]. A program a HMBC spektrumban elkerülhetetlenül megjelenő fáziskülönbségekre érzékeny, ezért ezek a spektrumok magnitúdó módban kerülnek felhasználásra: spektrumot önmagával megszorozva, majd az eredményből gyököt vonva csak pozitív csúcsok maradnak (ennek a spektrumformának természetes hátránya, hogy a jelek sokkal szélesebbek lesznek).

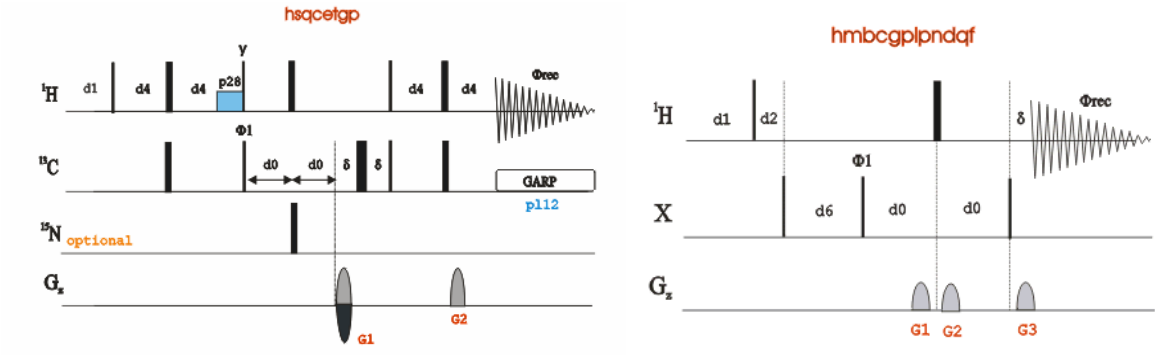
A már említett további ajánlott, de a program használatához nem feltétlenül szükséges spektrumok felvétele további struktúrális megkötéseket ad a fenti mérésekből nyerhetőkhöz, ezáltal a generált szerkezetek száma csökkenthető. Az ilyen spektrumok a 2. táblázatban találhatók, a megfelelő pulzusprogramok pedig a 2. ábrán láthatók.

2. táblázat. Az ajánlott, de nem szükséges spektrumok.

	^{1}H/^{1}H COSY spektrum	^{13}C spektrum	^{1}H/^{15}N HSQC spektrum	^{1}H/^{15}N HMBC spektrum
ajánlott pulzusprogram neve	cosygpmfppqf	zgpg30	hsqcetgp	hmbcgplpndqf
optimalizált paraméterkészlet	CMCse_COSY	CMCse_13C	CMCse_ 15NHSQCf2	CMCse_ 15NHMBCf2



2/1. ábra. ^{13}C spektrum (balra) és ^{1}H/^{1}H COSY spektrum (jobbra) felvételéhez ajánlott pulzusprogramok.



2/2. ábra. $^{1}H/^{15}N$ HSQC (balra) és HMBC (jobbra) spektrum felvételéhez ajánlott pulzusprogram.

A szénspektrum felvételéhez használt pulzusprogram a protonspektrum pulzusprogramjához nagyon hasonló, az egyetlen különbség, hogy a protonokat (az ábrán S magként feltüntetve) a mérés alatt lecsatoljuk egy CPD (Composite Pulse Decoupling, pl. WALTZ) szekvencia segítségével. Erre azért van szükség, mivel lecsatolás nélkül a szénspektrum jelei bonyolult multiplettekként jelennek meg, hiszen sok protonnal vannak csatolásban.

A megfelelő analízishez a szénspektrum kémiai eltolódás-tartományának két szélén a teljes tartomány 10%-ának megfelelő jelmentes régiónak kell lennie. A szénspektrum jelentősége megnő, ha sok jel van kis távolságra egymástól, mivel ilyen esetben a kétdimenziós spektrumokból nem látszik egyértelműen, hogy a korrelációkat egy vagy több szén jele adja. Az APT (Attached Proton Test) [10] és a DEPT (Distortionless Enhancement by Polarization Transfer) [11] pulzusokkal is felvehető szénspektrum. Mindkét spektrum képes megkülönböztetni a szénatomokat aszerint, hogy hány proton kapcsolódik hozzájuk. Az APT spektrumban a CH_3- és a CH-csoport jelei megfelelő fázisolás esetén pozitívak, míg a CH_2-csoport és azok a szenek, amikhez nem kapcsolódik hidrogén negatívak. A DEPT spektrumban paraméterfüggő, hogy a CH_3-, CH_2-, CH-csoportok közül melyik negatív és melyik pozitív. Ezeket a spektrumokat a program képes értelmezni, de nem ajánlottak, egyrészt mert a szénspektrum körülbelüli integrálértékei is felhasználásra kerülnek, amit ezek a pulzusok torzítanak, másrészt mivel a multiplicitás szerint szerkesztett HSQC már magában hordozza ezeket az információkat. Emellett a DEPT spektrumban a kvaterner szenek nem is jelennek meg.

További előnye a szénspektrumnak, hogy az analízis egyik utolsó pontjaként lehetőség van a kapott szerkezetek számolt szénspektrumának összehasonlítására a mért szénspektrummal, és ezalapján egy valószínűségi sorrend felállítására is. Mivel a kémiai eltolódások elméleti úton való meghatározása jelenleg még elég időigényes, ugyanakkor pontatlan is, ez a sorrend csak tájékoztató jellegű, de kétségtelenül hasznos.

A $^{1}H/^{1}H$ COSY spektrum kétdimenziós, mindkét tengelyen a protonok kémiai eltolódása szerepel, keresztcsúcsot pedig azoknál a protonoknál ad, melyek egymással csatolásban vannak [12; 13]. A legelterjedtebb fajta a DQF-COSY (Double-Quantum-Filtered COSY), vagyis a kétkvantum szűrésű COSY, ami teljesen elnyomja a csatolatlan protonok nagyon intenzív jeleit. Ha ennek a spektrumnak a méréséhez gradiens pulzusokat használunk, rövid idő alatt kaphatunk tisztán abszorpciós jeleket [14; 15]. A program működéséhez szükséges, hogy a COSY jelek a fázisra ne legyenek érzékenyek, amit magnitúdó módú spektrumokkal lehet elérni.

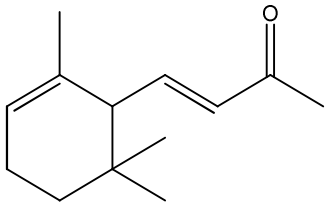
A $^{1}H/^{15}N$ HSQC esetében nincs értelme multiplicitás által szerkesztett spektrumot felvenni, ezért egy egyszerűbb szekvencia az ajánlott. A HMBC pulzust távolható csatolások (J_{NH} = 8-10 Hz) detektálására ajánlott optimálni.

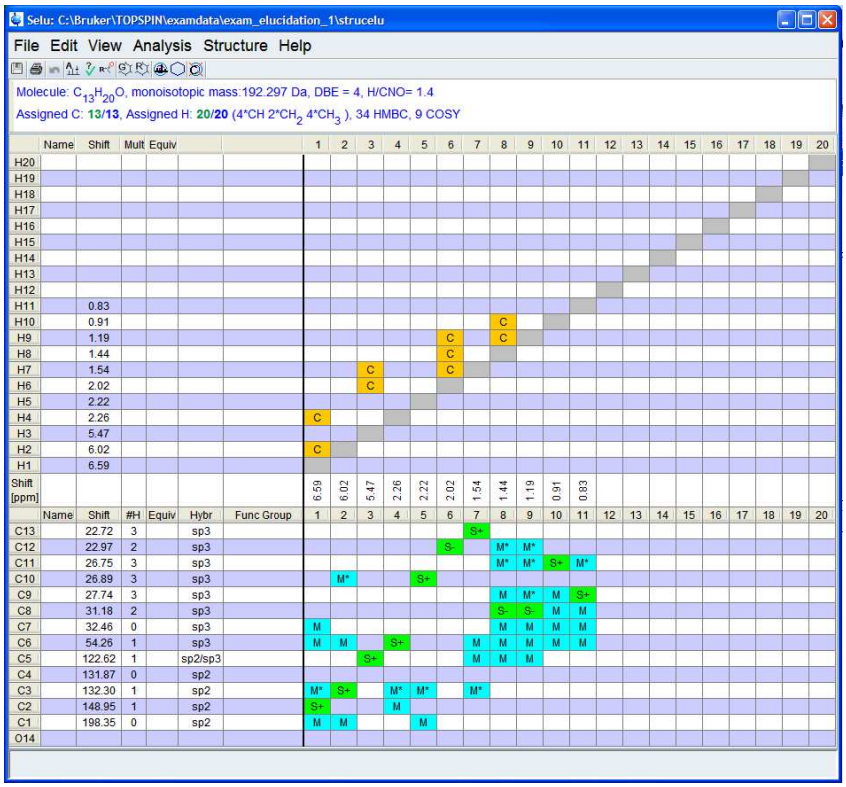
Fontos, hogy a spektrumok kalibráltak legyenek. Ezt egydimenziós spektrumoknál legegyszerűbben úgy lehet elérni, hogy a TMS jelét 0 ppm-re állítjuk, ezen kívül szükségszerű, hogy az egy- és kétdimenziós spektrumban lévő jelek kémiai eltolódásai egymással összhangban legyenek.

Ha a megfelelő spektrumok rendelkezésre állnak, a program elvégzi ezek analízisét. Először az egydimenziós spektrumok alapján készül egy csúcslista, majd a kétdimenziós spektrumokból egy korrelációs lista. A program ezután a HSQC spektrum integrálértékei alapján meghatározza a szénatomokhoz kapcsolódó protonok számát, a szénspektrum csúcsainak relatív intenzitásértékei alapján pedig a mágnesesen ekvivalens szénatomok számát. A kapcsolódó protonok számának ellenőrzésére lehet használni a protonspektrumot. A szénatomok hibridizációs állapotának meghatározása nem elengedhetetlen, de ha lehetséges, akkor a kémiai eltolódásértékek alapján történik.

A kinyert információk az úgynevezett Correlation Table-ben (korrelációs táblázatban) találhatók meg összefoglalva. A TopSpin 3.1 kézikönyvében található néhány példa, ezek közül az α-jonon mért spektrumai (proton, szén, COSY, HSQC és HMBC) alapján kapott korrelációs táblázata látható a 3. ábrán. A fejlécben a molekuláról találhatók információk (összegképlet, pontos tömeg, „H/CNO” hidrogén/egyéb atom arány, asszignált protonok és szenek száma, stb.), alatta két táblázat látható. A felső táblázatban a proton-proton korrelációk vannak jelölve a COSY spektrum alapján, narancssárga hátterű mezőben „C” betűvel. A táblázat mellett található sávban a számozott protonok mellett a hozzájuk tartozó jel kémiai eltolódása látható. Ekvivalens protonok esetében csak az egyik atomnál van feltüntetve a kémiai eltolódás és a korrelációt jelző cella, a többi atom csak fel van sorolva a felsőbb sorokban. Az alsó táblázatban a direkt-, illetve többkötéses csatolások láthatók a protonok és a

szenek között: a kék mezők „M” betűvel HMBC spektrumban jelentkező keresztcsúcsra utalnak, a zöld mezők „S” betűvel pedig HSQC spektrumbelire. Az „M*” jelölés életlen HMBC-beli jelre utal, az „S” betű mellett található „+” és „–” jelölés pedig az editált HSQC spektrum csúcsának pozitív vagy negatív voltát reprezentálja. A bal oldali sávban vannak felsorolva sorrendben a számozott szénatomok, (a felhasználó által megadható nevük), a kémiai eltolódásuk, a hozzájuk kapcsolódó hidrogénatomok száma, az ekvivalens szénatomok száma, és a valószínű hibridizációs állapot.





3. ábra. Az α-jonon és korrelációs táblázata.

A korrelációs táblázat interaktív, továbblépés előtt javasolt átnézni, hogy a rosszul interpretált korrelációkat kijavítsuk, vagy a fel nem dolgozott jelekkel kiegészítsük azt. A félreértelmezések például kalibrációs problémákra, oldószerjellel elfedett vagy egymással átfedő jelekre, szennyezőanyag jeleire, artifaktumokra (például direktcsatolások megjelenése a HMBC spektrumban), vagy fázishibákra vezethetők vissza, de akkor is problémákba ütközünk, ha van olyan csatolási állandó, ami a vártnál kisebb, tehát a keresztcsúcs kevéssé intenzív és ezért a program nem veszi figyelembe.

A szerkezetgenerálás előtt további kényszerek építhetők be: egyrészt megadható a protonok multiplicitása (szinglett, dublett, triplett, stb.), másrészt az, hogy milyen funkciós csoportnak vagy molekulafragmensnek (pl. benzolgyűrű, karbonil-csoport, stb.) kell feltétlenül szerepelnie a generált szerkezetekben. Amennyiben ezt az utolsó lehetőséget kihasználjuk, lehetőség van a molekularészlethez tartozó szén- és hidrogénatomokat asszignálni, és a következőkben a program ezekből indul ki a szerkezetgenerálásnál. Asszignáció nélkül az összes szerkezet generálódik, de csak a kijelölt részletet tartalmazók jelennek meg, asszignációval viszont csak olyan szerkezet generálódik, ami már tartalmazza a szükséges fragmenst, így az erre a lépésre fordított idő jelentősen csökkenhet. A fentiekhez hasonlóan az is beállítható, hogy milyen funkciós csoportok és molekularészletek nem jelenhetnek meg a szerkezetekben. Ennek az eszköznek egy speciális esete, hogy az is beállítható, hogy legyen-e a generált szerkezetben gyűrű, és ha igen, maximum vagy minimum milyen hosszú lehet.

Előfordulhat, hogy távolható csatolások miatt megjelenik keresztcsúcs olyan atomok között is a HMBC vagy a COSY spektrumban, amik több, mint 3 kötésre vannak egymástól. Mivel a jelek intenzitása a csatolási állandótól függ, ami nem feltétlenül arányos az atomokat elválasztó kötések számával, ezért ezek nem biztos, hogy figyelmen kívül lesznek hagyva. Viszont a program az ilyen atomokra is fellállítja azt a kényszert, hogy maximum 3 kötésre lehetnek egymástól, ami ellentmondáshoz vezet, és az igazi szerkezet nem kerül generálásra. Ennek a problémának az elkerülésére beállítható az is, hogy ezek közül a távolsági kényszerek közül maximum hányat sérthet meg az adott szerkezet anélkül, hogy kidobásra kerülne. Kiindulópontnak az összes korreláció 10%-át javasolják megsérthetőnek, de ez a molekula nagyságával változhat, mivel kevés atom között kevesebb korreláció fog megjelenni, és ekkor a 10% túl nagy flexibilitást ad a szerkezetmeghatározáshoz.

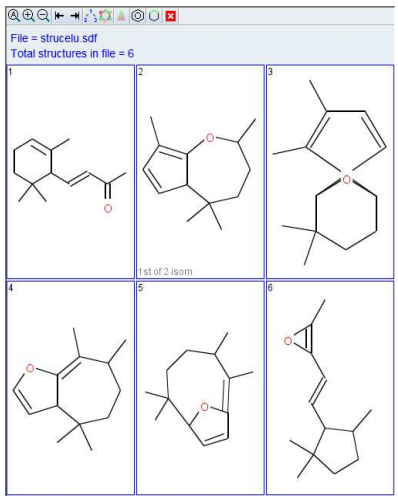
A korrelációk eliminálásával kapcsolatban beállítható egy integrálérték, ami alatt a csúcsokat a program nem veszi figyelembe egyáltalán. Ezen az értéken felüli csúcsok kezelésénél három mód közül választhatunk: a „optimal”, „fast” és „exact” módok közül. „Optimal” mód esetén a nagyobb integrálértékű csúcsok nem sérthetők meg semmiféleképpen, a „fast” mód ehhez hasonló, csak kevesebb csúcsot tart sérthetetlennek, az „exact” módban pedig az integrálérték nem befolyásolja az eliminálhatóságot [2].

A szerkezetek generálásához két algoritmus közül lehet választani: az LSD (Logics for Structure Determination), ami J. M. Nuzillard nevéhez köthető [16] és a Bruker saját szoftvere közül. A kettő között több különbség van, például hogy a Bruker szofvere képes a bóratomok, a kéntartalmú funkciós csoportok (az LSD csak az R_1-S-R_2 szulfidcsoportot támogatja) és az *sp* hibridizációs állapot kezelésére, valamint a fragmensek parciális

asszignációja is csak a Bruker szoftverével lehetséges. Különbség még, hogy azt az információt, mely szerint mágnesesen ekvivalens atomok vannak a molekulában, az LSD csak a szerkezetgenerálás után veszi figyelembe szűrőként, a Bruker szoftver viszont már a szerkezetépítésnél. Olyan esetekben viszont, ahol nincs ilyen, az LSD algoritmus gyorsabb lehet a Bruker szoftvernél. Beállítható az is, hogy maximum hány szerkezet kerüljön generálásra, vagy az is, hogy a program maximum mennyi időt tölthessen a szerkezetek előállításával.

Ha a szerkezetgenerálás nem hoz találatot, az származhat egymásnak ellentmondó feltételekből (például szükséges, hogy szerepeljen benzolgyűrű a molekulában, de a megengedett legnagyobb gyűrűtagszám 5), vagy a túl kevés számú korrelációból. Utóbbi esetben hasznos lehet átnézni a meglévő korrelációkat, valamint a gyengébb és ezért eddig nem vizsgált jeleket is bevenni. Ha nagyszámú találat van (akár több ezer), több kényszerfeltételt kell definiálni, akár újfajta spektrumok felvételével, hogy a találatok számát ezzel csökkentsük. Ha kevés és megbízhatónak tűnő találatot kapunk, akkor is érdemes újra elvégezni a szerkezetgenerálást egyes feltételek kihagyásával, hátha ekkor újabb fajta használható szerkezeteket találunk.

