

# MOLEKULASZIMMETRIA VEKTORALGEBRAI MEGHATÁROZÁSA ÉS ALKALMAZÁSA A MODELLEZÉSBEN

Doktori (Ph.D.) értekezés

GYEVI-NAGY LÁSZLÓ  
okleveles vegyész

Témavezető:  
DR. TASI GYULA  
egyetemi docens

Kémia Doktori Iskola



Alkalmazott és Környezeti Kémiai Tanszék  
Természettudományi és Informatikai Kar  
Szegedi Tudományegyetem  
Szeged, 2018

# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>1</b>
<b>2. Irodalmi áttekintés</b>	<b>4</b>
2.1 Molekulaszimmetria meghatározása . . . . .	4
2.1.1 A szimmetria jelentősége . . . . .	4
2.1.2 Korábbi módszerek . . . . .	7
2.1.2.1 Mátrix- és tenzoralképzési módszerek . . . . .	7
2.1.2.2 Gráfelméleti módszerek . . . . .	8
2.2 <i>Ab initio</i> termokémiai modellek . . . . .	10
2.2.1 Kompozit kvantumkémiai módszerek . . . . .	10
2.2.1.1 Gaussian- <i>n</i> és CBS- <i>n</i> modellek . . . . .	11
2.2.1.2 Weizmann- <i>n</i> és HEAT modellek . . . . .	11
2.2.1.3 További modellek . . . . .	12
2.2.2 Fókuszpont analízis . . . . .	13
2.3 QSPR modellezés . . . . .	13
2.3.1 Topológiai indexek . . . . .	14
2.3.2 Topografikus indexek . . . . .	15
2.3.3 Szimmetria deskriptorok . . . . .	16
2.3.4 Általánosan alkalmazott modellfüggvények . . . . .	17
<b>3. Célkitűzés</b>	<b>18</b>
<b>4. Elméleti összefoglaló</b>	<b>19</b>
4.1 Szimmetriaanalízis . . . . .	19
4.1.1 Szimmetriaműveletek . . . . .	19
4.1.2 A transzformációra vonatkozó megszorítások . . . . .	20
4.2 FPA modell entalpiakülönbségek becslésére . . . . .	23
4.2.1 Termokémiai mennyiségek számítása . . . . .	23
4.2.2 A teljes elektronenergia járulékaik meghatározása . . . . .	24
4.2.2.1 CCSD(T)/CBS energia . . . . .	24
4.2.2.2 Kis korrekciók . . . . .	25
4.2.3 A zérusponti rezgési energia számítása . . . . .	26

4.2.4	Hibas számítás . . . . .	26
4.2.5	Alkalmazott programok . . . . .	27
4.3	Az entalpiakülönbség kísérleti meghatározása . . . . .	27
4.4	QSPR modellezés . . . . .	29
4.4.1	Módosított topológiai indexek . . . . .	29
4.4.2	Nemlineáris modellfüggvények . . . . .	31
<b>5.</b>	<b>Eredmények és értékelésük</b>	<b>33</b>
5.1	Szimmetriaanalízis . . . . .	33
5.1.1	Algoritmus . . . . .	33
5.1.2	Implementáció . . . . .	39
5.1.3	Jellemzés, összehasonlítás más szoftverekkel . . . . .	41
5.2	A pentánkonformerek entalpiakülönbségei . . . . .	45
5.2.1	Az entalpiakülönbségek kvantumkémiai meghatározása . . . . .	45
5.2.1.1	A referencia geometria előállítása . . . . .	45
5.2.1.2	Az elektronenergia számítása . . . . .	46
5.2.1.3	A zérusponti rezgési energia és a hőmérsékleti korrekció . . . . .	47
5.2.1.4	Összehasonlítás korábbi számítási eredményekkel . . . . .	49
5.2.2	Az entalpiakülönbségek meghatározása kísérleti adatok alapján . . . . .	52
5.3	Az entalpiakülönbségek QSPR modellezése . . . . .	57
5.3.1	Molekulaszimmetrián alapuló deskriptorok . . . . .	57
5.3.2	Topografikus indexek . . . . .	59
5.3.3	Hőmérsékletfüggés modellezése . . . . .	63
5.3.4	Nemlineáris modellfüggvény . . . . .	66
5.3.5	Az elsőfajú nyeregponatok modellezése . . . . .	68
5.3.6	Butánkonformerek . . . . .	71
<b>6.</b>	<b>Összefoglalás</b>	<b>74</b>
<b>7.</b>	<b>Summary</b>	<b>77</b>
	<b>Köszönetnyilvánítás</b>	<b>80</b>
	<b>Irodalomjegyzék</b>	<b>81</b>
	<b>Függelék</b>	<b>93</b>