A kapott találatok egy sdf kiterjesztésű fájlban találhatók. Ennek tartalmát meg lehet jeleníteni a programon belül, például az α-jononra kapott találatok a 4. ábrán láthatók, de egy pdf fájl is készül a szerkezetekről. Bár az eredeti vegyület csak egy hattagú gyűrűt tartalmaz, a találatok között több, összetettebb gyűrűrendszert tartalmazó szerkezet is van, például a 2. és a 4. találat kondenzált gyűrűs, az 5. pedig áthidalt gyűrűs.



4. ábra. Az α-jononra generált szerkezetek.

A 2. találat ablakának alsó részén a „1st of 2 isom” felirat látható. Ez azt jelenti, hogy az adott szerkezethez két izomorf is tartozik, azaz két megegyező konnektivitású szerkezet, amiknek csak az asszignációja tér el egymástól.

Az egyes találatok valószínűségét kétféleképpen is lehet ellenőrizni. Egyrészt bármelyik szerkezet kereshető a University of Vienna szerverén [17], és ha az tartalmaz erre a szerkezetre NMR-spektrumot, az meg is jelenik. Másrészt, az összes (vagy csak több kiválasztott) találatra lehet generálni szénspektrumot, amik a mért spektrummal automatikusan összehasonlításra kerülnek, és a kettő közötti eltérés szórása alapján a találatok egyfajta sorszámot kapnak. Tipikusan a helyes szerkezetnek 5 ppm-nél kevesebb lesz a

szórása, a 10 ppm-es szórásnál nagyobb értékek pedig nagyon valószínűtlenek. Mégis, az NMR-spektrum számolás bizonytalan, ezért az így kapott sorrend csak tájékoztató jellegű, valamint a spektrumok számolása főleg nagy számú találat esetén hosszú időt vehet igénybe. A felhasználó is felállíthat saját rangsort, ami előtt ajánlott a figyelembe vett korrelációk teljesülését ellenőrizni a szerkezeteken, valamint a szerkezetgenerálás során kihagyott korrelációkról ellenőrizni, hogy valóban artifaktumok vagy távolható csatolások eredményei voltak-e.

Az sdf (structure-data file) formátum egy, a Molecular Design Limited Information Systems által fejlesztett és elterjedten használt formátum, ami egyszerre több kémiai szerkezetről képes információt tárolni. Az egyes szerkezetek egymástól a „$$$$” jel választja el. A *Small Molecule Structure Elucidation* output fájljának általános felépítésének vázlata az 5. ábrán látható.

Az első három sor az sdf formátumban az adott szerkezet fejléce; ebben az esetben minden szerkezetnél a „CDK” karaktersor és a fájl készültének dátuma szerepel. A negyedik sorban az első szám az atomok számát, a második szám a kötések számát adja meg (többszörös kötések is csak egy kötésnek számítanak). Az ezt követő blokk az atomok térbeli helyzetét írja le: minden atomra egy sorban szerepelnek az *x*, *y* és *z* térkoordináták a Descartes koordinátarendszerben, majd a negyedik oszlopban az atomok kémiai minősége található. Ezután a kötések leírása következik: minden kötésre az első két oszlopban található annak a két atomnak a sorszáma, amiket összeköt (a sorszám az előző blokkban a felsorolás sorrendjével egyezik meg), a harmadik oszlopban pedig hogy hányszoros kötésről van szó. Ezt a szakaszt a „M END” tartalmú sor zárja le, ezt követően pedig a szerkezet tulajdonságai következnek [18]. Ebben az output fájlban a „> <B_NISO>” sor után a már említett izomorfok száma található, a „> <B_ISOMAP>” részben pedig ezeknek az izomorfoknak a leírása (melyik atom asszignációja cserélhető fel melyik másik atoméval) [2].

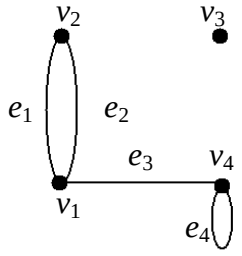
```
  CDK    2/9/12,3:24
 14 14  0  0  0  0  0  0  0  0999 V2000
   -3.9803   -0.0000   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.9803   -1.5320   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.6535   -2.2980   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.3268   -1.5320   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.3268   -0.0000   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.6535    0.7660   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000   -2.2980   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000   -0.7660   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.6535    2.2980   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.7660   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.3268   -0.0000   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.6535    0.7660   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.9803   -0.0000   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.6535    2.2980   -0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  1  0  0  0  0
  3  4  1  0  0  0  0
  4  5  1  0  0  0  0
  5  6  1  0  0  0  0
  6  1  2  0  0  0  0
  4  7  1  0  0  0  0
  4  8  1  0  0  0  0
  6  9  1  0  0  0  0
  5 10  1  0  0  0  0
 10 11  2  0  0  0  0
 11 12  1  0  0  0  0
 12 13  1  0  0  0  0
 12 14  2  0  0  0  0
M  END
> <IDENTICALS>

> <PRCGuess>

> <B_ISOMAP>

> <B_NISO>

$$$$
```

5. ábra. Az sdf formátumú output fájl felépítésének vázlata.

Fontos megjegyezni, hogy a program csak a molekulát alkotó atomok kapcsolódási sorrendjét (konnektivitását) tudja meghatározni. A molekula térbeli tulajdonságairól (konformáció, kiralitás, stb.) ezzel a módszerrel információ nem nyerhető, bár önmagában az NMR spektroszkópia széleskörűen használható ilyen célokra, elég csak azt megemlíteni, hogy fehérjék térszerkezetének meghatározásában is egyre nagyobb szerepet játszik.

2.2. A gráfelmélet és alkalmazása kémiai problémákban

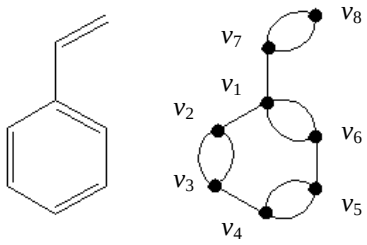
A gráfelmélet története a XVIII. századba nyúlik vissza. Az első dokumentált gráfelméleti problémával Kalinyingrád lakosai fordultak Eulerhez, aki akkor a szentpétervári akadémián tanított: a várost a Pregel folyó több részre osztotta, és a folyón hét híd épült. A kérdés az volt, hogy lehet-e olyan sétát tenni, hogy az útvonal minden hídon, de egy hídon csak egyszer haladjon át. Euler bebizonyította, hogy ilyen útvonal nem létezik, bizonyítása pedig az első gráfelméleti dokumentációvá vált. Ezután az 1840-es években Kirchhoff az elektromos hálózatokat próbálta gráfokként kezelni, és I. és II. törvényei, valamint a gráfok tulajdonságai között fontos összefüggést fedezett fel. Ez annyira jelentős volt, hogy lineáris elektromos hálózatok analízisének ma is ez az alapja. Meghatározó fontosságú volt még Cayley munkássága is a molekuláris diagramok területén. Mára a gráfelmélet önálló tudományággá fejlődött, eredményeit elterjedten használják (például nyomtatott áramkörök tervezésében, vagy a híres „utazó ügynök”-probléma esetében) [19].



6. ábra. Példa irányítatlan gráfstruktúrára.

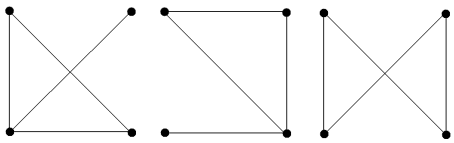
Gráfnak (6. ábra) definíció szerint egy rendezett párt, $G(V,E)$-t nevezzük, ahol V egy nem-üres halmaz, aminek elemeit pontoknak vagy csúcsoknak (v) nevezzük, E pedig ebből a halmazból képezhető párok halmaza, amiket éleknek (e) nevezünk. Az $e \in E$ él a $\{v_1, v_2\}$ pontpárnak felel meg, vagyis ez a két pont az él két végpontja. Ha a két pont megegyezik ($v_1 = v_2$), akkor az élet hurokélnek nevezzük (például ilyen a 6. ábrán az e_4 él), ha két pont között több él fut, akkor párhuzamos vagy többszörös élekről beszélünk (e_1 és e_2 él). Ha az éleknek van iránya, vagyis a hozzájuk tartozó pontok meg vannak különböztetve (egyik pontja a kezdőpontja, a másik a végpontja), akkor a gráf irányított gráf, más esetben irányítatlan. Az egy pontra illeszkedő élek számát a pont fokszámának nevezzük; ha ez 0, akkor a pont izolált pont (v_3). Izolált pontot nem tartalmazó gráfot összefüggőnek nevezünk. [20].

A definíció alapján látható, hogy a kémiában nap mint nap használunk gráfokat, anélkül, hogy akként tekintenénk rájuk: a felrajzolt szerkezetekben az atomok a csúcsoknak, a kötések pedig az éleknek felelnek meg. A molekula-szerkezetek tehát összefüggő irányítatlan gráfoknak tekinthetők, mint ahogy azt a 7. ábra mutatja (a gráf pontjainak számozása nem egyezik meg a sztirol szénatomjainak konvencionális kémiai számozásával).



7. ábra. A sztirol molekula-szerkezete (balra) és ugyanez a szerkezet gráfként felrajzolva (jobbra).

Az ilyen konstitúciós gráfokon kívül használatosak még úgynevezett reakció-gráfok. Ezekben a csúcsok vegyületeknek felelnek meg, az élek pedig elemi reakciólépéseknek, és reakciókinetikai problémák megoldására használhatók. Gráfok alkalmazhatók kvantitatív szerkezet-aktivitás (QSAR, Quantitative Structure-Activity Relationship) és kvantitatív szerkezet-tulajdonság (QSPR, Quantitative Structure-Property Relationship) összefüggések meghatározásánál is. Az elemi szén és a fullerének vizsgálatához pedig végtelen molekuláris gráfok nyújthatnak segítséget [21]. A molekulák gráfként való kezelését több területen is használják, és főként az izomorf és a részgráfok tulajdonságainak van nagy jelentősége. Ezen kívül a szerves vegyületek körében nagyon sok a gyűrűs molekula, amik körökként jelennek meg a gráfban. Ezeknek a fogalmaknak szükséges tehát a tisztázása.



(a) ***(b)*** ***(c)***

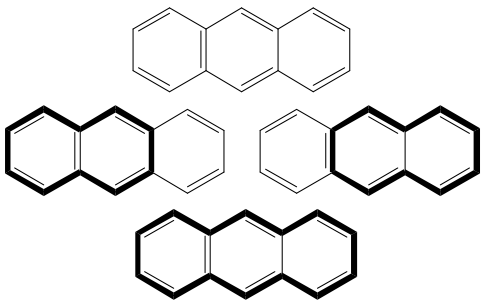
8. ábra. Izomorf gráfok (*a* és *b*).

A $G(V,E)$ és a $G'(V',E')$ gráfok izomorfak, ha van olyan egyértelmű megfeleltetés a pontjaik között, hogy G-ben pontosan akkor szomszédos két pont, ha G'-ben is szomszédosak a nekik megfelelő pontok, és a szomszédos pontpárok között ugyanannyi él fut. Például a 8. ábrán az *(a)* és a *(b)* gráfok izomorfak egymással, de a *(c)* gráffal nem. A $G''(V'',E'')$ gráfot $G(V,E)$ részgráfjának nevezzük, ha $V'' \in V$ és $E'' \in E$, valamint egy pont és egy él pontosan akkor illeszkedik egymásra G''-ben, ha G-ben is illeszkednek. A $G''(V'',E'')$ részgráf feszítő részgráf, ha $V''=V$. A $(v_0,e_1,v_1,e_2,v_2 \ldots v_{k-1},e_k,v_k)$ sorozatot élsorozatnak nevezzük, ahol e_i a v_{i-1} és v_i pontokat köti össze. Ha egy ponton sem megy keresztül egynél többször (vagyis minden élen is csak egyszer), akkor ezt az élsorozatot útnak nevezzük v_0 és v_k között. Ha minden él és minden csúcs különbözik és $v_0 = v_k$, akkor az élsorozat egy kört alkot. A kör hosszán az azt alkotó élek számát értjük. A kört nem tartalmazó összefüggő gráfot fának nevezzük.

A gráfokat célszerű mátrixreprezentációikként tárolni. A két legkézenfekvőbb mátrixreprezentáció a szomszédsági és az illeszkedési mátrix. Az *A* szomszédsági mátrix egy *n* csúcsot tartalmazó gráfra $n * n$-es: az a_{ij} eleme 0, ha az *i*-edik és a *j*-edik csúcs nem szomszédos; *k*, ha az *i*-edik és a *j*-edik csúcs között *k* darab él fut; *l*, ha $i = j$, és hozzá *l* darab hurokél csatlakozik. A *B* illeszkedési mátrix egy *n* csúcsot és *e* darab élt tartalmazó mátrixra $n * e$-es: b_{ij} eleme 0, ha a *j*-edik él nem illeszkedik az *i*-edik pontra; 1, ha illeszkedik rá. [19; 20]. Például a sztirol vegyület *A* szomszédsági és *B* illeszkedési mátrixa:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 2 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \end{pmatrix}; \; B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Molekuláris gráfokat használhatnak számítógéppel végzett szintézis tervezéshez, valamint egy adott szerkezet adatbázisban való kereséséhez is. Utóbbinál az izomorfizmus problémát jelent: az adott szerkezetet és a lehetséges találatot reprezentáló gráfokról be kell látni, hogy izomorfak-e. Ez a feladat nagyon időigényes lehet, akár *v*! számú esetet is meg kell vizsgálni. Ehhez hozzávéve, hogy az adatbázisokban nagyon sok szerkezet szerepel, a keresés hosszú ideig is eltartana, ezért szükséges, hogy a lehetséges szerkezetek közül a lehető legtöbbet kizárjuk valamilyen információ alapján. A gyűrűinformáció alkalmas erre a célra, ráadásul a körök a gráfelmélet egyik legfontosabb objektumai. A gráfelméletben, de a kémiai irodalomban is többféle körkeresési algoritmust írtak le, utóbbiban azonban meglehetősen sok téves állítás szerepel, egyrészt a helytelen matematikai terminológia használat következtében, másrészt mivel a szerzők a „kémiailag értelmes” gyűrűk megtalálására és definiálására törekedtek, amit viszont a gyűrűk felhasználásának célja befolyásolhat [22; 23].



9. ábra. A matematikai és a kémiai szemlélet eltérése.

A „kémiailag értelmes” gyűrű kifejezésnek azért van létjogosultsága, mivel nem minden matematikailag jelen lévő kör felel meg gyűrűnek a kémiai szemlélet szerint. Például az antracén (9. ábra) egy policiklusos aromás szénhidrogén, amit (kémiai szemlélet szerint) három benzolgyűrű alkot. Ezzel

szemben matematikailag hat gyűrű van jelen benne: a három hatos gyűrűn kívül két tíztagú gyűrű és egy tizennégytagú gyűrű.

Az, hogy a gráfban van-e kör, megjelenik a szomszédsági mátrix hatványaiban: az *A* mátrix *k*-adik hatványában az $a_{ij}^{(k)}$ elem a *k* hosszúságú, v_i és v_j közötti élsorozatok számát adja meg, így a diagonális elemek a *k* hosszúságú körök számát mutatják (egy kör az összes pontjának megfelelő diagonális elemben megjelenik kétszer, mivel két irányból is lehetséges az út). Ez a módszer azonban egyrészt sok időt igényel, másrészt nem ad felvilágosítást arról, hogy az adott körnek mely pontok az elemei pontosan (mivel ha két, ugyanolyan hosszú kör is van, akkor a tagjaik nem különíthetők el). Általánosabb módszer, ha a gráfot pontról pontra úgymond bejárjuk, és a szerkezetét így derítjük fel. Erre a célra két bejárási algoritmus van, a mélységi (Depth-First-Search, DFS), illetve a szélességi (Breadth-First-Search, BFS) bejárás. Mindkét bejárásnál megszámozzuk a pontokat a bejárás sorrendjében. Mélységi bejárás esetén a startpont után annak egyik szomszédja következik, majd a startpont szomszédjának egyik még be nem járt szomszédja, és így tovább. Ha egy pontnak nincs még be nem járt szomszédja, akkor az algoritmus visszalép arra a pontra, ahonnan ide jutott. A szélességi keresésben a startpont összes szomszédját megszámozzuk, majd sorban végighaladunk ezeken a szomszédokon és ezeknek a szomszédjait számozzuk meg. Már bejárt pontra ennél a keresésnél sem lépünk. Mindkét keresés eredménye egy fa lesz [19; 20].

A legtöbb körkeresési algoritmus csak a jelen lévő körök egy adott tulajdonságokkal bíró alcsoportját hivatott meghatározni, amely alcsoport bázisként használható a teljes gyűrűrendszer leírásához, így az algoritmusokat ezek alapján a tulajdonságok alapján lehet kategóriákba rendezni. A meghatározott alcsoportnak és a hozzá tartozó algoritmusnak többféle feltételnek kell eleget tennie: *egyedinek* kell lennie abban az értelemben, hogy projekciótól és az algoritmus implementációjától függetlenül ugyanazt a körhalmazt kell adnia egy szerkezetre; *komplettnek* kell lennie, tehát a teljes körrendszert le kell írnia, az összes többi körnek levezethetőnek kell lennie a körhalmaz elemeiből; *megkülönböztetőnek* kell lennie, hogy a különböző gyűrűrendszerek között könnyen különbséget tudjon tenni; valamint könnyen *kiszámíthatónak* kell lennie (ami, figyelembe véve, hogy a kémiai szerkezetek nem állnak matematikai értelemben túl sok pontból, nem feltétlenül igényel polinom időben lefutó algoritmust). Az algoritmusok legfontosabb kategóriái a következők [22; 23]:

❖ Fundamentális körök halmazát kereső algoritmusok

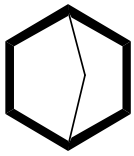
Legyen $T(V_T, E_T)$ a $G(V, E)$ gráf feszítőfája (azaz *T* feszítő részgráfja *G*-nek és fa), amit a mélységi vagy a szélességi bejárással lehet kapni. Ekkor minden olyan $e \in E$ élhez, amire igaz, hogy $e \notin E_T$, egy egyedi kör tartozik, amit a gráfelméletben fundamentális körnek nevezünk. A T-hez rendelhető összes ilyen kör alkotja a fundamentális körök bázisát. Ez a bázis általában nem teljesíti az egyediség követelményét, mivel függ a használt fától, ezért általában csak más körbázisok előállításához szolgál kiindulólépésként.