# Alkalmazott gyakoribb rövidítések

<b>CNP</b>	<i>complete nuclear permutation group,</i> teljes magpermutációs csoport
<b>CNPI</b>	<i>complete nuclear permutation inversion group,</i> teljes magpermutációs–inverziós csoport
<b>MS</b>	<i>molecular symmetry group,</i> molekuláris szimmetriacsoport
<b>FC</b>	<i>frozen core,</i> befagyasztott mag közelítés
<b>CBS</b>	<i>complete basis set,</i> teljes báziskészlet
<b>FPA</b>	<i>focal point analysis,</i> fókuszpont analízis
<b>HF</b>	Hartree–Fock módszer
<b>CC</b>	<i>coupled cluster,</i> csatolt-klaszter módszer
<b>MP<math>n</math></b>	$n$ -edrendű Møller–Plesset perturbációs módszer
<b>DBOC</b>	<i>diagonal Born–Oppenheimer correction,</i> diagonális Born–Oppenheimer korrekció
<b>DKH</b>	Douglas–Kroll–Hess Hamilton-operátor szerinti skaláris relativisztikus korrekció
<b>ZPVE</b>	<i>zero point vibrational energy,</i> zérusponti rezgési energia
<b>OLLS</b>	<i>ordinary linear least squares,</i> közönséges lineáris legkisebb négyzetek módszere
<b>WLLS</b>	<i>weighted linear least squares,</i> súlyozott lineáris legkisebb négyzetek módszere
<b>ONLS</b>	<i>ordinary nonlinear least squares,</i> közönséges nemlineáris legkisebb négyzetek módszere
<b>QSPR/QSAR</b>	<i>quantitative structure property/activity relationship,</i> kvantitatív szerkezet-tulajdonság/aktivitás összefüggés
<b>MLR</b>	multilineáris regresszió

# 1. Bevezetés

A kvantumkémia jelentős fejlődésen esett át az elmúlt néhány évtizedben. Az elméleti kémia alkalmazása az informatika gyors fejlődésének köszönhetően egyre rutinszerűbbé válik, melynek hiányában komoly eredmények elérése számos területen már szinte elképzelhetetlen. E fejlődés következtében a számítási eredmények pontossága manapság már vetekszik a kísérleti eredményekkel, sőt bizonyos esetekben a kvantumkémiai számítások sokkal konzisztensebb adatokat szolgáltatnak.<sup>1,2</sup> Ez a precizitás szükséges az egyes rendszerek egyre jobb modellezéséhez valamint technológiai folyamatok tervezéséhez. A kvantumkémia segítségével akár kísérletileg meg nem határozható mennyiségek – például reakcióutak<sup>3-5</sup> – is számíthatók. Ebből kifolyólag a modern kvantumkémia egyik feladata, hogy kis molekulákra nagy pontosságú referenciaértékeket szolgáltatson. A pontosság növelése azonban egyre összetettebb számítások végrehajtását követeli meg, amelyek megvalósítása jelentős időbefektetést igényel.

A számítások pontosságának, illetve elméleti szintjének emelése mellett a futásidő csökkentése például további egyszerűsítések bevezetésével lehetséges. Ezek az elhanyagolások jelentősen csökkenthetik a modell struktúrájának komplexitását, lehetővé téve magasabb szintű számítások végrehajtását is. Az alkalmazott megszorítások egy adott szinten természetesen rontják az eredmény minőségét, azonban hozzájárulnak a hatékonyság, ezáltal az elméleti szint növeléséhez. Az ezzel járó nyereség sokszor meghaladja a közelítések által okozott pontatlanság növekedését.

A rendelkezésre álló kvantumkémiai szoftverek is sokféle közelítő módszer alkalmazását támogatják, melyek felhasználhatók például egy iterációs módszer konvergenciájának biztosítására. Erre a célra több extrapolációs és interpolációs módszer áll rendelkezésre. Ilyen eljárás többek között a Pulay-féle DIIS (*direct inversion in the iterative subspace*) módszer,<sup>6-10</sup> annak módosított változataival, valamint az SCF eljárás első néhány lépésében gyakran alkalmazott csillapítás (*damping*)<sup>11</sup> módszere.

A modell lényegesen leegyszerűsíthető a vizsgált rendszer particionálásával, majd annak csak egyes részeire bevezetett megszorítások, illetve közelítések segítségével. Ez a megközelítés lehetővé teszi a fontosabb régiókra történő pontosabb számítások végrehajtását anélkül, hogy a rendszer méretét feleslegesen növelnénk. Az erre a célra gyakran használt egyik egyszerűsítő módszer a befagyasztott mag (*frozen core*, FC) közelítés,<sup>12,13</sup> melynek alkalmazásakor a számítások elvégzése során egy bizonyos mag régióban (általában az atomtörzsben) változatlan betöltési számokat alkalmazva határozzuk meg a korrelációs korrekciót. A választott mag elektronjainak hatása egy effektív potenciál (*effective core potential*, ECP)<sup>14,15</sup> segítségével is modellezhető. Amennyi-

ben szükséges, a mag régió és a többi (vegyérték)elektron kölcsönhatásából származó hiba (*core-core*, *core-valence* korrekció) ezután alacsonyabb szinten becsülhető az összes elektron korrelációjának figyelembevételével.

A kvantumkémiai számítások során a sebességhatározó lépés a molekuláris integrálok előállítás, mely folyamat egy segéd bázis felhasználásával jelentősen leegyszerűsíthető. Erre a bázisra projektálva a bázisfüggvények szorzatai, a kételektron integrálok számításához szükséges műveletek száma jelentősen redukálható. Ez a módszer a kapott egyenletekben szereplő művelet egységfelbontáshoz való hasonlósága miatt RI (*resolution of identity*) vagy DF (*density fitting*) módszer<sup>13,16,17</sup> néven ismert.

A számítandó integrálok számának csökkentésére azok szimmetriájának kihasználása szintén alkalmas, mégpedig (a) az integrálok permutáció invarianciájának<sup>11</sup> figyelembevételével, mely jelenség a Hamilton-operátor hermiticitásának és a valós bázisfüggvények használatának tudható be, illetve (b) a magkonfiguráció szimmetriájának kiaknázásával,<sup>11</sup> ami a konvencionális (atomon centrált) bázisfüggvények alkalmazásából eredeztethető.