❖ Az SSSR-t (Smallest Set of Smallest Rings) és a κ-gyűrűket kereső algoritmusok

Az SSSR egy minimális hosszúságú fundamentális körbázis. (A körbázis hosszán az azt alkotó körök hosszának összegét értjük.) Az SSSR sok félreértés alapja, mivel sokan csak egy minimális hosszúságú körbázissal azonosítják, ez azonban nem feltétlenül fundamentális. Meg kell jegyezni, hogy van néhány olyan szerkezet, ami sok szimmetrikusan ekvivalens gyűrűt tartalmaz, és ezért nincs egyedi SSSR-je, de ettől eltekintve hasznos bázis. Sok algoritmus először egy fundamentális bázist generál, és ebből kiindulva határozza meg az SSSR-t, ilyen például E. Corey *et al.* [24], W. Wipke *et al.* [25], valamint J. Gasteiger *et al.* [26] munkája, amiket rendre a LHASA, a SECS és az EROS szintézisprogramokban használtak fel. J. Figueras szélességi keresésen alapuló munkáját [27] például a nyílt forráskódú Chemistry Development Kit (CDK) tartalmazza, ami a kémiai és a bioinformatikában fontos algoritmusok könyvtára [28]. Erről az algoritmusról később bebizonyították, hogy némely esetben nem generálja megfelelően az SSSR-t [23]. Ennél gyorsabb és nagyobb rendszerekre alkalmazható a C. Lee *et al.* nevéhez fűződő algoritmust is [29], de az irodalom még a felsoroltakon kívül is nagyon sok SSSR kereső algoritmust tartalmaz (például [30; 31]). Az SSSR hibáit kijavítandó definiálta M. Plotkin a κ-gyűrűk halmazát: ebbe a halmazba az összes lehetséges SSSR gyűrűt sorolta [32]. Ezt a gyűrűbázist a Chemical Information and Data System (CIDS) keretében használták szerkezetek keresésére.

❖ A β-gyűrűket kereső algoritmusok

Az SSSR egyik problémája, hogy például a norbornánban levő hattagú kört (10. ábra) nem tartalmazza, csak a két öttagút, viszont a kémiai szemléletben a hattagú gyűrű a főgyűrű. H. Nickelsen a β-gyűrűk definiálásával akarta ezt a problémát áthidalni, anélkül, hogy például az antracénben levő hatnál több tagú köröket is bevenné [23; 33]. A β-gyűrűk olyan egyszerű körök (egyszerű kör pontjait nem köti össze olyan él, ami nincs benne a körben), amelyek csak három vagy négy csúcsot tartalmaznak, vagy amelyeket nem lehet előállítani három vagy több kisebb



10. ábra. Norbornán.

egyszerű kör kombinációjaként. A felsoroltakon kívül definiálták még az ESSR (Extended Set of Smallest Rings), az SSCE (Set of Smallest Cycles at Edges) valamint a SER (Set of Elementary Rings) nevű körbázisokat is [23]. A körkereső algoritmusok utolsó nagy kategóriáját képezik az olyan algoritmusok, amik az összes kör megkeresésére irányulnak.

❖ Az összes kört megkereső algoritmusok

Ezeknél az algoritmusoknál általában a meghatározott köröket utólag sorolják olyan kategóriákba, amik segítik a kémiai jelentés megállapítását. J. Corey *et al*. munkájukban egy mélységi kereséssel előállított feszítőfa alapján adták meg a köröket, amiket ezután két fő kategóriába soroltak: a valódi és a pszeudo-gyűrűk kategóriájába. A valódi gyűrűk halmazában (1) az összes gyűrűélnek benne kell lennie úgy, hogy bármely gyűrű elhagyása a halmazból már azt eredményezi, hogy ez a feltétel nem teljesül; (2) a lehető legrövidebb köröknek kell szerepelnie; (3) az előzőeket kielégítő lehető legtöbb kört tartalmaznia kell [34].

S. Fujita nevéhez fűződik az ESER (Essential Set of Essential Rings) definiálása. Az összes kör meghatározása után azok esszenciális és nemesszenciális körökre választhatók szét. A nemesszenciális körök lehetnek kötöttek, többszörösen kötöttek, illetve függők. A kötött körök pontjait csak egy olyan él köti össze, ami nem szerepel a körben, többszörösen kötött köröknél több ilyen él van. Azt, hogy egy kör függő vagy esszenciális, befolyásolja, hogy csak szénatomot tartalmaz-e, vagy heteroatomot (N, O, S, P) is, vagy „abnormális” (minden egyéb) atomot [35; 36].

A. T. Balaban *et al*. munkája bevezette a homeomorfikusan redukált gráf (HRG, Homeomorphically Reduced Graph) fogalmát. A homeomorfia tulajdonképpen topológiai izomorfia, vagyis két objektum homeomorf, ha topológiai szempontból azonosak. Ezt a körkeresésre úgy használták ki, hogy első lépésként a gráf egy egyszerűbb homeomorfját állították elő, és a körkeresést már ezen hajtották végre [37].

T. Hanser *et al*. a redukált gráfokhoz hasonló módszert fejlesztett ki: a molekuláris gráfból (*M*-gráf) egy „útvonal gráfot” képeznek (*P*-gráf), ami megegyezik a molekuláris gráffal, de az éleknek címkéjük van. Ezután sorban távolítják el a pontokat, és azok között a pontok között, amik között az adott ponton keresztül vezetett út, a pontot egy éllel helyettesítik. A pontok eltávolítását addig folytatják, amíg az összes elfogy, ezt nevezik a gráf „összeomlásának” (collapsing). A módszer neve Collapsing *P*-graph (összeomló *P*-gráf) algoritmus [38]. Ezt később gyors és hatásos algoritmusnak találták [23].

3. Célkitűzések

Mivel a *Small Molecule Structure Elucidation* több esetben nagyon sok szerkezetet generál, fellép a találatok rendszerezésének igénye. Ugyan több lehetőség is be van építve erre a célra a szoftverben, általános megoldást egyik sem nyújt: a meglévő adatbázisban történő keresés, valamint a számolt és a mért szénspektrumok összehasonlítása egymással nagyon időigényes. A megadható beépítendő molekulafragmenst meghatározó eszköz hatékonyan leszűkítheti a találatok számát, de a használatához előzetes ismeretek kellenek a molekuláról.

Mivel sok szerves vegyület tartalmaz gyűrűs szerkezetet, és mivel ezeket elterjedten használják az egész molekula jellemzésére, ennek a meghatározása jó szempont lehet a találatok rendszerezésénél. A gyűrűket legegyszerűbben a molekulák gráfreprezentációján végrehajtott körkeresési algoritmusokkal lehet meghatározni, amikben az irodalom bővelkedik. Tekintettel a különböző alcsoport-kereső algoritmusok ismertetett bizonytalanságaira és hiányosságaira, a gyűrűk keresésére az adott algoritmusok közül olyat érdemes választani, amely az összes kört garantáltan megtalálja.

Az elmondottak miatt a szakdolgozatom céljául a *Small Molecule Structure Elucidation* által generált találatok automatikus rendszerezését elvégző algoritmus megalkotását tűztem ki Java nyelven, a *Structure Elucidation* modul sdf formátumú output fájljában található információkat használva kiindulópontnak. A rendszerezés szempontjának a struktúrában jelen lévő gyűrűrendszerek szerkezeti minőségét választottam, amihez először egy összes gyűrű megkeresésére irányuló algoritmust fejlesztettem, majd a talált gyűrűk alapján meghatároztam a gyűrűrendszer jellegét.

4. Gyűrűrendszerek jellemzése

A szerkezeti találatok elemzéséhez az 5. ábrán látható felépítésű sdf output fájlt használtam. A felépítéséből adódóan célszerű soronként beolvasni, a „CDK” karaktersort kijelölve a szerkezetelemzés startpontjának, és a beolvasást addig folytatni, amíg a fájl véget nem ér. A startpont után a második sorból első elemként kiolvasható az atomok száma, második elemként pedig a kötések száma. Az ezt követő blokk első részéből (az első *n* sorból, ahol *n* az atomok száma) meghatározható az atomok kémiai minősége, amit célszerűen egy egydimenziós tömbként tárolok, második részéből (az ezután következő *e* sorból, ahol *e* az élek száma) pedig egyfajta élmátrix, ami a konstitúciós információt hordozza, és könnyen átalakítható szomszédsági mátrixxá, amit a gyűrűkeresési algoritmus fel tud használni. Ha az élek beolvasásának végére értünk, a következő sor második elemének az „END” karaktersornak kell lennie; ha ez így van, akkor nem volt probléma az adott találat beolvasásásával. Ha a szerkezet beolvasásra került, el lehet kezdeni a körkeresést.

4.1. A gyűrűk keresésére használt algoritmus

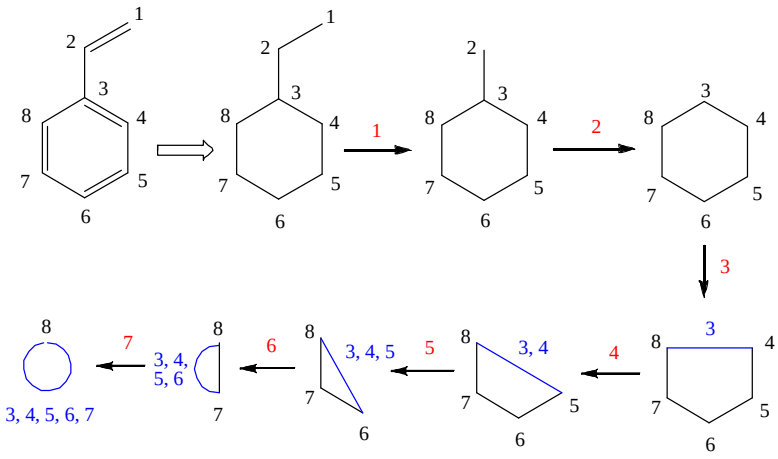
A gráfban lévő körök kereséséhez Hanser *et al*. összeomló *P*-gráf algoritmusát [38] használtam, az alkalmazás mikéntjét pedig a következőkben ismertetem. A molekuláris gráfot (*M*-gráf) első lépésben út-gráffá (path graph, *P*-gráf) kell alakítani. Ebben a *P*-gráfban az élek *M*-gráfban szereplő utaknak felelnek meg. Kezdetben a két gráf megegyezik, a különbség csupán annyi, hogy a *P*-gráfban az élekhez címkéket rendelünk. Ezek a címkék az élnek megfelelő út pontjait fogják tartalmazni, az algoritmus koncepciója ugyanis a következő: a gráfból egyesével távolít el pontokat és hozzá tartozó éleket úgy, hogy a topológiai információ megmaradjon. Ezt úgy éri el, hogy ha az eltávolítandó pont (legyen ez *x*) összeköt egy *y* és egy *z* pontot (azaz út vezet közöttük, aminek része *x*), akkor *x* eltávolításához egy új élet kell létrehozni *y* és *z* között, aminek a címkéjébe az *x* kerül. Ha a megszűnő éleknek is van már a címkéjükben valami, akkor a keletkező él címkéjébe a megszűnő él címkéjének tartalma is bekerül, kivéve, ha a két megszűnő él címkéjében szerepel azonos pont. Ekkor ugyanis nem szabad új élt létrehozni, mivel az él által reprezentált út nem volna igazi út, hiszen egy ponton többször is átmenne. Ezzel a módszerrel pontról pontra csökkenteni lehet a gráfot, egészen addig, amíg „össze nem omlik”, azaz nem marad pontja. A körök az algoritmus során hurokélekként fognak megjelenni, amik a címkéjüben tartalmazzák a kört

alkotó pontokat. Az algoritmus lépésszáma erősen függ az eltávolított pontok sorrendjétől, mivel például egy olyan pont eltávolítása során, ami nem kapcsolódott, csak egy másik ponthoz, nem jön létre új él, de egy olyan pont esetében, ami másik négy ponttal kapcsolódott, a lehetséges új élek száma akár hat is lehet. Ebből az okból célszerű a pontokat a növekvő konnektivitás sorrendjében eltávolítani [38].

Mivel a *P*-gráfban az éleknek címkéjük van, a mátrixreprezentációt (*P* mátrixot) érdemes az sdf formátumban lévőhöz (5. ábra) hasonlóan felépíteni: az első és a második oszlop tartalmazza a pontok sorszámát, amik között az él fut; a harmadik oszlop és az azt követők pedig a címkében található információkat. Célszerű a mátrixot úgy megalkotni, hogy a mátrix első (mondjuk 30) sora az első ponthoz, a második 30 sora a második ponthoz tartozó éleket tartalmazza (azaz ennek a pontnak a sorszáma legyen az első oszlopban), és így tovább, ekkor ugyanis könnyen meg lehet állapítani, hogy egy ponthoz hány él tartozik, mivel az egy ponthoz rendelhető élek egy meghatározott helyen találhatók a mátrixban. További előny, hogy így az egyes lépésekben keletkező élek számára is van hely. Ebben a mátrixban egy él kétszer szerepel, egyszer ott, ahol az egyik, másszor ott, ahol a másik pont sorszáma van az első oszlopban. Mivel a *P*-gráfot egy molekuláris gráfból állítjuk elő, a többszörös kötéseket egyszeres kötésre kell változtatni, különben azok a feldolgozás során a kettőskötésű atomok közötti kéttagú körként jelennének meg. Ez a mátrixreprezentációk szintjén annak felel meg, hogy a szomszédsági mátrixxal konstitúciós szempontból azonos szerkezeti mátrixot hozunk létre, aminek az elemei 0 vagy 1 értéket vehetnek fel: ha a szomszédsági mátrix megfelelő eleme 0, akkor a szerkezeti mátrix eleme is 0, egyébként 1.

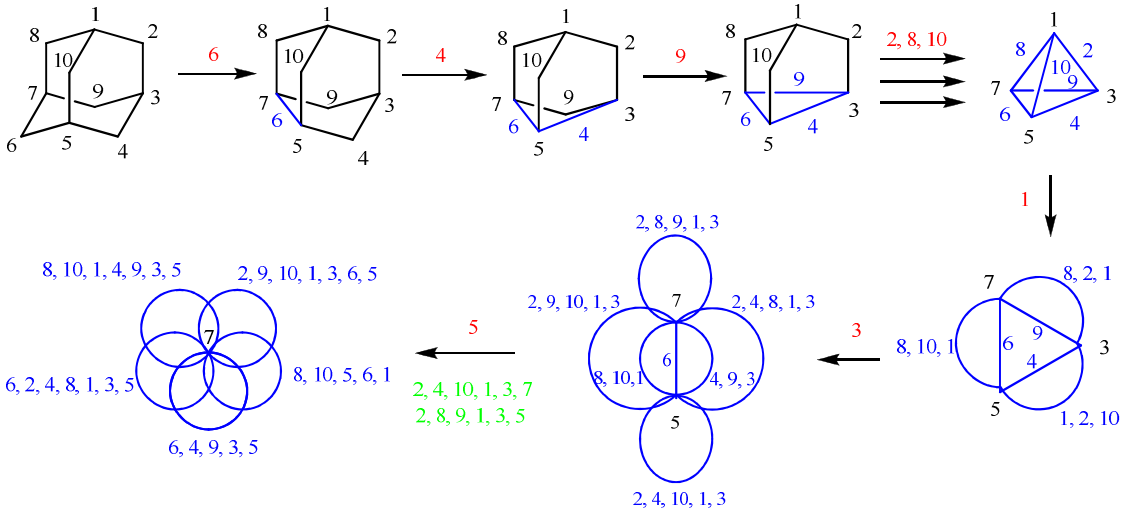
Ezután minden lépésnél a legkisebb konnektivitású pontot kell megkeresni, amíg a gráf össze nem omlik. Ezek (mivel a molekuláris gráfokban a hidrogénatomokat nem szokták feltüntetni) kezdetben a láncvégi metilcsoportok. Praktikus, ha a legkisebb konnektivitású pont megkeresésénél először az előző lépésben eltávolított pont szomszédait vizsgáljuk, és lehetőség szerint ezeket választjuk ki. Ekkor ugyanis például egy szubsztituált benzolgyűrű esetében először az oldalláncok tűnnek el, és a gráf a hattagú körre egyszerűsödik. A 11. ábra bemutatja az algoritmust egy egyszerű példán, a sztirolon keresztül.

A nyilak itt természetesen nem reakciót jelentenek, hanem egy-egy lépést az algoritmusban. Az első lépés a kettőskötések eltávolítása, a már ismertetett okból. A további lépésekben eltávolított pont a nyíl felett látható piros színnel, a címkével rendelkező él mellett pedig kékkel szerepel a címke tartalma. A legutolsó lépésben már csak a 8. pont marad, amin keletkezett egy hurokél. A hurokél címkéjéhez hozzávéve a pontot, amihez kapcsolódik, a címke tartalma 3, 4, 5, 6, 7, 8 lesz, ami megfelel a gráfban jelen lévő körnek.