Egy mód tehát a számítások komplexitásának csökkentésére, valamint ezáltal komolyabb számítások végrehajtására, a vizsgált molekula szimmetriájának kihasználásában rejlik. A molekula pontcsoportjának detektálását követően ugyanis megadhatók az eltűnő, valamint az egymással megegyező integrálok, melynek oka az alkalmazott bázis alakjában keresendő. Mivel a gyakorlatban legtöbbször az atommagokon centrált bázisfüggvények<sup>18</sup> használata jellemző, a bázis a magkonfiguráció szimmetriájával, legalábbis annak egy részcsoportjával rendelkezik. Továbbá a megoldás megszorítható úgy, hogy visszatükrözze a molekula szimmetriáját, így nagymértékű sebességnövekedés érhető el. Bár a számított molekulapályák nem minden esetben rendelkeznek a magkonfiguráció szimmetriájával, közönséges molekulák esetén – melyeknél a Fock-operátor egyelektronos, szimmetrikus része dominál – ez a szimmetria mégsem sérül, ezt kihasználva pedig jelentősen egyszerűsíthető a rendszert leíró modell.

A molekulaszimmetria ismerete statisztikus termodinamikai számítások során is elengedhetetlen, hiszen a forgási állapotösszeg a korrekt forgási szimmetriaszám hiányában nem állítható elő, vagyis a rosszul felismert szimmetria következtében a hőmérsékleti korrekciókra téves értékek adódnak.

Bár a molekulamodellőző szoftverek általában képesek felhasználni a molekula pontcsoportját<sup>19–24</sup> a számítások során, sajnos nem mindegyik program tudja azt meghatározni. Amennyiben valamely alkalmazásban implementálásra is került egy a szimmetria meghatározására szolgáló eljárás, az sokszor nem a helyes vagy nem a legnagyobb pontcsoportot állítja elő, így a molekula szimmetriája nem használható ki a legteljesebb mértékben – hiányzó pontcsoport esetén egyáltalán nem –, ez pedig rontja a szoftver hatékonyságát.

Egy vegyület különféle tulajdonságait az azt felépítő molekulák szerkezete – mag-

konfigurációja és elektronszerkezete – determinálja. Az elektronszerkezet leírása rendkívül bonyolult feladat, ezért kézenfekvőnek tűnhet annak mellőzése a modellépítés során. Szélsőséges esetben a modell addig egyszerűsíthető, amíg már csak a magkonfigurációtól függő ún. deszkriptorok szolgálnak a modellezni kívánt tulajdonság leírására. Ennek alapja, hogy a Hamilton-operátor, melynek sajátfüggvényei a rendszer lehetséges állapotait és tulajdonságait határozzák meg, a magkoordinátáktól és az elektronkoordinátáktól függ, ezek a koordináták pedig determinálják egy kvantumkémiai rendszer jellemzőit. A leírt közelítés – egyéb paraméterek bevezetése mellett – lényegében az elektronkoordináták elhagyásának felel meg. Végső soron tehát az elektronszerkezetet a magkonfiguráció függvényének tekintjük. Az elektronkoordináták elhagyása a leírásból értelemszerűen a modell szignifikáns egyszerűsítését jelenti, azonban a gyakorlat azt mutatja, hogy még a leggyakrabban alkalmazott multilineáris modell is képes lehet fizikai-kémiai tulajdonságok vagy biológiai aktivitás nagy pontosságú jellemzésére. Az imént bemutatott eljárást kvantitatív szerkezet-tulajdonság/aktivitás összefüggés (*quantitative structure property/activity relationship*, QSPR/QSAR)<sup>25</sup> módszernek nevezik.

Mint látjuk, a Hamilton-operátor szimmetriája fontos szerepet tölt be a molekulapályák jellegének alakulásában, így feltételezhető, hogy a szimmetriából származtatott deszkriptorok között a kvantumkémiai számításokból származó értékek (pl. energiaszintek, termokémiai mennyiségek, spektroszkópiai tulajdonságok, elektronsűrűség) becslésére alkalmas mennyiségek is szerepelnek. A QSPR/QSAR modellezés során mind a deszkriptorokban, mind pedig a paraméterekben lineáris modellfüggvény használata igen elterjedtnek számít. Nemlineáris modell használatára is található példa az irodalomban,<sup>26–28</sup> azonban az ilyen típusú modellezés során nem ez a legjellemzőbb eljárás, hiszen az efféle modellek sok esetben igen bonyolultak és nehezen vagy egyáltalán nem értelmezhetők. Az említett deszkriptorok és nemlineáris modellek vizsgálata érdeklődésre tarthat számot, mivel így a korábbi nagy pontosságú eredményekből – a felállított modell precizitásától függő pontossággal – más specieszek tulajdonságaira is predikciót lehetne tenni.

Egy termokémiai tulajdonságok leírására felállított modell a konformációs flexibilitás megértése szempontjából is előnyös lehet. Amennyiben rendelkezésre állna egy teljes konformációs potenciális energiafelület jellemzésére alkalmas modell, lehetővé válna az adott molekula tulajdonságainak tetszőleges konformációs állapotában történő becslése bonyolult, időigényes számítások végrehajtása nélkül. Ezen tulajdonságok gyors és pontos meghatározása az intra- és intermolekuláris kölcsönhatások, valamint az azok által vezérelt folyamatok (pl. fehérjefeltekeredés) értelmezését is elősegíthetné.