11. ábra. Az összeomló P-gráf algortimus működése a sztirol példáján keresztül.

Ha a kiválasztott pont konnektivitása csak egy, akkor eltávolításához elég kitörölni a ponthoz tartozó éleket a *P* mátrixból. Ha ennél több, akkor a pont szomszédjait meg kell keresni és meg kell vizsgálni, hogy mely szomszédpárok között kell új élet létrehozni. A maximálisan létrehozandó élek száma egyenlő az ismétlés nélküli kombinációval, azaz két szomszédos pontnál egy, három szomszédos pontnál három, négynél hat, és így tovább. A 12. ábra az adamantán váz példáján keresztül mutatja be azt az esetet, amelynél nem kell az összes élt létrehozni a címkék tartalmának átfedése miatt. Vegyük észre, hogy azért van értelme négynél több szomszédos pontról beszélni, mert a *P*-gráfban az élek már nem a kémiai kötéseknek felelnek meg, így nem köti őket semmilyen vegyértékszabály.



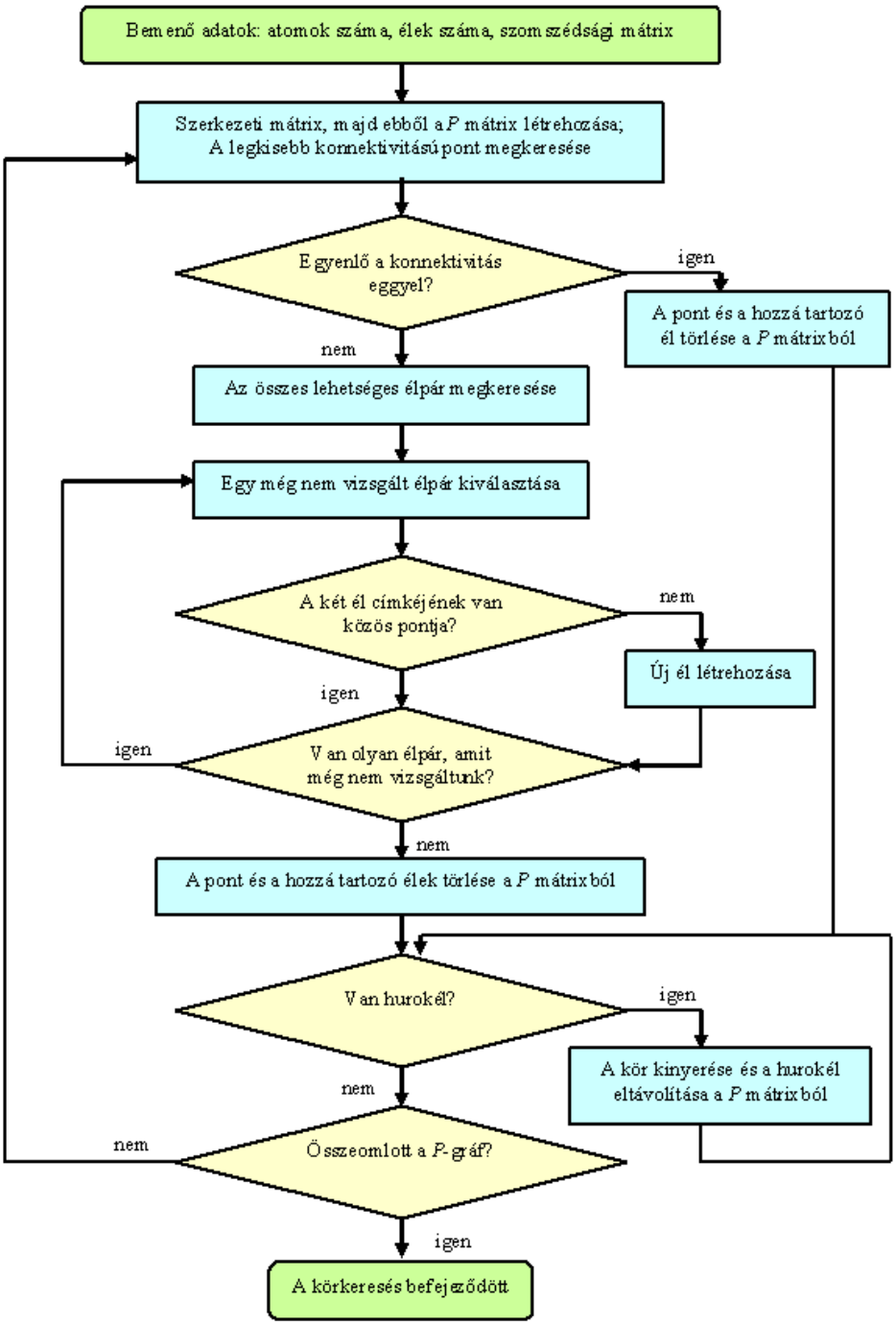
12. ábra. A gyűrűkereső algortimus lépései az adamantán váz esetében. (A jelölések a 11. ábráéval megegyeznek, zölddel a hurokélként megjelent és a lépésben eltávolított körök tagjai vannak feltüntetve a lépést jelző nyíl alatt.)

Az új élet csak abban az esetben szabad létrehozni, ha azoknak az éleknek a címkéjében, amiket helyettesít, nincs közös pont. Ebben az esetben a *P* mátrixban mindkét szomszédnál

létre kell hozni egy új élet, aminek a címke részébe kell másolni mindkét, a pontot és a szomszédot összekötő él címkéjét, az eredeti éleket pedig törölni kell.

Egy pont eltávolítása után ellenőrizni kell, hogy keletkezett-e hurokél (vagyis a *P* mátrixban van-e olyan sor, amiben az első és a második oszlopban lévő szám megegyezik). Ha igen, akkor a hozzá tartozó kört el kell tárolni egy gyűrűmátrixban, a hurokélt pedig ki kell venni a *P* mátrixból. Ha több hurokél is keletkezett, akkor ezt mindegyikkel meg kell tenni.

A körkereső algoritmus blokkdiagrammja a 13. ábrán látható.



13. ábra. A körkereső algoritmus blokkdiagrammja.

4.2. A meghatározott gyűrűk osztályozása és felismerése

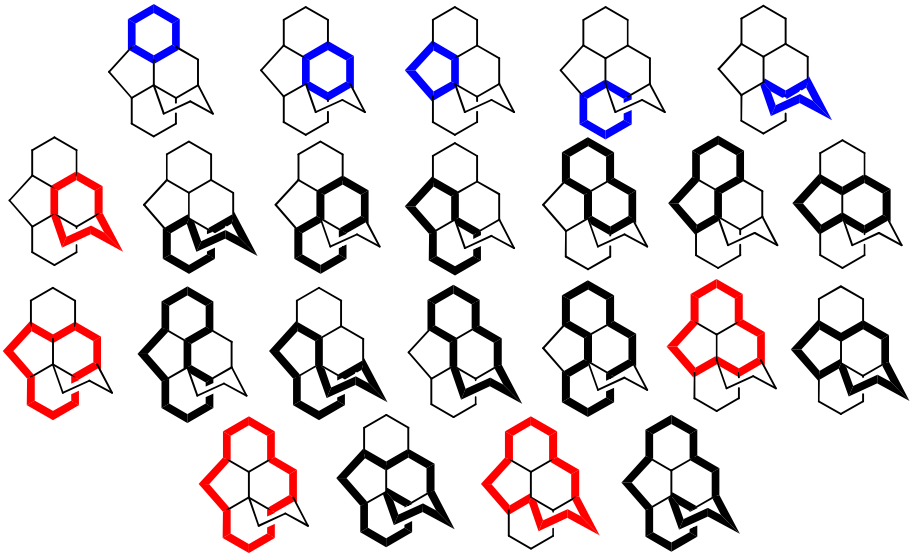
Mint azt az antracén és a norbornán (9. és 10. ábra) példáján láttuk, a matematikai és a kémiai szemlélet nem egyezik meg a molekuláris gráfban lévő gyűrűket tekintve, ezért fontos, hogy a meghatározott köröket valamilyen szisztéma szerint kategóriákba soroljuk, amely kategóriák segítenek a kémiai gyűrűrendszer felismerésében.

Ha több gyűrű is van egy szerkezetben, akkor azokat az egymáshoz viszonyított helyzetük alapján négy osztályba lehet sorolni:

- izolált gyűrűk: nincs közös atom a két gyűrűben;
- spirogyűrűk: egy közös atom van a két gyűrűben;
- kondenzált gyűrűk: két közös atom van a két gyűrűben;
- áthidalt gyűrűk: kettőnél több közös atom van a két gyűrűben.

Ezek alapján először azt kell megállapítani, hogy a szerkezetben hány izolált gyűrű (vagy gyűrűcsoport) van. Ehhez egy olyan *C* gyűrűkonnektivitási mátrixot definiáltam, ami $n * n$-es, ha a molekulában *n* darab gyűrű van, és a c_{ij} eleme megadja, hogy az *i*-edik és a *j*-edik gyűrűt hány atom köti össze. Ha ebben a mátrixban kiindulunk az egyik gyűrűről és megnézzük, hogy mely másik gyűrűkkel van kapcsolatban, majd az ezekkel a gyűrűkkel kapcsolatban lévő gyűrűket is megkeressük, és így tovább, akkor el tudjuk különíteni a molekulában izoláltan elhelyezkedő gyűrűket (vagy csoportjaikat).

A következő lépésben az egyes izolált gyűrűcsoportokat kell elemezni. A 14. ábrán látható a morfin alapvázának példáján, hogy milyen sok gyűrűt lehet egy szerkezetben találni, és hogy ezek nagy része nem játszik szerepet a gyűrűrendszer meghatározásában. Az elemzéshez a gyűrűket két nagy kategóriára osztottam: az egyszerű gyűrűkre és a komplex gyűrűkre. Az egyszerű gyűrűk a gráfelméleti definíciónak megfelelő egyszerű gyűrűk, azaz pontjaikat nem köti össze olyan él, ami nincs benne a gyűrűben, komplex gyűrűk esetében viszont van ilyen él. Ezen meggondolás szerint a komplex gyűrűk biztosan két kisebb gyűrűből állnak össze. Már a morfin alapvázának példáján is látszik, hogy az egyszerű gyűrűk kategóriáját tovább kell finomítani, mert az öt alapvető gyűrűn (a 14. ábra legfelső sorában láthatók kék színnel) kívül az pirossal jelzett gyűrűk is egyszerű gyűrűnek számítanak, pedig ezek csak három vagy négy másik gyűrű együtteséből állnak össze. Az ábrán a feketével jelölt gyűrűk a komplex gyűrűk.



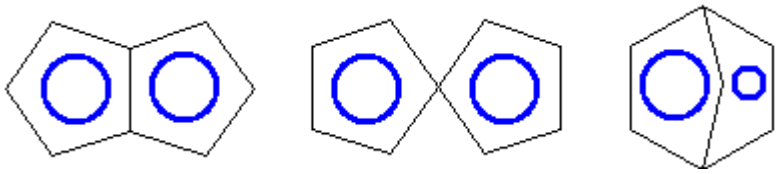
14. ábra. A morfin alapvázában jelen lévő körök (az egyszerűség kedvéért a morfinvázban egyébként jelen lévő heteroatomokat itt nem tüntettem fel).

Ennek a problémának a megoldására az egyszerű gyűrűket újabb két részre osztottam: valódi egyszerű gyűrűkre és pszeudo-egyszerű gyűrűkre, J. Corey *et al.* [34] munkájában találhatóhoz hasonló módon. A két részre osztáshoz az egyszerű gyűrűket a hosszuk szerint növekvő sorrendben kell vizsgálni. Az első gyűrű (vagyis a legrövidebb) mindenképpen valódi gyűrű lesz, az alkotó éleit pedig felvesszük egy olyan listába, ami a gyűrűrendszert leíró éleket tartalmazza. A következő gyűrűknél megvizsgáljuk, hogy az éleik között van-e olyan él, ami még nem tagja ennek a listának: ha van, akkor valódi egyszerű gyűrűről van szó, és az új éleket szintén felvesszük a listára, ha nem, akkor pszeudo-egyszerű gyűrűről beszélünk. Ezzel a módszerrel a 14. ábrán látható kék gyűrűk valódi egyszerű gyűrűk lesznek, a pirosak pszeudo-egyszerű gyűrűk. A valódi egyszerű gyűrűkön kívül a pszeudo-egyszerű gyűrűknek lesz jelentősége a gyűrűrendszer meghatározásában, mivel az áthidalt szerkezeteknél a főgyűrű a leghosszabb, ezért ez pszeudo-gyűrű lehet (lásd a 10. ábrán a norbornán esetét).

A gyűrűrendszert a valódi egyszerű gyűrűk egymáshoz viszonyított helyzete határozza meg. Ha csak egy valódi egyszerű gyűrű van, akkor monociklus a vegyület (vagy legalábbis egy izolált monociklust tartalmaz). Néhány előre leírt monociklust képes az algoritmus felismerni, ezek egy külön fájlban találhatók. A felismerés miként jének ismertetésére később kerül sor.

Ha kettő valódi egyszerű gyűrű található egy izolált csoporton belül, akkor a gyűrűrendszer lehet spiro, kondenzált vagy áthidalt. Ezeket a szerkezeteket a *C* gyűrűkonnektivitási mátrix alapján lehet megkülönböztetni, ahogy az a 15. ábrán látszik két öttagú gyűrűt tartalmazó

szerkezetek esetére (a *C* mátrix diagonális elemei a praktikus kezelés miatt –100-as értékre vannak állítva).



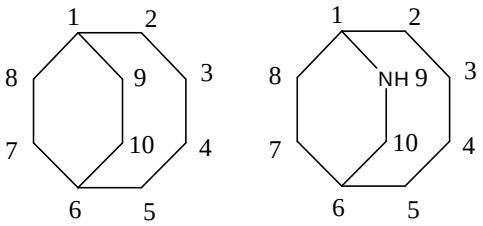
(a) ***(b)*** ***(c)***

$$C_a = \begin{pmatrix} -100 & 2 \\ 2 & -100 \end{pmatrix};\ C_b = \begin{pmatrix} -100 & 1 \\ 1 & -100 \end{pmatrix};\ C_c = \begin{pmatrix} -100 & 3 \\ 3 & -100 \end{pmatrix}.$$

15. ábra. Két valódi gyűrűt tartalmazó szerkezetek és gyűrűkonnektivitási mátrixaik. (A kék körök a két valódi gyűrűt jelölik, nem az aromaticitást.)

A kondenzált és spiro-rendszer esetében tehát a gyűrűszerkezet a gyűrűkonnektivitási mátrix ismeretében megadható, azaz outputként kiírható, hogy a molekula két adott hosszúságú gyűrűt tartalmaz, egymással spiro- vagy kondenzált kötésen keresztül kapcsolódva. A spiro-esetben érdemes még a két gyűrűre külön-külön rákeresni a monociklusokat tartalmazó fájlban. Ezen felül, ha a gyűrűrendszer valamelyik atomja heteroatom, a leíráshoz ennek a megadása is fontos, ezért a gyűrűalkotó atomokat végig kell nézni, hogy vannak-e köztük heteroatomok, és ha vannak, akkor meg kell adni a minőségüket és a gyűrűrendszerben elfoglalt helyzetüket. Ha a heteroatom a közös atomok egyike, akkor hídfőatomról beszélünk, ha nem, akkor azt a kört kell megadni, amelyikben megtalálható.

Áthidalt szerkezet esetében a főgyűrű a legtöbb esetben pszeudo-egyszerű gyűrűként jelenik meg, a gyűrűrendszerben ugyanis ekkor matematikai értelemben három gyűrű található, amik közül a leghosszabb a kémiai értelemben vett főgyűrű (például a norbornán esetében a főgyűrű a 10. ábrán látható pszeudo-egyszerű gyűrű, a két valódi egyszerű gyűrű pedig a 15. ábrán van jelölve). A gyűrűszerkezet leírásához tehát a pszeudo-gyűrűként nyilvántartott gyűrűt, pontosabban annak a hosszát kell megadni, ezen kívül pedig a hidat alkotó atomok számát, amit úgy kaphatunk meg, hogy a *C* mátrix offdiagonális eleméből kivonunk kettőt, ami a két hídfőatomot jelképezi.



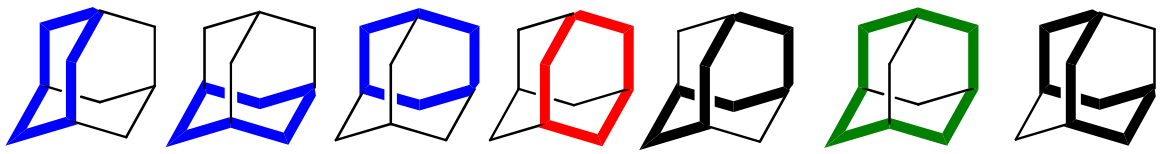
16. ábra. Biciklo[4.2.2]dekán (balra) és 7-azabiciklo[4.2.2]dekán (jobbra).