## 2. Irodalmi áttekintés

### 2.1. Molekulaszimmetria meghatározása

#### 2.1.1. A szimmetria jelentősége

A modern fizika alapvető jelentőségű megállapítása a Noether-tétel, mely kimondja, hogy egy fizikai rendszerben megjelenő folytonos szimmetriához megmaradási tétel, illetve megmaradó mennyiség (Noether-töltés) tartozik.<sup>29</sup> Ez a törvény lényeges szerepet játszik például a speciális relativitáselméletben. A folytonos vagy akár a diszkrét szimmetriatulajdonságok általában is fontos szerepet töltenek be a részecskefizikában, valamint a kémiában és a szilárdtestfizikában. Ennek oka abban rejlik, hogy egy anyag számos fizikai tulajdonságára nézve pusztán szimmetria megfontolások alapján – akár a rendszer teljes Hamilton-operátorának ismerete nélkül is – érdemi következtetéseket vonhatunk le.<sup>30</sup>

Egy kvantumrendszer dinamikai viselkedését annak Hamilton-operátora határozza meg. Ha ez az operátor egy tetszőleges transzformációra nézve invariáns, akkor a transzformált rendszer megkülönböztethetetlen az eredetitől, vagyis térben és időben azzal azonos módon fejlődik. Definíció szerint szimmetria transzformációnak nevezzük az unitér, Hamilton-operátorral kommutáló operátorok ( $\hat{\mathbf{R}}$ ) hatását. Ezek az operátorok a rendszer egyik sajátállapotát egy azonos energiasajátértékhez tartozó sajátállapotba transzformálják át a 2.1 egyenlet szerint.

$$\hat{\mathbf{H}}\hat{\mathbf{R}}\Psi_n = \hat{\mathbf{R}}\hat{\mathbf{H}}\Psi_n = \hat{\mathbf{R}}E_n\Psi_n = E_n\hat{\mathbf{R}}\Psi_n \quad (2.1)$$

A Hamilton-operátorral kommutáló unitér operátorok csoportot alkotnak, melyet a Hamilton-operátor csoportjának nevezünk.<sup>30</sup> Bizonyítható, hogy a kinetikus energia operátora bármely unitér operátorral kommutál. A Hamilton-operátor szimmetriája tehát a potenciális energia operátor pontos alakjától függ. Ez a leggyakoribb esetben – külső potenciál nélkül, kizárólag a Coulomb-kölcsönhatást figyelembe véve – megegyezik a magkonfiguráció szimmetriájával. A Hamilton-operátor csoportja a leggyakoribb esetben tehát a molekula pontcsoportjával egyezik meg.

Amennyiben a 2.1 egyenletben  $\hat{\mathbf{R}}\Psi_n \neq \Psi_n$ , degenerált állapotról beszélünk. Tehát  $\hat{\mathbf{R}}\Psi_n$  is a Hamilton-operátor  $E_n$  sajátértékhez tartozó sajátfüggvénye. Ebből az következik, hogy az  $E_n$  sajátértékhez tartozó összes  $\hat{\mathbf{R}}\Psi_j$  sajátfüggvény felírható a következő



alakban:

$$\hat{\mathbf{R}}\Psi_j = \sum_{k=1}^d R_{jk}\Psi_k, \quad (2.2)$$

ha  $\Psi_j$ -hez tartozó  $\lambda_j$  sajátérték  $d$ -szeres multiplicitású. Ezek az  $\mathbf{R}$  mátrixok egy reprezentációt képeznek. Egy adott energiához tartozó sajátfüggvények a Hamilton-operátor csoportjának egy invariáns alterét feszítik ki. A kvantummechanika irreducibilitási axiómája kimondja, hogy a Hamilton-operátor sajátfüggvényei a Hamilton-operátor csoportjának irreducibilis reprezentációinak bázisfüggvényei. Az  $\hat{\mathbf{R}}$  szimmetriaműveleteket a nemdegenerált sajátállapotokra alkalmazva tehát azok irreducibilis reprezentációi állíthatók elő, az így nyert reprezentációk dimenziója pedig a megfelelő sajátérték multiplicitásával egyezik meg. Ezek szerint a degeneráció a rendszer szimmetriájának következménye, a degenerált állapotokat pedig a szimmetriatranszformációk kapcsolják össze. Megjegyzendő, hogy a szimmetriaműveletekből álló csoport irreducibilis reprezentációinak ismeretében a sajátállapotokat is a szimmetria alapján csoportosíthatjuk.

Ez az összefüggés lehetőséget teremt számos kvalitatív vizsgálat elvégzésére. A szimmetria és a csoportelmélet egyik lehetséges alkalmazását a kvantummechanikában a sajátállapotok degeneráltságának és a normálrezgések szimmetriájának a Schrödinger-egyenlet mellőzésével történő meghatározása jelenti. Egy további alkalmazási területnek a szimmetria bizonyos perturbáció általi csökkenése következtében fellépő degenerációváltozás elemzése számít. Ez a degeneráció megszüntethetőségének megállapítását, illetve a perturbált és az eredeti állapotok közötti összefüggés megismerését is lehetővé teszi.