Ha a két hosszabb gyűrű egyenlő hosszúságú (például a biciklo[4.2.2]dekán esetében a 16. ábrán az 1-2-3-4-5-6-7-8 és az 1-2-3-4-5-6-9-10 gyűrűk), akkor ezek a gyűrűk ekvivalensek mind gráfelméleti, mind kémiai szempontból. Ilyen esetben az atomok sorszámozása dönti el, hogy melyik gyűrű lesz

valódi és melyik lesz pszeudo-gyűrű a feldolgozás során a két ekvivalens gyűrű közül. Ezzel szemben, ha van olyan heteroatom, amit az egyik gyűrű tartalmaz, a másik gyűrű viszont nem (mint például a 7-azabiciklo[4.2.2]dekán esetében a 16. ábrán), akkor a két gyűrű kémiai szempontból már nem lesz ekvivalens. Mivel a IUPAC ajánlása a gyűrűs molekulák számozására az, hogy a heteroatomok a lehető legkisebb számot kapják, ezért annak a gyűrűnek kell a főgyűrűnek lennie, ami a (több heteroatom esetén a legtöbb) heteroatomot tartalmazza. A heteroatomszámot csak olyan esetekben kell figyelembe venni, ahol a hosszúság nem dönti el egyértelműen, hogy melyik kör lesz a főgyűrű. Ha a főgyűrűt már meghatároztuk, a heteroatomok lehetnek tagjai a főgyűrűnek; lehetnek hídfő atomok; valamint lehetnek a híd részei.

Kettőnél több valódi egyszerű gyűrű jelenléte esetén a gyűrűkonnektivitási mátrixot kell megvizsgálni abból a szempontból, hogy hány darab 1, 2, illetve 2-nél nagyobb értékű eleme van. Ha az előbbi három értékből csak egy fordul elő, akkor a szerkezet egyértelműen egymással csak spiro- / csak kondenzált / csak áthidalt kapcsolatú gyűrűket tartalmaz. A csak spiro- és a csak kondenzált gyűrűt tartalmazó szerkezeteknél a két valódi egyszerű gyűrűt tartalmazó szerkezetekhez hasonlóan elég a gyűrűk számát és hosszát megadni a jellemzéshez, emellett pedig az esetleges heteroatomokat és azok helyzetét. Spirovegyületeknél az egyes gyűrűket itt is érdemes ellenőrizni a monociklusokat tartalmazó fájlban.

A többtagú tisztán áthidalt gyűrűszerkezetre jó példa az adamantánváz. A lehetséges gyűrűket a 17. ábra mutatja be. Ezeknél a rendszereknél is meg kell keresni első körben a főgyűrűt, majd definiálni a benne lévő hidak számát és hosszát.



17. ábra. Példa többtagú áthidalt gyűrűs szerkezetre: az adamantánváz.

A 17. ábrán a hattagú gyűrűk közül a kék színnel kiemeltek lesznek a valódi egyszerű gyűrűk, a piros színű pedig biztosan pszeudo-egyszerű gyűrű lesz. A nyolctagú gyűrűk szintén pszeudo-egyszerű gyűrűk lesznek az algoritmus szerint, de valójában az egyik az adamantán váz főgyűrűje (szimmetriaokokból bármelyik lehet, de az ábrán a középső gyűrű van zölddel kiemelve példaként). Általános esetben is igaz, hogy ahhoz, hogy a főgyűrűt megtaláljuk, a pszeudo-egyszerű gyűrűk között kell megkeresni a leghosszabbakat. Ha a valós egyszerű gyűrűk között is van ilyen hosszúságú, akkor azt a pszeudo-gyűrűkkel együtt kell vizsgálni.

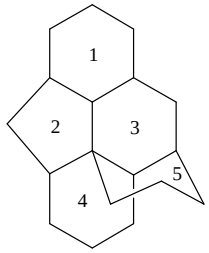
Az ugyanúgy maximális hosszúságú gyűrűk közül a bennük lévő heteroatomok mennyisége alapján lehet kiválasztani a főgyűrűt.

Ha a főgyűrű nem valódi egyszerű gyűrű, akkor az egyszerű gyűrűk felosztását újra el kell végezni úgy, hogy a főgyűrű mindenképpen valódi egyszerű gyűrű legyen, praktikusan a gyűrűk vizsgálatának sorrendjét úgy kell megváltoztatni, hogy az első gyűrű a főgyűrű legyen, a többi gyűrű pedig növekvő hosszúsági sorrendben kövesse. Erre azért van szükség, mivel a hidak hosszát a valódi gyűrűk és a főgyűrű átfedése alapján lehet megadni: a valódi gyűrű olyan pontjai lesznek a híd elemei, amik nem találhatók meg a főgyűrűben. Azzal viszont, hogy egy eddig pszeudo-gyűrűnek számító gyűrűt valódiként kezelünk, lehet, hogy eddig valódinak számító gyűrűnek már nincs olyan élhozzájárulása a gyűrűélekhez, ami miatt valódi lenne. Például ha a 17. ábrán zöld színnel feltütetett gyűrűt választjuk főgyűrűnek, akkor balról az első és a második valódi gyűrű és a főgyűrű különbsége megadja a két hidat, a harmadik valódi gyűrű és a főgyűrű különbsége viszont nem új híd, mert ezeket az atomokat a második valódi gyűrűnél már vizsgáltuk egyszer. Ha viszont újrakategorizáljuk az egyszerű gyűrűket, akkor a harmadik gyűrű már pszeudo-gyűrű lesz, és ezért a hidak meghatározásánál nem lesz figyelembe véve.

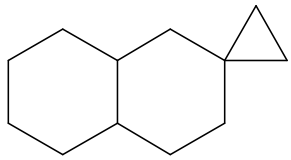
Ha a főgyűrű a valódi gyűrűk közül kerül ki, akkor az egyszerű gyűrűk újrakategorizálására nincs szükség. A szerkezet leírására ekkor is a főgyűrű hosszát, a hidak számát és hosszát, valamint a heteroatomokat és helyzetüket (főgyűrűben, hídban, vagy hídfőatomként) kell megadni.

Ha a gyűrűkonnektivitási mátrixban nemcsak egyféle érték szerepel, akkor komplexebb szerkezettel van dolgunk. Például a morfinváz három hattagú és egy öttagú gyűrű kondenzált rendszerének tekinthető, amiben egy háromtagú híd is szerepel (bár az alapváz a benzilizokinolin kondenzációjával képezhető). A gyűrűkonnektivitási mátrixa az egyszerű gyűrűkre (az alapvázban az 1-gyel jelölt gyűrű kapcsolatai az első sorban és oszlopban, a 2-vel jelölté a másodikban, stb. találhatók):

$$C_{morfinváz} = \begin{pmatrix} -100 & 2 & 2 & 0 & 0 \\ 2 & -100 & 2 & 2 & 1 \\ 2 & 2 & -100 & 2 & 3 \\ 0 & 2 & 2 & -100 & 2 \\ 0 & 1 & 3 & 2 & -100 \end{pmatrix}$$



A gyűrűkonnektivitási mátrixból látszik, hogy nemcsak egy értéket vesznek fel az offdiagonális elemek, illetve, hogy bár van 1 értékű elem, a vegyületre nem mondhatjuk, hogy spiroszármazék. Ebből levonható az a következtetés, hogy az 1 értékű elem szükséges, de nem elégséges feltétele annak, hogy spiro-helyzetű gyűrűről legyen szó. Ezzel szemben a 2 és a 3 értékű elem egyértelműen kondenzált gyűrű, illetve híd jelenlétére utal, bár a hidat jelző elem annál a gyűrűnél is megjelenik, ami maga nem híd, de hozzá kapcsolódik híd. (A példamátrixban a hidat egy 3 értékű elem jelzi, de általános esetben ez természetesen nagyobb értéket is felvehet).



18. ábra. Példa komplexebb szerkezetre.

Ahhoz, hogy egy ilyen, vagy még komplexebb szerkezetről információt tudjak adni, először több, egymással spiro-helyzetű komponensre bontottam a gyűrűrendszert, az izolált gyűrűrendszerek kereséséhez hasonló módon. Például a 18. ábrán látható szerkezet két, egymással spiro-helyzetben lévő gyűrűrendszert tartalmaz, az egyik két hattagú gyűrű által alkotott kondenzált gyűrű, a másik pedig egy háromtagú gyűrű. Ez a tárgyalásmód azért praktikus, mert egyrészt az egymástól majdnem független gyűrűket külön lehet elemezni, másrészt a kémiai nevezéktannal is összhangban van, mivel a 18. ábrán lévő vegyület egy perhidro-spiro[ciklopropán-naftalin].

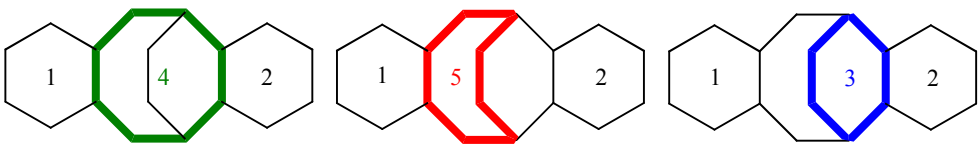
A további vizsgálathoz az egyes valódi gyűrűk adott spiro-komponensben betöltött „szerepét” kell meghatározni. Ez jelentheti azt, hogy az adott gyűrű egy monociklus, vagy egy kondenzált rendszer taggyűrűje, vagy egy híd rendelhető hozzá a molekulán belül. Ha az adott gyűrűhöz tartozó sor offdiagonális elemei a gyűrűkonnektivitási mátrixban csak egyféle értéket vesznek fel nullán kívül, akkor a gyűrű egyértelműen spiro-/kondenzált/áthidalt szerepet tölt be. A morfinváz esetében ilyen az 1. és a 4. gyűrű: mindkettő offdiagonális elemei 2-es értéket vesznek fel, így ezek mindenképpen egy kondenzált rendszer részei. Ha a sorban csak 1 és 2 szerepel, akkor az adott gyűrű szintén kondenzált rendszer része lesz (például a 2. gyűrű), ha csak 1 és 3, akkor a gyűrű egyértelműen áthidalt szerepet tölt be, mivel, mint ahogy a morfin példáján is látszik, az egy atomos átfedés nem jelenti feltétlenül, hogy különálló spirogyűrűkről van szó. Ezek alapján csak az olyan gyűrűk funkciója eldöntetlen, amikhez tartozó sorokban 2-es és 3-as értékű elemek is vannak. Ilyen gyűrűk a morfin esetében a 3. és az 5. gyűrűk: ezek egymással kapcsolódnak több, mint két atomon keresztül, a többi gyűrűhöz pedig két, vagy annál kevesebb atomon keresztül, ezért ezek közül kell eldönteni a mátrix alapján, hogy melyik tartalmazza a hidat, és melyik része a kondenzált rendszernek.

Ha nincs egy olyan gyűrű sem, aminek a fentiekkel meg lehetne határozni a funkcióját, akkor ki kell választani azt a gyűrűt, ami a legtöbb másik gyűrűvel van kondenzált viszonyban, és ezt tekinteni kondenzáltnak, mivel ebből kiindulva lehet a teljes kondenzált rendszert felderíteni.

A 2 és 3 értékekkel jellemezhető gyűrűk funkciójának eldöntéséhez, ha van, akkor az egyértelműen hidat reprezentáló gyűrűkkel való kapcsolatukat figyelmen kívül kell hagyni. Ha ekkor nincs 2-nél nagyobb érték a gyűrűkonnektivitási mátrix megfelelő sorában, akkor a vizsgált gyűrű kondenzált. Ha nincs hidat reprezentáló gyűrű, ami alapján ezt meg lehetne tenni (például a morfinváz esetén), akkor a kérdéses funkciójú gyűrűk között sorrendet kell felállítani, aszerint, hogy melyik milyen valószínűséggel reprezentál hidat. Ennél a sorrendnél a legfontosabb szempont az, hogy minél kevesebb biztosan kondenzált gyűrűvel legyen 2 atomja közös, mivel ha sok kondenzált gyűrűvel van kondenzált kötése, akkor valószínűbb, hogy a kondenzált rendszer része. Ha több kérdéses gyűrű is ugyanannyi kondenzált gyűrűvel van kapcsolatban, akkor a következő sorrendet befolyásoló tényező a gyűrű hosszúsága. A rövidebb gyűrűk kedvezményezettek, mivel az áthidalt szerkezeteknél a leghosszabb gyűrű a főgyűrű. Ha a gyűrűk hossza is azonos, akkor a gyűrűben lévő heteroatomok száma dönt: ennél is a kisebb szám tartozik a hidat valószínűbben reprezentáló gyűrűhöz. A morfin példájában a 3. gyűrű a kondenzált gyűrűk (1., 2. és 4.) mindegyikéhez két közös atommal kapcsolódik, az 5. gyűrű pedig csak a 4. gyűrűhöz, így a 3. gyűrű lesz a kondenzált rendszer része, az 5. gyűrű pedig a hidat írja le.

Ezt követően a sorrend szerint legvalószínűbb gyűrűt hidat reprezentáló gyűrűnek tekintjük. Ha ennek megfelelően a többi kérdéses funckiójú gyűrű esetében az ezzel a gyűrűvel való kapcsolatot nem vesszük figyelembe, és még így is vannak kérdéses funkciójú gyűrűk, akkor a következő legvalószínűbb gyűrűt tekintjük hidat reprezentáló gyűrűnek, és így tovább, egészen addig, amíg nem marad ismeretlen funkciójú gyűrű. Ekkor az adott izolált gyűrű adott spirokomponensében csak kondenzált vagy híd szerepet betöltő gyűrűk vannak jelen.

Ha a híd a kondenzált rendszernek csak egy gyűrűjéhez kapcsolódik (mint például a 19. ábrán), akkor ugyan egy biciklusos alegységet képeznek, de nem biztos, hogy az a gyűrű lesz a főgyűrű, ami az alegység önálló kezeléséből következne.



19. ábra. Csak egy kondenzált gyűrűhöz kapcsolódó híd esete.

Ha a 19. ábrán lévő szerkezetben a középső gyűrűrendszert vizsgálnánk önmagában, azt találnánk, hogy a zöld és a piros színnel jelölt gyűrű ekvivalens, bármelyik lehet az áthidalt szerkezet főgyűrűje. Ebben az esetben viszont csak a zölddel jelölt gyűrű lehet a főgyűrű, mivel ha a pirossal jelölt lenne a főgyűrű, akkor megszakadna a kondenzált gyűrűrendszer. Ráadásul, ha a fenti szerkezetben a híd több, mint két atomot tartalmazna, egyértelműen a kondenzált rendszert megtörő gyűrű lenne a főgyűrű, mivel az a leghosszabb. Ezért, ha egy ilyen komplex szerkezetben a híd csak egy kondenzált gyűrűhöz kapcsolódik, akkor további vizsgálatok szükségesek.

A gyűrűkonnektivitási mátrixok, ha a 4. gyűrű (ekkor a sorok/oszlopok rendre az 1., 2., 3., 4. gyűrűkhoz tartoznak), illetve, ha az 5. gyűrű a valódi egyszerű gyűrű (a sorok/oszlopok sorrendje ekkor: 1., 2., 3., 5.):

$$C_4 = \begin{pmatrix} -100 & 0 & 0 & 2 \\ 0 & -100 & 2 & 2 \\ 0 & 2 & -100 & 4 \\ 2 & 2 & 4 & -100 \end{pmatrix}; \; C_5 = \begin{pmatrix} -100 & 0 & 0 & 2 \\ 0 & -100 & 2 & 0 \\ 0 & 2 & -100 & 4 \\ 2 & 0 & 4 & -100 \end{pmatrix}.$$

A két mátrixot összehasonlítva látszik, hogy a $c_{4,2}$ elemben van különbség: ez a két gyűrű a C_4 mátrix szerint egymással kondenzált kapcsolatban van, a C_5 mátrix alapján viszont nem. Általános esetben az összes olyan hídnál, ami csak egy gyűrűhöz kapcsolódik a kondenzált rendszerben, ellenőrizni kell, hogy a megfelelő gyűrű lett-e kiválasztva a kondenzált rendszer részének. Ehhez meg kell keresni azt a pszeudo-gyűrűt, ami az adott áthidalt szerkezethez tartozik, és megvizsgálni, hogy ha ez lenne valódi gyűrű, akkor a gyűrűkonnektivitási mátrix hogy változna. Ha a gyűrűkonnektivitási mátrixban nem veszik el információ, azaz a nemnulla elemek nem válnak nullává, illetve nem is csökkennek, egyes nulla értékű elemek viszont nagyobb értékűek lesznek, akkor a pszeudo-gyűrű jobb választás a meglévőnél, különben nem. A két mátrix alapján jól látszik, hogy az 4. gyűrűt kell a főgyűrűnek tekinteni.

Ezek után a komplex gyűrűrendszer jellemzéséhez meg kell adni a benne szereplő spirokomponensek leírását. Az egyes spirokomponenseket jellemzi az őket alkotó monociklus, vagy ha van, a kondenzált gyűrűrendszer, azok tagjainak hosszúsága, a hidak mennyisége és hosszúsága (amit a hidat reprezentáló gyűrűk olyan atomjainak a száma adja meg, amik más valódi gyűrűben nem szerepelnek), és a heteroatomok a gyűrűkben.

Ha a szerkezet valamilyen formában tartalmaz monociklust, akkor lehetőség van néhány ismertebb, hagyományos névvel rendelkező vegyülettípus azonosítására. Ezeknek a monociklusoknak a listája a Függelékben található. Ehhez az azonosításhoz az ismert szerkezetek egy külön fájlban vannak tárolva, a 20. ábra szerinti formában. Egy-egy monociklus tulajdonságainak leírása a „new” és az „end” karaktersort tartalmazó sorok között helyezkedik el. Itt az első sorban a gyűrűk száma található (monociklusoknál ez mindig 1), a második sorban a konkrét vegyület hagyományos neve foglal helyet, a harmadik sor pedig azt tartalmazza, hogy hány tagú a gyűrű. A negyedik sor a heteroatomok számát tünteti fel, ezután pedig annyi sorban, ahány heteroatom van, az egyes heteroatomok kémiai minősége szerepel, valamint az, hogy az előttük felsorolásra került egyéb heteroatomoktól rendre hány atom választja el őket. Az utolsó sorban a telítetlenségre vonatkozó információ található: mivel az összegyűjtött szerkezetek mindegyike vagy telített, vagy aromás, ezért ezt a kettő esetet kell csak megkülönböztetni. A –1-es érték az aromásságot, a 0-s érték a telítettséget jelenti.

```
new
        1
        pyrazole
        5
        heteroatom: 2
        N
        N 1
        unsaturation: -1
end
new
        1
        imidazole
        5
        heteroatom: 2
        N
        N 2
        unsaturation: -1
end
```

20. ábra. A monociklusokat tartalmazó fájl részlete.

A szerkezetben lévő monociklust az összes, fájlban található monociklussal össze kell hasonlítani. Legelőször azt lehet ellenőrizni, hogy a talált gyűrű és a fájlban leírt egyenlő hosszaságú-e. Ha igen, akkor a következő lépésben a benne található heteroatomok számának kell megegyeznie. Ezen belül meg kell egyeznie a különböző kémiai minőségű heteroatomok számának külön-külön is, valamint az egymáshoz viszonyított helyzetüknek is egyeznie kell. Ezt a fájlban a heteroatomokat leíró tömb tartalmazza, aminek sorai és oszlopai az egyes heteroatomoknak fellenek meg. A szerkezetben lévő monociklusban az egymáshoz viszonyított helyzetet a heteroatomokról indított szélességi kereséssel lehet meghatározni, és a meghatározott távolságokat egy mátrixban tárolni. Ha a két mátrix megegyezik, akkor a heteroatomok helyzete a két gyűrűben ugyanolyan. Ha kettőnél több heteroatom van, akkor viszont a mátrix sorainak és oszlopainak sorrendje függ attól, hogy milyen sorrendben vizsgáltuk az egyes heteroatomokat, ezért az összes lehetséges sorrendnek megfelelő mátrixot meg kell vizsgálni. Ez tulajdonképpen megfelel a gráfizomorfizmus problémájának, ami ugyan sok időt igényel, de figyelembe véve, hogy a heteroatomok száma a listában szereplő monociklusokban maximum négy, ezért még megengedhető. A következő lépésben a gyűrű telítetlenségére vonatkozó információkat kell

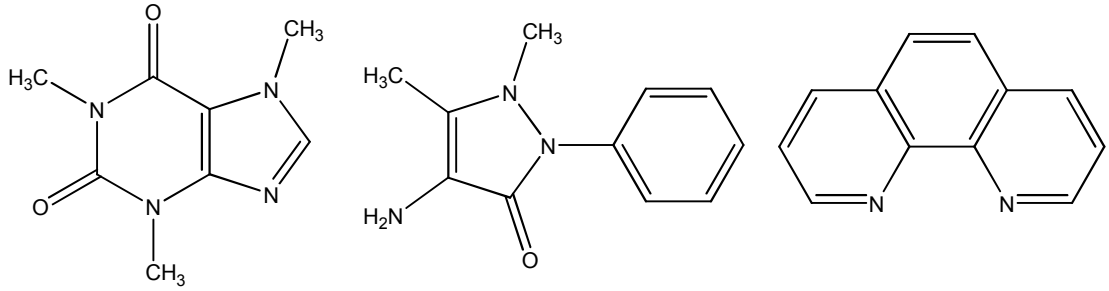
ellenőrizni. A telítettség feltétele, hogy a gyűrű tagjai között csak egyszeres kötések lehetnek, az aromaticitást pedig a szénatomok sp-hibridizációs állapota alapján lehet meghatározni.

Ezzel a módszerrel lehet jellemezni az egyes generált szerkezetek gyűrűrendszereit. Az esetleges nagymennyiségű adat átláthatóságát segíti, ha egy összefoglalást adunk a szerkezetekben lévő gyűrűrendszerekről. Ebben feltüntettem, hogy összesen

- hány monociklus fordult elő a találatokban;
- hány, tisztán spirogyűrűket tartalmazó származék fordult elő a találatokban;
- hány, tisztán kondenzált gyűrűrendszereket tartalmazó származék fordult elő a találatokban;
- hány, tisztán áthidalt gyűrűket tartalmazó származék fordult elő a találatokban;
- hány olyan származék fordult elő a találatokban, amiben a kondenzált rendszerhez híd kapcsolódik;
- hány, a fentebb felsoroltaknál is komplexebb szerkezet fordult elő a találatokban;
- illetve, hogy hány vegyület tartalmazott egynél több, egymástól izolált gyűrűrendszert.

4.3. Tesztvegyületek mérése

Az algoritmus teszteléséhez kiválasztottam három olyan kismolekulás szerves vegyületet, aminek érdekes lehet az automatikus szerkezetanalízise. Ezek a vegyületek a koffein, a 4-amino-antipirin (szebályos nevén 4-amino-1,5-dimetil-2-fenil-pirazol-3-on) és az 1,10-fenantrolin voltak, a szerkezetüket a 21. ábra mutatja be.



21. ábra. Tesztvegyületek (balról jobbra): koffein, 4-amino-antipirin és 1,10-fenantrolin.

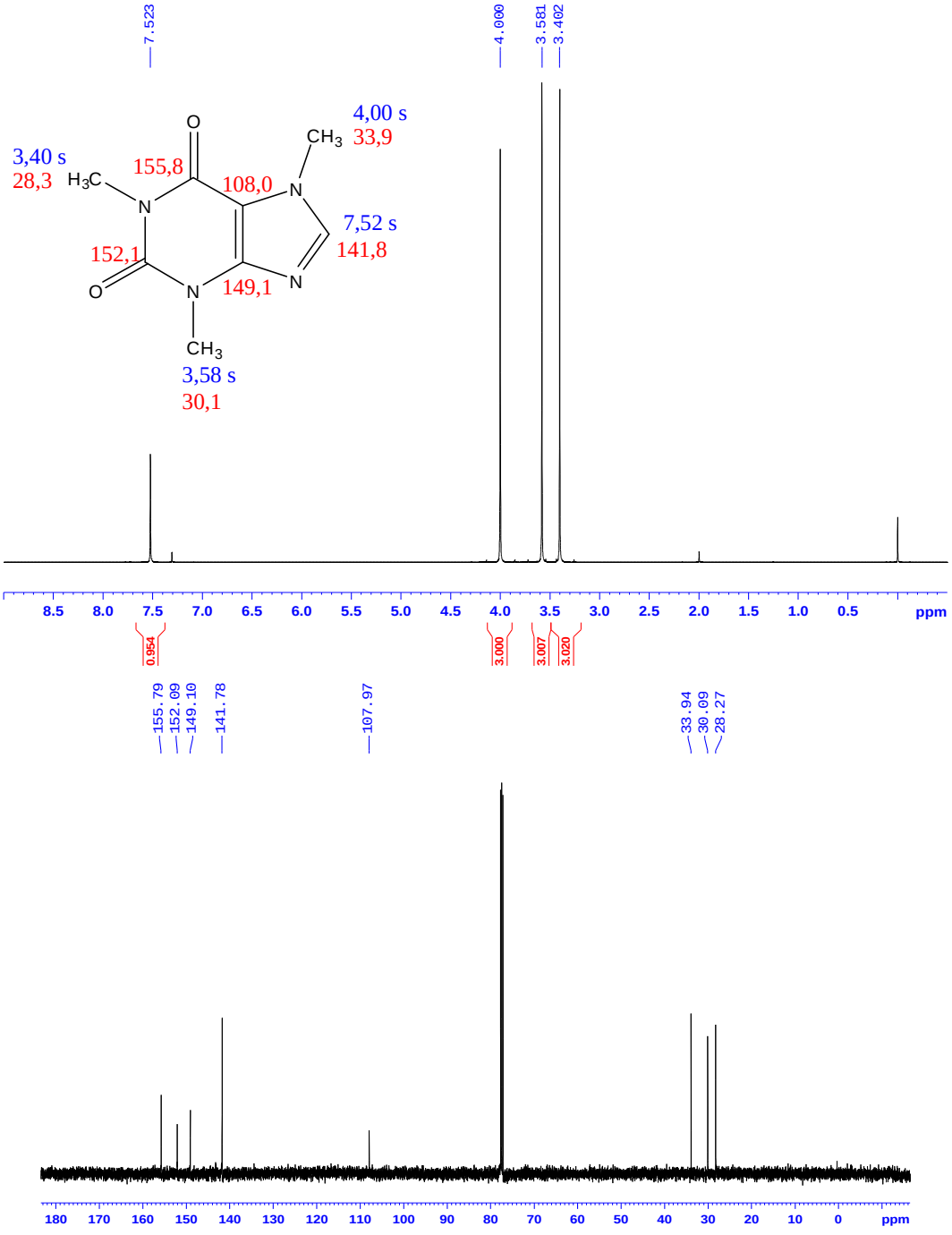
Mindhárom vegyületről a szükséges ^{1}H, HSQC és HMBC spektrumokon kívül felvettem ^{13}C spektrumot is. A spektrumok az ELTE TTK Kémiai Intézet Bruker DRX 500 MHz-es NMR spektrométerén készültek a 3. táblázatban részletezett paraméterekkel (5 mm-es 1H/13C/15N mérőfejjel, 300 K-en). Az oldószernek használt $CDCl_3$ 1% TMS-t is tartalmazott.

	koffein	4-amino-antipirin	1,10-fenantrolin
bemért mintamennyiség	30 mg	39 mg	15 mg
oldószer	600 µl $CDCl_3$	600 µl $CDCl_3$	600 µl $CDCl_3$
^{1}H spektrum	✓	✓	✓
pulzusprogram	zg30	zg30	zg30
d1	1 sec	1 sec	1 sec
scan-ek száma	16	16	16
spektrum szélessége	20 ppm	20 ppm	20 ppm
spektrum középpontja	6 ppm	6 ppm	6 ppm
^{13}C spektrum	✓	✓	✓
pulzusprogram	zgpg30	zgpg30	zgpg30
d1	2 sec	2 sec	2 sec
scan-ek száma	88	64	255
spektrum szélessége	240 ppm	240 ppm	240 ppm
spektrum középpontja	100 ppm	100 ppm	100 ppm
HSQC spektrum	✓	✓	✓
pulzusprogram	hsqcedetgpsp.3	hsqcedetgpsp.3	hsqcedetgpsp.3
scan-ek száma	2	1	1
^{1}H tengely szélessége	10 ppm	10 ppm	10 ppm
^{1}H tengely középpontja	5,5 ppm	5 ppm	5 ppm
^{13}C tengely szélessége	200 ppm	220 ppm	100 ppm
^{13}C tengely középpontja	95 ppm	105 ppm	140 ppm
HMBC spektrum	✓	✓	✓
pulzusprogram	hmbcetgpl3nd	hmbcetgpl3nd	hmbcetgpl3nd
scan-ek száma	8	4	4
^{1}H tengely szélessége	10 ppm	10 ppm	10 ppm
^{1}H tengely középpontja	5,5 ppm	5 ppm	5 ppm
^{13}C tengely szélessége	200 ppm	220 ppm	100 ppm
^{13}C tengely középpontja	95 ppm	105 ppm	140 ppm

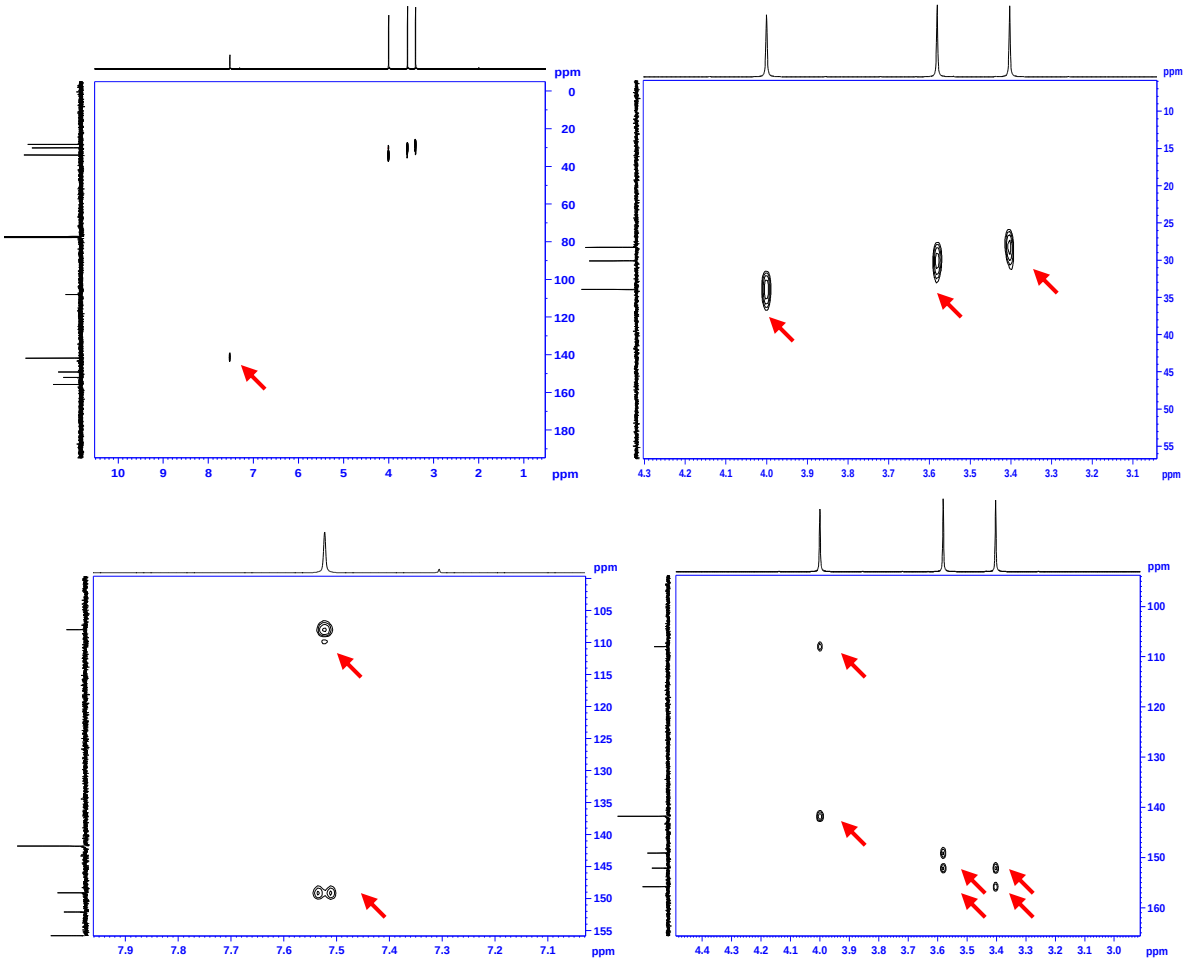
3. táblázat. A tesztvegyületek spektrumainak felvételénél használt főbb paraméterek.

4.3.1. A koffein mérése

A koffein proton- és szénspektruma a 22. ábrán látható, a molekula asszignációjával együtt (a kék számok a protonok eltolódásértékeit jelentik, a pirosak pedig a szenekét). A HSQC és a HMBC spektrumok fontos részei a 23. ábrán láthatók, rajtuk piros nyilakkal ki vannak emelve azok a keresztcsúcsok, amik felhasználásra kerültek a *Small Molecule Structure Elucidation* automatikus csúcsanalízisében (egy csúcsra csak egyszer mutat nyíl, akkor is, ha két spektrumon szerepel).



22. ábra. A koffein protonspektruma (felül) és szénspektruma (alul).

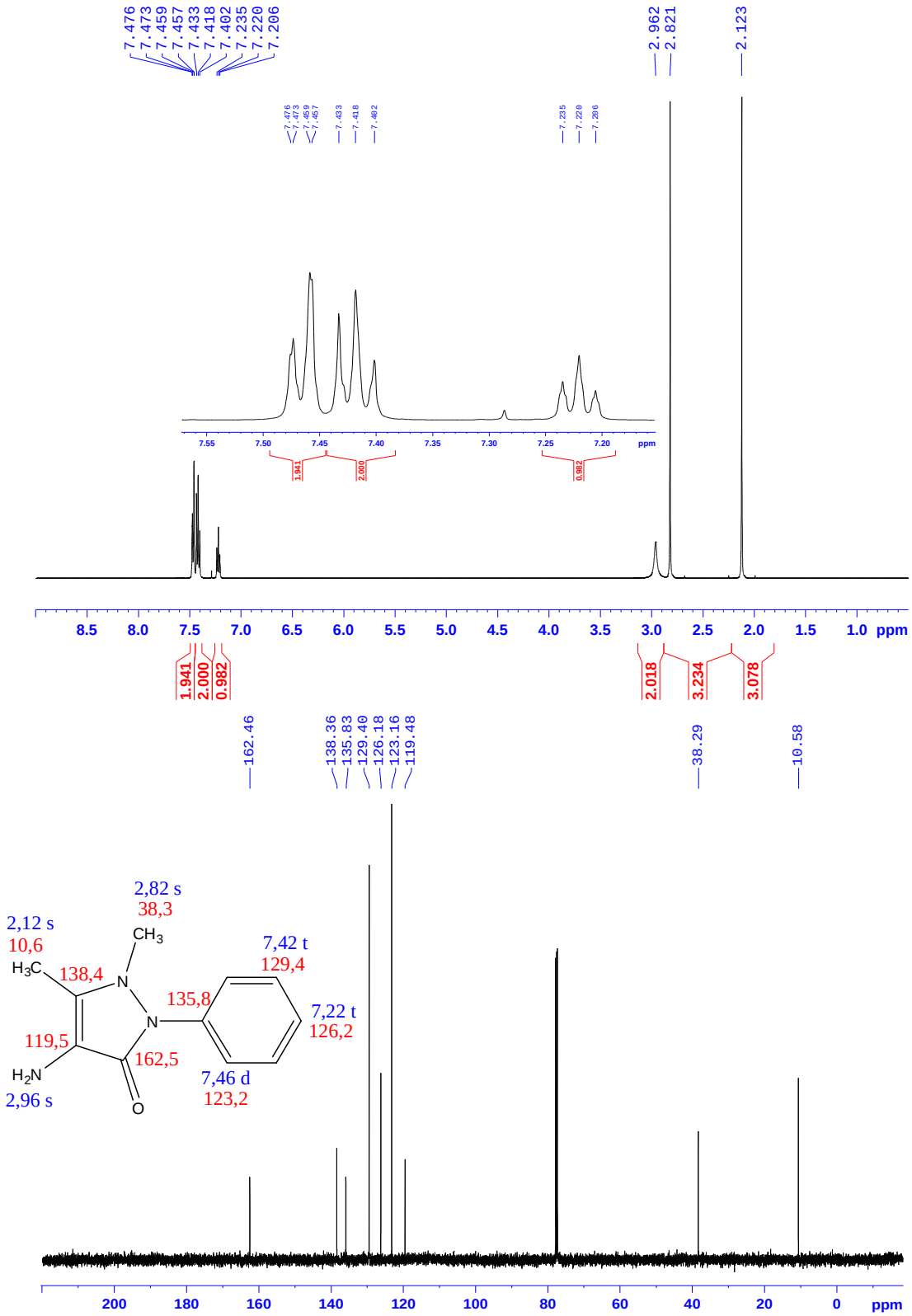


23. ábra. A HSQC spektrum (felül) és a HMBC spektrum (alul) részletei.