A szimmetria kvantitatív számítások esetén is sikeresen alkalmazható, például a kvantumkémiaiában szimmetriaadaptált bázisok előállítására<sup>13</sup> vagy a molekuláris integrálok számának redukációjára.<sup>11</sup> Fontos szerepe van a szimmetriának az egymáshoz közeli frekvenciájú rezgések közötti csatolások típusának megállapításában is: azonos szimmetriájú rezgések között Fermi-kölcsönhatások, eltérő szimmetriájú rezgések között pedig Coriolis-csatolások lépnek fel.

A szimmetriatulajdonságok két csoportra oszthatók. Az egyik csoportba (folytonos szimmetria) a térbeli és az időbeli transláció (a tér és az idő homogenitását fejezi ki), valamint a térbeli rotáció (a tér izotropiájának következménye) tartozik.

A másik csoport a diszkrét szimmetriákat foglalja magába, melyekkel kapcsolatban négy transzformációt kell megemlítenünk. Az időtükrözés (T), mely az idő irányának megfordítását jelenti, a tértükrözés (P), mely a térbeli koordináták inverzióját takarja, és a töltéstükrözés (C), mely ellentétes töltésű részecskék, antirészecskék alkalmazását jelenti, önmagukban nem szimmetriatranszformációk. Általános esetben a T-szimmetria, a P-szimmetria, a C-szimmetria, valamint az utolsó kettő kombinációjaként kapott CP-szimmetria is sérül,<sup>31</sup> a három művelet szimultán végrehajtására (CPT) nézve azonban egy tetszőleges fizikai rendszer szimmetrikusnak bizonyul.<sup>32,33</sup>

Diszkrét transzformációként fogható fel az azonos részecskék permutálása is. Ezek a műveletek permutációs csoportokat alkotnak, melyek direkt szorzata egy molekula teljes magpermutációs csoportját (*complete nuclear permutation*, CNP) adja. A Hamilton-operátor az inverzióra (tértükrözés) szintén invariáns, mely művelet az ún. identikus transzformációval alkotja az inverziós csoportot. Természetesen a Hamilton-operátor a CNP és az inverziós csoport direkt szorzatának elemeire nézve is változatlan, ezt a direkt szorzat csoportot teljes magpermutációs-inverziós csoportnak (CNPI) nevezzük.

A CNPI számossága  $K$ -féle elemből felépülő molekula esetén  $\prod_{i=1}^K N_i!$ , ha  $N_i$  az azonos rendszámú atomok száma. Ennek a csoportnak az előállítása és alkalmazása a kombinatorikus robbanás miatt nagyobb molekulák esetében nem lehetséges. A valószínűségben egyébként sem figyelhető meg minden transzformáció a CNPI elemei közül, bizonyos transzformációknak megfelelő átmeneteket leküzdhetetlen potenciálgát választhat el egymástól. Ezeket az elemeket elhagyva kapjuk a megvalósuló transzformációkból álló molekuláris szimmetriacsoportot (MS). Az MS tehát a CNPI egy alkalmasan megválasztott részcsoportha.

A Hamilton-operátor a részecskék – térbeli és spin – koordinátáitól függ. Ha bázisnak az elektronkoordinátákat, valamint a magkoordináták helyett a tömegközéppont koordinátáit, a molekula orientációját megadó Euler-szögeket és  $3N - 6$  (lineáris molekuláknál  $3N - 5$ ) rezgési kitérítési koordinátát választunk, az MS csoport további redukciójára is lehetőség nyílik. Az említett koordináták három csoportra oszthatók:

1. vibronikus koordináták (rezgési kitérítések és elektronkoordináták),
2. forgási koordináták (Euler-szögek),
3. magspin koordináták.

A MS csoport minden eleme a fenti három osztályba sorolható koordinátákra ható egy-egy transzformációnak megfelelő permutáció szorzataként írható fel.

Az első csoportra ható transzformációk halmaza azonos egy molekula pontcsoportjával. Ezek alapján egy merev, nemlineáris molekula pontcsoportja és szimmetriacsoportja izomorf, így a pontcsoport is alkalmazható a vibronikus átmenetek vizsgálatára.

A kvantumkémiai alkalmazott atomi bázisfüggvények általában a magkonfigurációénál nagyobb szimmetriával rendelkeznek, ezért a szimmetriaműveletek ezek lineáris kombinációit állítják elő. Ezek a lineáris kombinációk a molekuláris integrálok egyedi integrálokból történő előállítását is lehetővé teszik.<sup>11</sup> Egyszerűbb esetben – egy Abel-részcsoporthoz alkalmazásával – a szimmetriaműveletek hatására a bázisfüggvények permutációi állnak elő. Ez a kezelésmód a szimmetria figyelembevételét – az Abel-részcsoporthoz sokkal kisebb rendje, valamint a permutációk egyszerűbb szerkezete miatt – lényegesen egyszerűbbé teszi. A gyakorlatban ezért általában ezt a módszert használják, ami azonban a szimmetria teljes kihasználására nem alkalmas.

A szimmetria hatással van a teljes Hamilton-mátrix alakjára is, ugyanis ez a mátrix gyakran invariáns blokkokból álló blokkdiagonális formára hozható,<sup>34</sup> ami a számítások során jelentős teljesítménynövekedést eredményezhet. A molekula szimmetriájának használata más szempontból is fontos, lehetővé teszi ugyanis a megfelelő szimmetriával rendelkező hullámfüggvény előállítását.<sup>35</sup> Különböző okokból ez a szimmetria sérülhet. Ilyen ok lehet a külső elektromos vagy mágneses erőterrel való kölcsönhatás, az alkalmazott módszer bizonyos sajátságai – pl. megszorítás nélküli Hartree–Fock-módszer (*unrestricted Hartree-Fock*, UHF) az elektronspin miatt –, vagy a teljes pontcsoport helyettesítése annak egy részcsoportjával, mely az iteratív megoldási mód miatt a kezdeti kis eltéréseket a numerikus pontatlanság miatt felnagyíthatja, ami akár kvalitatíve rossz megoldást is eredményezhet.