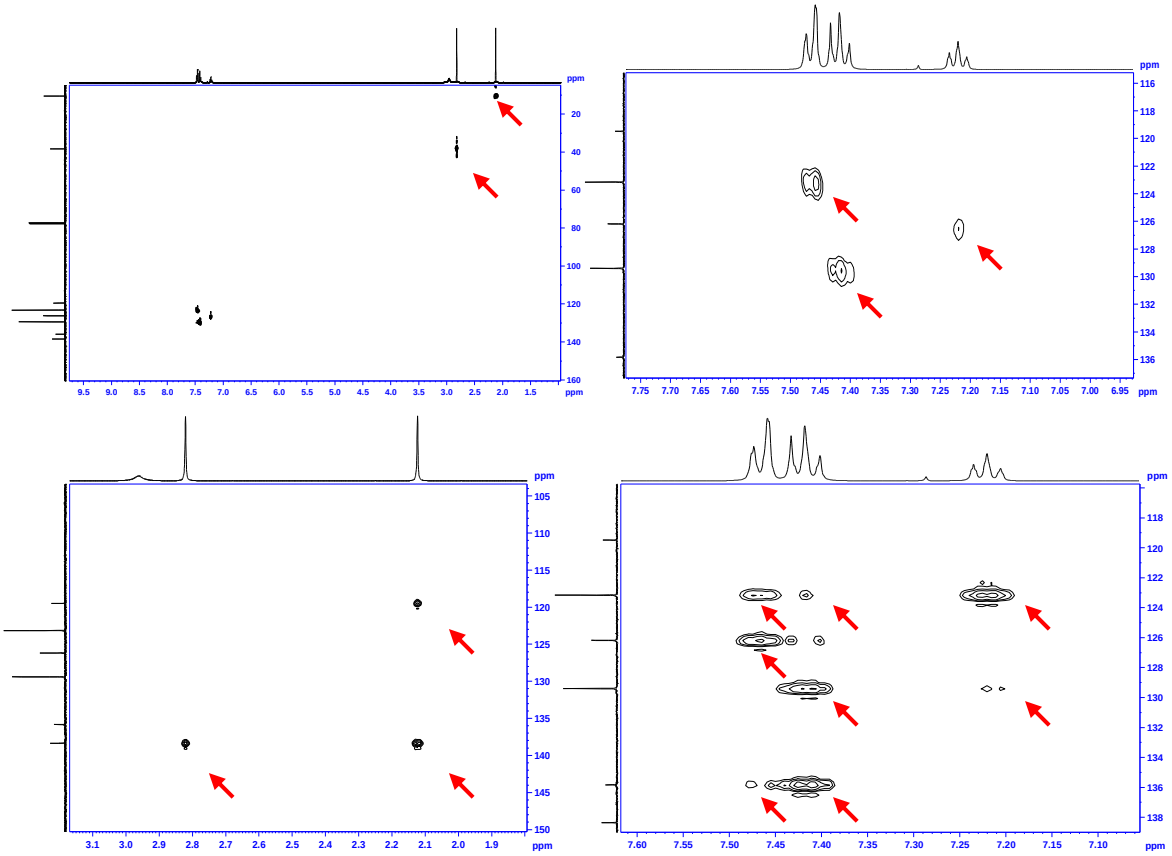
A a *Small Molecule Structure Elucidation* az automatikus analízis során figyelmeztető üzenetet küldött, miszerint alacsony a H/CNO arány a molekulában, és sok találatra lehet számítani a kevés számú keresztcsúcs miatt. A szerkezetgenerálásnál a Bruker saját generátorát használtam, és nem állítottam be semmilyen kényszerítő fragmenst, sem maximális gyűrűhosszt. Ezen kívül az első szerkezetenerálásnál meghagytam a lehetőségét annak, hogy a HMBC korrelációk közül az összes keresztcsúcs 10%-át (jelen esetben 1 korrelációt) megsérthessen a generált találat, a csúcsok eliminálásához pedig az „optimal” módot választottam. A szerkezetgenerálásra fordítható maximális időt 5 percre állítottam, amit az első próbálkozásnál teljes mértékben ki is töltött a program: 34568 találat adódott. Ezért második próbálkozásnál már nem engedtem, hogy bármely keresztcsúcs eliminálható legyen. Így az időigény 5 percen belül maradt, a generált szerkezetek száma pedig 4165-re csökkent. A generált szerkezetek között az ötvenedikként megjelent a koffein valódi szerkezete (a találatok sorrendje véletlenszerű, ha nem használunk valamilyen rangsoroló metódust).

4.3.2. A 4-amino-antipirin mérése

A 4-amino-antipirin proton- és szénspektruma a 24. ábrán látható, a HSQC és a HMBC spektrum pedig a 25. ábrán.



24. ábra. A 4-amino-antipirin proton- (felül) és szénspektruma (alul).



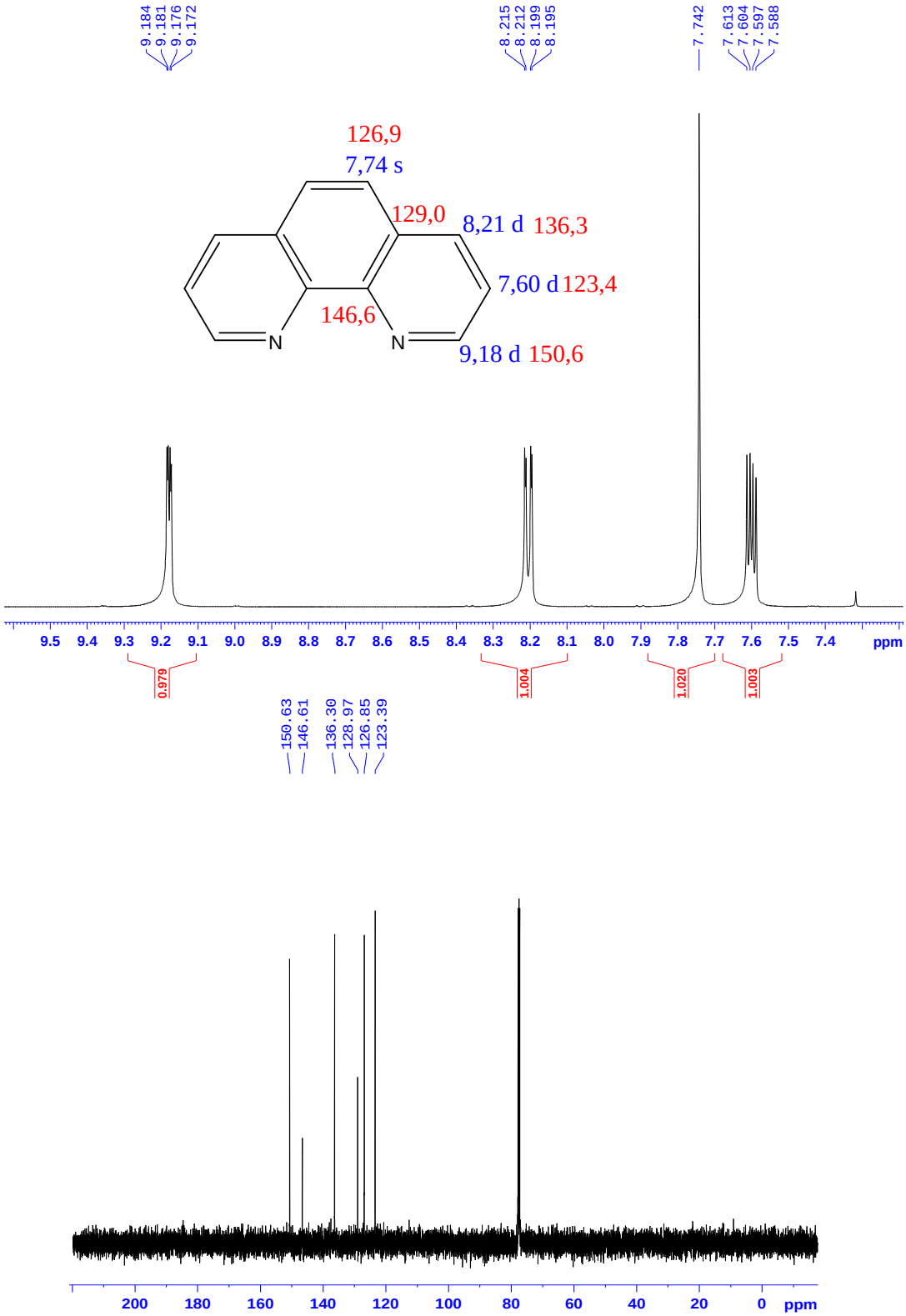
25. ábra. A 4-amino-antipirin HSQC (felül) és HMBC (alul) spektrumrészletei.

A 4-amino-antipirin esetében is alacsony a H/CNO arány. A szerkezetgenerálásnál ugyanazokat a beállításokat használtam, mint a koffein első próbálkozásánál: a Bruker szerkezetgeneráló algoritmusát, „optimal” módú keresztcsúcs eliminációs eljárást és semmilyen fragmens-kényszert. Ezekkel a feltételekkel 1313 találat generálódott, és a futás az időkorláton belül lezajlott. Ebben az esetben is generáltam úgy szerkezeteket, hogy a keresztcsúcs eliminációt nem engedélyeztem: ekkor az előző esettel megegyező számú, 1313 darab találatot kaptam, tehát ebben az esetben ez a kényszerfeltétel nem csökkenti a találatok számát. A 4-amino-antipirin tényleges szerkezete a 109. találatként generálódott.

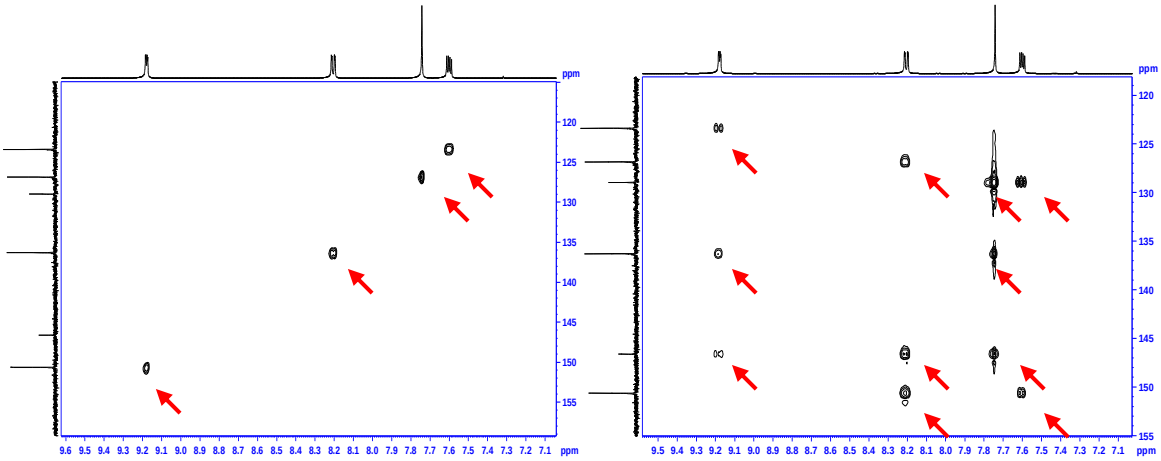
4.3.3. Az 1,10-fenantrolin mérése

Az 1,10-fenantrolin proton- és szénspektruma a 26. ábrán, HSQC és HMBC spektrumai pedig a 27. ábrán láthatók. Ennek a vegyületnek az esetében is alacsony a H/CNO arány, ami miatt most is sok találat keletkezésére kell számítani. Az automatikus keresztcsúcskeresésnél csak a hidrogén- és a szénatomok fele lett asszignálva, mivel a molekula szimmetrikus voltát nem ismerhette a program. Mégis, mivel a molekulatömeg ismert, lehet feltételezni, hogy az

egyes csúcsok megjelenéséért tulajdonképpen két atom a felelős a molekulában, ezért az automatikus asszignációt manuálisan korrigáltam.



26. ábra. Az 1,10-fenantrolin proton- (felül) és szénspektruma (alul).

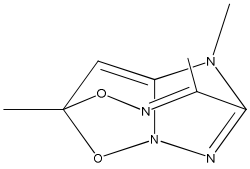
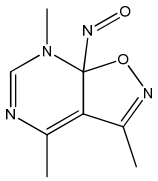
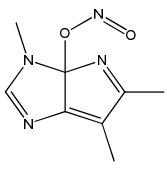


27. ábra. Az 1,10-fenantrolin HSQC (balra) és HMBC (jobbra) spektruma.

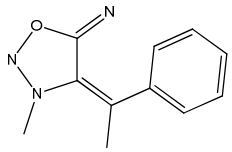
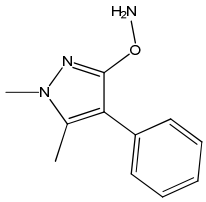
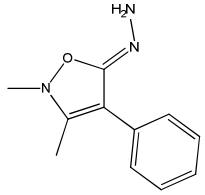
A szerkezetgeneráláshoz itt is a Bruker generáló szoftverét használtam, illetve az „optimal” módú keresztcsúcs eliminációs eljárást, aminek eredményeképpen 398 szerkezet generálódott, köztük elsőként a tényleges szerkezet.

Terjedelmi okokból csak az első 3 találat jellemzését és a találatok összegzését mutatom itt be. A találatok jellemzése a 4., 5. és a 6. táblázatban láthatók.

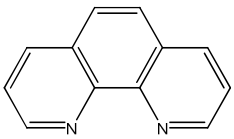
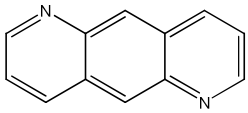
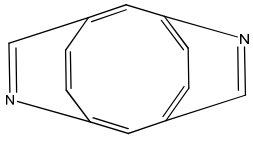
Már ezekből a találatokból is látszik, hogy van, amikor kevéssé valószínű, instabil szerkezetek is keletkeznek és van, amikor a kérdéses szerkezet már az első találattal megegyezik (például az 1,10-fenantrolin esetében). A konkrét találatokra adott jellemzés a táblázatok alapján helyes, az összhangban van a tényeges kémiai gyűrűrendszerrel.

Szerkezet	Output
1. találat: 	1. hit: It's a complex ringsystem: a(n) 5-membered ring and a(n) 5-membered ring condensed to each other containing a(n) O atom in the 5-membered ring a(n) N atom in the 5-membered ring, and the condensed system has 1 bridge(s): a(n) 3-membered bridge containing a(n) N atom a(n) O atom.
2. találat: 	2. hit: It's a condensed compound: a(n) 5-membered ring and a(n) 6-membered ring condensed to each other containing a(n) N atom in the 5-membered ring and a(n) O atom in the 5-membered ring and a(n) N atom in the 6-membered ring and a(n) N atom in the 6-membered ring.
3. találat: 	3. hit: It's a condensed compound: a(n) 5-membered ring and a(n) 5-membered ring condensed to each other containing a(n) N atom in the 5-membered ring and a(n) N atom in the 5-membered ring and a(n) N atom in the 5-membered ring.

4. táblázat. A koffeinre generált első 3 szerkezet.

Szerkezet	Output
1. találat: 	1. hit:The structure contains 2 isolated ring systems: The 1. isolated ring system: It's a 5-membered ring containing a(n) N atom and a(n) O atom and a(n) N atom. The 2. isolated ring system: It's a 6-membered ring.
2. találat: 	2. hit:The structure contains 2 isolated ring systems: The 1. isolated ring system: It's a(n) pyrazole ring. The 2. isolated ring system: It's a 6-membered ring.
3. találat: 	3. hit:The structure contains 2 isolated ring systems: The 1. isolated ring system: It's a 5-membered ring containing a(n) O atom and a(n) N atom. The 2. isolated ring system: It's a 6-membered ring.

5. táblázat. A 4-amino-antipirinre generált első 3 szerkezet.

Szerkezet	Output
1. találat: 	1. hit: It's a condensed system: a(n) 6-; a(n) 6-; and a(n) 6-membered ring condensed to each other containing a(n) N atom in the 6-membered ring and a(n) N atom in the 6-membered ring.
2. találat: 	2. hit: It's a condensed system: a(n) 6-; a(n) 6-; and a(n) 6-membered ring condensed to each other containing a(n) N atom in the 6-membered ring and a(n) N atom in the 6-membered ring.
3. találat: 	3. hit: It's a bridged system: a(n) 10-membered ring with a(n) 2-membered bridge and a(n) 2-membered bridge containing a(n) N atom in the main chain and a(n) N atom in the main chain.

6. táblázat. Az 1,10-fenantrolinra generált első 3 szerkezet.

A találati lista összegzése az egyes vegyületek esetében:

Koffein:

Results contain:
- 1733 monocycles;
- 31 spiro compounds;
- 1828 condensed compounds;
- 566 bridged compounds;
- 740 condensed compounds with bridges;
- 54 more complex ringsystems.
- 784 compounds have more than one isolated ring.

4-Amino-antipirin:

Results contain:
- 2291 monocycles;
- 29 spiro compounds;
- 249 condensed compounds;
- 14 bridged compounds;
- 0 condensed compounds with bridges;
- 0 more complex ringsystems.
- 1245 compounds have more than one isolated ring.

1,10-Fenantrolin:

Results contain:
- 5 monocycles;
- 0 spiro compounds;
- 271 condensed compounds;
- 24 bridged compounds;
- 166 condensed compounds with bridges;
- 0 more complex ringsystems.
- 66 compounds have more than one isolated ring.

Az összegző listák segítségével könnyen át lehet látni, hogy milyen típusú szerkezetekre lehet számítani.

4.4. Eredmények és következtetések

A Bruker BioSpin Corporation *Small Molecule Structure Elucidation* alkalmazásának működéséről a tesztvegyületek mérési eredményeinek tükrében is megerősítést nyert, hogy még jó felbontású spektrumok használatakor is sok találat generálódhat. Ez például olyan vegyületek esetében történhet, amelyeknél alacsony a H/CNO arány, mivel a kevés számú proton nem tudja „felderíteni” a szerkezetet. Ezeket a nagyszámú találati listákat a *Small Molecule Structure Elucidation* jelenlegi formájában nem tudja csökkenteni, mivel a szénspektrumok generálása több órába telne és a manuális spektrumkönyvtárban való keresés se kivitelezhető belátható időn belül több ezer spektrumra. Ezekben az esetekben tehát különösen hasznos, ha van alternatív módja annak, hogy a találatokról információhoz jussunk, illetve, hogy egy összegző áttekintést lássunk a találati listáról, ráadásul ezeket rövid idő alatt kapjuk meg.

Ilyen áttekintés és találati jellemzés lehetőségét teremtettem meg, az alkalmazás sdf kiterjesztésű outputfájljait felhasználva. Ehhez Java nyelven készítettem egy algoritmust, ami első lépésében a Collapsing *P*-graph metódus alapján meghatározza a gráfként adott molekulában az összes gyűrűt. Ezután készít egy gyűrűkonnektivitási mátrixot, melynek elemei megadják, hogy két körnek hány közös atomja van, és a továbbiakban ezt használja a gyűrűk osztályozására. Az osztályozás után az algoritmus képes meghatározni, hogy monociklusos, spiro-, kondenzált, áthidalt, híddal rendelkező kondenzált vagy ezeknél összetettebb gyűrűrendszer szerepel-e a molekulában. Ha egy gyűrű nem kapcsolódik más gyűrűkkel (vagyis monociklusos), akkor az algoritmus több ismert szerkezettel is összehasonlítja azt, és ha egyezést talál, azt meg is adja. A gyűrűvizsgálatot minden egyes találatra elvégzi, majd azt is kijelzi, hogy az egyes gyűrűrendszer fajtákból hány fordul elő a találatok között. A tesztvegyületekre kapott eredmények alátámasztották, hogy az algoritmus helyesen elemzi a gyűrűrendszereket.

A monociklusos, spiro-, kondenzált, áthidalt, híddal rendelkező kondenzált vagy ezeknél összetettebb gyűrűrendszer felosztást érdemes volna még tovább árnyalni és kisebb alcsoportokat is létrehozni. Ilyen lehetne például kondenzált rendszerek esetében, hogy hány gyűrű építi fel a kondenzált rendszert, illetve hogy ezek milyen hosszúak; áthidalt szerkezetek esetében pedig, hogy hány hidat tartalmaznak. Ezek a kiegészítések még tovább differenciálnák a találatokat, és így közelebb vinnének a helyes szerkezet megtalálásához.

5. Összefoglalás

Szakdolgozati munkámhoz a kiindulópontot a Bruker Corporation *Small Molecule Structure Elucidation* szoftverje jelentette, melynek feladata, hogy egy adott anyagról mért NMR-spektrumokból lehetséges szerkezeteket állítson elő. Ennek a szoftvernek nagy problémája, hogy sok esetben (főleg kevés hidrogént tartalmazó anyagokra) nagyszámú találatot generál, amik között csak nagyon idő- és munkaigényes módokon lehet megkülönböztetni a valószínűbb és a kevésbé valószínű szerkezeteket.

Emiatt a szakdolgozatom céljául tűztem ki, hogy a *Small Molecule Structure Elucidation* output fájlját használva ezekről a találatokról a bennük levő gyűrűrendszer alapján egy elsődleges információt határozzak meg, ami hozzásegítheti a felhasználót a tényleges szerkezet megtalálásához. Ehhez gráfelméleti algoritmusokat alkalmaztam, kihasználva azt a két tényt, hogy a molekuláris szerkezetek kémiai gráfoknak is tekinthetők, illetve, hogy a gráfokban előforduló körök meghatározása rég foglalkoztatja mind a matematikusokat, mind a kémikusokat. A legcélravezetőbb megoldás volt az összes lehetséges kör meghatározása, majd ezek csoportosítása valamilyen szisztéma szerint. Alapvetően a gyűrűk két csoportját különböztettem meg: az egyszerű és a komplex gyűrűket, azon tulajdonságuk alapján, hogy a tagjaik között van-e olyan kötés, ami nem szerepel magában a gyűrűben. Az egyszerű gyűrűket célszerű volt további két alcsoportra bontani, a valódi egyszerű és a pszeudo-egyszerű gyűrűk alcsoportjára. A kémiai gyűrűrendszer pontos leírásához mind a két alcsoport szükségesnek bizonyult.

A gyűrűrendszerben a gyűrűk egymáshoz viszonyított helyzetét a valódi egyszerű gyűrűk egymással közös atomjainak száma határozza meg. Ennek alapján beszélhetünk izolált, spiro-, kondenzált vagy áthidalt gyűrűkről, vagy olyan komplexebb gyűrűrendszerekről, amikben ezek közül többféle gyűrű is megtalálható. Az egyes találatok gyűrűrendszerének jellemzésén túl az összes találatról egy áttekintő összefoglalást is készítettem.

Az alkalmazott algoritmust három kismolekulás szerves vegyületen teszteltem: a koffeinen, a 4-amino-antipirinen, illetve az 1,10-fenantrolinon. A *Small Molecule Structure Elucidation* mindegyik vegyületre nagyszámú (több száztól több ezerig terjedő) találatot generált. Az egyes találatok feldolgozása során meghatározott gyűrűrendszer a ténylegesen a szerkezetben lévővel megegyezett, az összefoglaló listák pedig segítenek a nagyszámú találat áttekintésében.

6. Summary

My thesis work is based on the *Small Molecule Structure Elucidation* software of the Bruker Corporation, which generates possible structures from measured NMR-spectra of a compound. This software's biggest problem in many cases is that it generates (mainly for compounds containing small amounts of hydrogen) a great number of structures, among which it is hard to distinguish between the least probable and more probable ones.

For this reason, the goal of my thesis is to define a primary information based on the ring system in the generated structures using the output file of the *Small Molecule Structure Elucidation* software. This primary information helps the user in determining the real structure. I implemented graph theory algorithms taking advantage of two facts: that molecular structures can be viewed as chemical graphs; and that the cycle perception has been a preoccupation of many scientists in the fields of both mathematics and chemistry. The most effective solution is the determination of all the possible cycles and their categorization based on a system. I have differentiated two basic groups of rings: simple and complex, based on their property whether there is an edge between the members of the ring, that is not a part of the ring. It is effective to divide simple rings into two subcategories: real simple rings and pseudo simple rings. For the description of the chemical ring system both subcategories were proved to be necessary.

The relation of the chemical rings is given by the number of the mutual atoms of the real simple cycles. According to this we can speak of isolated, spiro, condensed and bridged compounds, or more complex ring systems that include more of the latter. In addition to the characterization of the ring system of each structure, I also provided an insightful summary for all of the structures.

I tested the algorithm on three organic compounds: caffeine, 4-amino-antipyrine and 1,10-phenanthroline. *Small Molecule Structure Elucidation* generated a great number (ranging from one hundred to a thousand) of hits for all of the compounds. In the case of each and every hit the real ringsystems were identical with the determined system, and the summary lists help in the review of the large amount of structures.

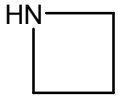
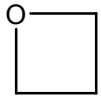
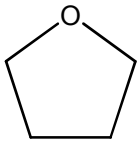
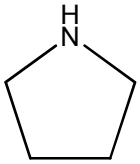
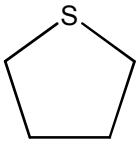
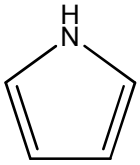
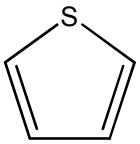
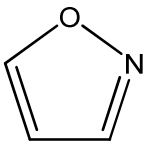
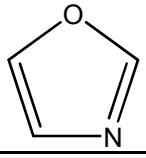
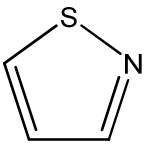
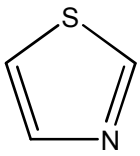
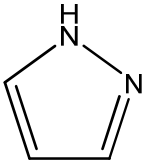
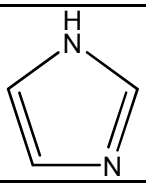
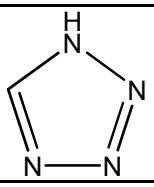
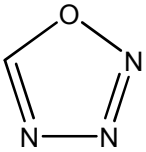
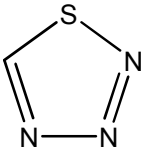
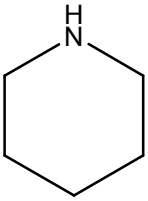
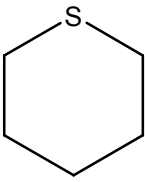
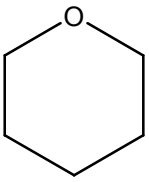
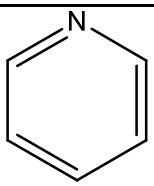
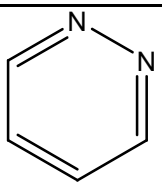
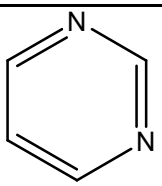
Irodalomjegyzék

[1] http://www.bruker.com 2012.04.14.
[2] Structure Elucidation with CMC-se™ - Instruction Manual, Bruker BioSpin GmbH. (2011)
[3] P. Sohár: Mágneses magrezonencia-spektroszkópia. Budapest, Akadémiai Kiadó. (1976)
[4] G. Bodenhausen, D. J. Ruben: "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy" *Chem. Phys. Lett.* **69** 185-189 (1980).
[5] W. Willker, D. Leibfritz, R. Kerssebaum, W. Bermel: "Gradient selection in inverse heteronuclear correlation spectroscopy" *Magn. Reson. Chem.* **31** 287-292 (1993).
[6] R. D. Boyer, R. Johnson, K. Krishnamurthy: "Compensation of refocusing inefficiency with synchronized inversion sweep (CRISIS) in multiplicity-edited HSQC" *J. Magn. Reson.* **165** 253-259 (2003).
[7] C. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat, L. E. Kay: "Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λ N-peptide/*boxB* RNA complex" *J. Am. Chem. Soc.* **119** 6711-6721 (1997).
[8] A. Bax, M. F. Summers: "Proton and carbon-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR" *J. Am. Chem. Soc.* **108** 2093-2094 (1986).
[9] D. O. Cicero, G. Barbato, R. Bazzo: "Sensitivity enhancement of a two-dimensional experiment for the measurement of heteronuclear long-range coupling constants, by a new scheme of coherence selection by gradients" *J. Magn. Reson.* **148** 209-213 (2001).
[10] S. L. Patt, J. N. Shoolery: "Attached proton test for carbon-13 NMR" *J. Magn. Reson.* **46** 535-539 (1982).
[11] D. M. Doddrell, D. T. Pegg, M. R. Bendall: "Distortionless enhancement of NMR signals by polarization transfer" *J. Magn. Reson.* **48** 323-327 (1982).
[12] K. Nagayama, A. Kumar, K. Wüthrich, R. R. Ernst: "Experimental techniques of two-dimensional correlated spectroscopy" *J. Magn. Reson.* **40** 321-334 (1980).
[13] W. P. Aue, E. Bartholdi, R. R. Ernst: "Two-dimensional spectroscopy. Application to nuclear magnetic resonance" *J. Chem. Phys.* **64** 2229-2246 (1976).
[14] A. A. Shaw, C. Salaun, J. F. Dauphin, B. Ancian: "Artifact-free PFG-enhanced Double-Quantum-Filtered COSY experiments" *J. Magn. Reson. A* **120** 110-115 (1996).
[15] B. Ancian, I. Bourgeois, J. F. Dauphin, A. A. Shaw: "Artifact-free pure absorption PFG-enhanced DQF-COSY spectra including a gradient pulse in the evolution period" *J. Magn. Reson.* **125** 348-354 (1997).
[16] http://eos.univ-reims.fr/LSD 2012.04.14.
[17] http://nmrpredict.orc.univie.ac.at 2012.04.14.
[18] A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grier, A. Leland, J. Laufer: "Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited" *J. Chem. Inf. Mod.* **32** 244-255 (1992).
[19] B. Andrásfai: Gráfelmélet. Szeged, Polygon. (1994)
[20] G. Katona, A. Recski, C. Szabó: A számítástudomány alapjai. Budapest, Typotex. (2006)
[21] A. T. Balaban: "Chemical graphs: Looking back and glimpsing ahead" *J. Chem. Inf. Comput. Sci.* **35** 339-350 (1995).

[22] G. M. Downs, V. J. Gillet, J. D. Holliday, M. F. Lynch: "Review of ring perception algorithms for chemical graphs" *J. Chem. Inf. Comput. Sci.* **29** 172-187 (1989).
[23] F. Berger, C. Flamm, P. M. Gleiss, J. Leydold, P. F. Stadler: "Counterexamples in chemical ring perception" *J. Chem. Inf. Comput. Sci.* **44** 323-331 (2004).
[24] E. J. Corey, G. A. Petersson: "An algorithm for machine perception of synthetically significant rings in complex cyclic organic structures" *J. Am. Chem. Soc.* **94** 460-465 (1972).
[25] W. T. Wipke, T. Dyott: "Use of ring assemblies in a ring perception algorithm" *J. Chem. Inf. Comput. Sci.* **15** 140-144 (1975).
[26] J. Gasteiger, C. Jochum: "An algorithm for the perception of synthetically important rings" *J. Chem. Inf. Comput. Sci.* **19** 43-48 (1979).
[27] J. Figureas: "Ring perception using breadth-first search" *J. Chem. Inf. Comput. Sci.* **36** 986-991 (1996).
[28] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen: "The chemistry development kit (CDK): An open-source Java library for chemo- and bioinformatics" *J. Chem. Inf. Comput. Sci.* **43** 493-500 (2003).
[29] C. J. Lee, Y.-M. Kang, K.-H. Cho, K. T. No: "A robust method for searching the smallest set of smallest rings with a path-included distance matrix" *PNAS* **106** 17355-17358 (2009).
[30] A. Zamora: "An algorithm for finding the Smallest Set of Smallest Rings" *J. Chem. Inf. Comput. Sci.* **16** 40-43 (1976).
[31] J. B. Hendrickson, D. L. Grier, A. G. Toczko: "Condensed structure identification and ring perception" *J. Chem. Inf. Comput. Sci.* **24** 195-203 (1984).
[32] M. Plotkin: "Mathematical basis of ring-finding algorithms in CIDS" *J. Chem. Doc.* **11** 60-63 (1971).
[33] H. Nickelsen: "Ringbegriffe in der chemie-dokumentation" *Nachr. Dok.* **3** 121-123 (1971).
[34] E. J. Corey, W. T. Wipke, R. D. Cramer, W. J. Howe: "Techniques for perception by a computer of synthetically significant structural features in complex molecules" *J. Am. Chem. Soc.* **94** 431-439 (1972).
[35] S. Fujita: "A new algorithm for selection of synthetically important rings. The Essential Set of Essential Rings for organic structures" *J. Chem. Inf. Comput. Sci.* **28** 78-82 (1988).
[36] S. Fujita: "Logical perception of ring opening, ring closure, and rearrangement reactions based on imaginary transition structures. Selection of the Essential Set of Essential Rings (ESER)" *J. Chem. Inf. Comput. Sci.* **28** 1-9 (1988).
[37] A. T. Balaban, P. Filip, T. S. Balaban: "Computer program for finding all possible cycles in graphs" *J. Comput. Chem.* **6** 316-329 (1985).
[38] T. Hanser, P. Jauffret, G. Kaufmann: "A new algorithm for exhaustive ring perception in a molecular graph" *J. Chem. Inf. Comput. Sci.* **36** 1146-1152 (1996).

Függelék

A felismerhető monociklusok:

Szerkezet	Output	Szerkezet	Output	Szerkezet	Output
	oxirane		aziridine		thiirane
	azetidine		oxetane		thietane
	tetrahydrofuran		pyrrolidine		tetrahydrothiophene
	furan		pyrrole		thiophene
	isoxazole		oxazole		isothiazole
	thiazole		pyrazole		imidazole
	tetrazole		oxatriazole		thiatriazole
	piperidine		tetrahydro-thiopyran		tetrahydro-pyran
	pyridine		pyridazine		pyrimidine