A szimmetria tovább pontosítható a Pople-féle vázcsoport felvételével.<sup>36</sup> Ez a struktúra lehetővé teszi azonos pontcsoportba tartozó molekulák további megkülönböztetését annak pontos specifikálásával, hogy mely atomok mely szimmetriaelemeken találhatóak. Bár a vázcsoport koncepciója még nem terjedt el széleskörűen a kvantumkémiai gyakorlatban, néhány ígéretes alkalmazással azért találkozhatunk az irodalomban.<sup>37–40</sup> A legtöbb program nem képes a vázcsoport meghatározására vagy bizonyos esetekben hibásan adja meg azt, ami tovább nehezíti az abban található többlet információ feldolgozását.

## 2.1.2. Korábbi módszerek

### 2.1.2.1. Mátrix- és tenzoralképzési módszerek

Az első program, amely molekulák szimmetriáját automatikusan határozta meg a SYMAPPS<sup>41</sup> volt. Az ebben implementált algoritmus a potenciális szimmetriaelemek meghatározásával kezdődik. Magasabb rendű tengelyek esetén a program olyan atomokat keres, melyek (a) a tömegközépponttól azonos távolságra találhatóak, (b) egy síkban fekszenek, valamint (c) távolságuk a legutóbb megtalált atomtól megegyezik az első két atom távolságával. Ha az így meghatározott szabályos  $n$ -oldalú sokszög középpontján és a tömegközépponton áthaladó tengely körüli  $360^\circ/n$  szögű forgatásra nézve a molekula invariáns, szimmetriatengelyről beszélünk.

A leírt algoritmus több szempontból sem hatékony. Az  $n$ -fogású szimmetriatengelyek generálásához  $N_i$  azonos rendszámú atom esetén az összes lehetséges  $\binom{N_i}{n}$  számú kombináció vizsgálata szükséges, így egy sok azonos rendszámú atomot tartalmazó molekulánál ez a megközelítés nem célszerű.

Az eljárás másik problémája, hogy a lehetséges szimmetriaműveletek elvégzését a tengely irányvektorának vagy a sík normálvektorának  $z$ -tengely irányába történő forgatásával kezdi. Bár a művelet végrehajtása ezután valóban egyszerű, a molekula közvetlen transzformációja sokkal hatékonyabb lenne.

Egy másik megközelítés<sup>42</sup> a koordinátarendszertől független Euklideszi-távolságmátrixot használja a CNPI előállítására. A távolságmátrix – illetve az annak megfelelő gráf – automorfizmus csoportja ugyanis pont a molekula teljes magpermutációs csoportjával egyezik meg. Ennek előállításához nagyszámú permutáció végrehajtására van szükség. Az automorfizmus csoport felírására több algoritmus létezik, valamint a módszer egyéb eszközök felhasználásával is felgyorsítható. A kapott eredmény viszont egyáltalán nem szemléletes és sok esetben nem is szükséges a teljes magpermutációs csoport előállítása, amely nagyméretű molekulák esetén a nagyszámú permutáció miatt igen időigényessé is válhat.

A Pilati és Forni által közölt *symmol*<sup>43–46</sup> szoftver a főtehetetlenségi momentumokat használja ki a molekulaszimmetria felismerésének megkönnyítésére. Egy standard orientáció megállapítása után a program valamelyik tengely mentén felépíti a forgatási mátrixot, majd a megfelelő forgatásokat végrehajtva igyekszik azonosítani a valódi forgástengelyeket. A potenciális szimmetriaelemek analízise során az atomi toleranciaszintek különféleképpen súlyozhatók. Ez a kód egyébként – a pontcsoport és a szimmetriaelemek mellett – a szimmetrizált koordinátákat is meghatározza.

Egy újabb módszer<sup>47</sup> a szimmetriaelemek detektálása közben szintén a főtehetetlenségi momentumokra szorítkozik. Ezen algoritmus arra a tényre van alapozva, hogy a szimmetriaelemek irány-, illetve normálvektorai a tehetetlenségi tenzor sajátvektorai, ugyanis ez a tenzor a szimmetriaműveletekre nézve – az Euklideszi-távolságmátrixhoz hasonlóan – invariáns.

Míg a legmagasabb rendű forgástengely a főtengelyek körüli forgatásokkal képezhető, addig a többi tengely a tömegközépponton, illetve vagy egy atomon vagy pedig egy olyan egyenes mentén fekszik, amely a két azonos atomot összekötő szakasz felezőpontján halad át. A szabályos csoportok (O, I, T) azonosítása további lépéseket igényel, melyekhez a  $C_2$  tengelyek irányvektorai vagy azok vektori szorzatai körüli forgatásokat kell elvégeznünk. A szimmetriasíkok kijelölése a szimmetriatengelyekre merőleges síkok meghatározásával kezdődik. A további lehetséges síkok valamely atomot és a tömegközéppontot tartalmazzák, vagy két azonos rendszámú atom felezőpontját foglalják magukban. Ezek után a lehetséges pontcsoportokat a megtalált szimmetriaelemek és az egyes pontcsoportok szimmetriaelemeinek összehasonlításával azonosítjuk.

Az ismertett algoritmust megvalósító program a pontcsoport megadását követően képes a geometria szimmetrizálására is. Ha az egyedi atomokra, melyek egy-egy szimmetriaelemen található, az összes szimmetriaműveletet végrehajtjuk, akkor ezáltal a szimmetriának megfelelő, pontos geometriához juthatunk.

### 2.1.2.2. Gráfelméleti módszerek

Az eddig ismertett módszerek mátrixalgebrai alapon detektálták egy molekula szimmetriáját, forgató- és tükrözőmátrixokat vagy a molekulagráf egy mátrixreprezen-

tációját vették igénybe a feladat megoldása során. Léteznek azonban olyan módszerek is, melyek főként gráfelméleti eszközöket alkalmaznak a szimmetria felismerésére.

Ezek a módszerek a szimmetriaanalízist a molekulagráf automorfizmusainak segítségével végzik el, melyek nagyobb molekulák esetén kombinatorikus robbanáshoz és a feladat számítógépes megoldhatatlanságához vezetnek. A gráfelméleti megközelítés azonban egyéb problémákat is felvet, hiszen a molekulagráf csak a molekulát alkotó atomok konstitúciójáról ad információt, a háromdimenziós szerkezeti eltérésekről viszont nem képes számot adni. Egyértelmű tehát, hogy a pusztán konnektivitásra hagyatkozó megfontolások nem alkalmasak a molekula statikus szimmetriájának meghatározására. Az automorfizmusokból ezenfelül csak a megfelelő permutációk nyerhetők, a szimmetriaelemek típusa pedig más módszerek felhasználásával pontosítható.

Az egyes eljárások leginkább az automorfizmus csoport előállításának módjában térnek el egymástól. A konstitúciós szempontból megkülönböztethetetlen atomok orbitokra osztják a molekulagráfot. Ezek az orbitok definíció szerint azon csúcsok halmazai, melyek automorfizmussal egymásba transzformálhatók. A kezdeti módszerek<sup>48</sup> az orbitok összes elemének permutációit vizsgálták, ez azonban az említett kombinatorikus robbanás miatt csak kis molekulák esetén alkalmazható.

A későbbi módszerek<sup>49–51</sup> általában a gráf egy speciális rendezésére építenek, illetve igyekeznek tovább korlátozni a lehetséges permutációk számát a sebesség növelése érdekében. A rendezés megválasztása azonban problémás, ugyanis a nem megfelelő rendezés a szerkezet nem egyértelmű leképezéséhez vezethet, ami olyan molekulákat tesz megkülönböztethetetlenné, amelyek a valóságban eltérnek egymástól.<sup>49</sup>

Egy lehetséges rendezés a molekulát alkotó atomok konnektivitását veszi alapul.<sup>50,51</sup> Ennél a módszernél a gráf csúcsainak számozása a legkevesebb szomszédal rendelkező csúcsnál kezdődik, majd minden lépésben a legutóbbi csúcs legkisebb konnektivitású szomszédjával folytatódik. Amennyiben már egy csúcs összes szomszédja szerepel a permutációban, az utolsó nem vizsgált szomszédal rendelkező csúcsnál folytatódik a keresés. Ez a számozás azonban még nem egyértelmű, így egy olyan konvenciót kell bevezetnünk, mely az ekvivalens számozások közül mindig ugyanazt a permutációt választja ki, amely alapján egy kétdimenziós szerkezetet determináló „név” generálható. A szerkezet pontos reprezentálásához viszont az így előállított karakterláncnak a háromdimenziós szerkezetből származó információkat is tartalmaznia kell, melyekből a szimmetriaelemek és azok típusa a felírt karakterlánc alapján azonosítható.

Elképzelhető a szerkezet más reprezentációjának vizsgálata is. A szimmetria meghatározható egy a molekulára jellemző súlyozott gráf alapján is. Ilyen gráf lehet például az Euklideszi-távolságmátrix mint súlymátrix által reprezentált gráf.<sup>52</sup> Ezen a struktúrán a molekula teljes magpermutációs csoportja előállítható, amely (a) sokkal nagyobb, mint a pontcsoport, (b) sok esetben felesleges információkat hordoz, illetve (c) előállítható a vizsgált pontcsoport felhasználásával is.

## 2.2. Ab initio termokémiai modellek

### 2.2.1. Kompozit kvantumkémiai módszerek

Az utóbbi évtizedekben számos új, kompozit kvantumkémiai módszert dolgoztak ki, melyek alkalmazásával nagyfokú pontosságnövekedés érhető el, amely alkalmassá teszi azokat termokémiai számítások végrehajtására. Az említett protokollok célja a kémiai pontosság ( $1 \text{ kcal mol}^{-1}$ ) elérése a számításigény csökkentése mellett. A kompozit módszerek elterjedése a kereskedelmi szoftverekben azok „fekete dobozként” való implementációjával magyarázható.

A kompozit modellek alapja az, hogy az egyes energiajárulékok különböző elméleti szinten konvergálnak, lehetővé téve kisebb pontosságú, de hatékonyabb számítási lépések igénybevételét egyetlen pontos, de időigényes módszer használata helyett. Az így kapott energiakorrekciókat ezután additívnak tekintve a teljes energia egyszerűen azok összegeként írható fel. Ez az additivitás azonban nem minden esetben áll fenn,<sup>53</sup> illetve a megfelelő pontosság elérése érdekében egyes modellek esetén empirikus korrekciós tagok belefoglalása is szükséges. Az empirikus paraméterek elméleti megfontolások alapján<sup>54</sup> vagy gyakrabban kísérleti adatokra történő illesztéssel becsülhetők. Amennyiben egy adott molekula egyes konformerei közötti energiakülönbséget számítjuk, a korrekciók, amelyek a  $G_n$ -protokollcsalád esetén csak a vegyértékelektronok számától függenek, kiejtik egymást a végső mennyiségben. Ennek eredményeképp konformációs energiakülönbségek esetén még az empirikus paramétereket alkalmazó  $G_n$  modelleknél is *ab initio* eredményekről beszélhetünk.

Az elektronenergia javítása mellett természetesen további korrekciók figyelembevételére is van lehetőség. A teljes energia tovább pontosítható például a Born–Oppenheimer közelítés diagonális korrekciójával, skaláris vagy többkomponensű relativisztikus korrekciókkal, esetleg a harmonikus zérusponti energia anharmonikus korrekciójával. A megoldás véges bázison történő felírása mind egyedi molekulák esetén, mind több rendszer kölcsönhatásának számításakor problémát okoz. Az előbbi esetben egy további módot a pontosság növelésére az egyes járulékok extrapolációja jelent, mellyel a teljes, végtelen bázison elérhető eredmények közelíthetők. Kölcsönhatási energia számításakor az alrendszerek bázisának szuperpozíciója további hibát eredményez (*basis set superposition error*, BSSE), melyet többféle módon korrigálhatunk. Egyes termokémiai modellek az előbb említett korrekciók némelyikét is tartalmazhatják.

Belátható azonban, hogy az elektronszerkezet leírásának bonyolultsága miatt nem adható meg egy általános, minden rendszer kezelésére alkalmas modell. Az egyes molekulák esetén különböző energiajárulékok dominálnak, ami csak egy bizonyos, egymáshoz hasonló molekulákból álló halmazra teszi lehetővé egy-egy ilyen módszer kidolgozását. Ha a számításokat olyan molekulákra szeretnénk végezni, amelyek nem

hasonlíttanak a modellalkotás során figyelembe vett szpécieszekhez, akkor az eredmény pontatlan, nagyobb elhanyagolások esetén alapvetően hibás lehet. Ez magyarázza a nagyszámú kompozit módszer megjelenését.

### 2.2.1.1. Gaussian- $n$ és CBS- $n$ modellek

A legelső *ab initio* termokémiai módszereknek a  $G_n$  (Gaussian- $n$ )<sup>53–56</sup> és a CBS- $n$ <sup>57–63</sup> modellek tekinthetők. A cél kezdetben különféle termokémiai mennyiségek (pl. disszociációs energiák, reakcióhők, képződési entalpiák) 2 kcal mol<sup>-1</sup> pontossággal történő becslése volt. Ezek a modellek megalkotásuk óta számos változtatáson estek át. A legújabb Gaussian modell, a G4-módszer, a vizsgált tesztalmazra 0.83 kcal mol<sup>-1</sup> szórást eredményezett, elérve ezzel a kémiai pontosságot. A módosított változatok leginkább az alkalmazott módszerekben, illetve a referencia geometriában térnek el egymástól. A  $G_n$  módszerek esetén például a korábbi modellekben alkalmazott QCI (*quadratic configuration interaction*, kvadratikus konfigurációs kölcsönhatás) számításokat modernebb csatolt klaszter (*coupled cluster*, CC) számításokkal váltották ki, valamint a G4-módszer korrekciós tagjainak számítása a korábbi másodrendű Møller-Plessett (MP2) módszerrel előállított szerkezettel szemben már B3LYP sűrűségfunkcionál módszerrel optimalizált geometriákon történik. Bár a B3LYP geometria általában jól egyezik pontosabb *ab initio* számítási vagy kísérleti eredményekkel, a sűrűségfunkcionál módszerek eredménye sajnos nem minden esetben megbízható.<sup>64,65</sup>

Megfelelő, szisztematikusan változtatott bázissal az energia a végtelen bázison számított energiához konvergál. Ez alapján a végtelen bázison vett teljes energia különféle extrapolációs formulák<sup>57–60,66–75</sup> segítségével közelíthető, így kiküszöbölve a véges bázisból származó hibát. Ezt az eljárást alkalmazzák a CBS (*complete basis set*) módszerek is, melyek az egyszerűbb korrelációs járulékokat, illetve olykor a referencia HF energiát extrapolációval határozzák meg, valamint a G4 módszer, amelynél az extrapoláció a Hatree-Fock-limitre történik.

### 2.2.1.2. Weizmann- $n$ és HEAT modellek

A további modellek kidolgozása során már nem használtak empirikus paramétereket, így ezek az eljárások a kvantumkémiai paramétereket tisztán *ab initio* számítások segítségével állítják elő.

Az első Weizmann modellek<sup>76</sup> kidolgozása során a cél az 1 kJ mol<sup>-1</sup> átlagos pontosság elérése volt, mégpedig oly módon, hogy a hiba a legrosszabb esetben se haladja meg a 4 kJ mol<sup>-1</sup> értéket. A később megalkotott modellek<sup>77</sup> esetén, melyek már csatolt klaszter számításokon alapulnak, a még nagyobb, szub-kJ mol<sup>-1</sup> pontosságot tűzték ki célul. A legpontosabb Weizmann-4 protokollban<sup>78</sup> az energiakorrekciók CCSD(T) referenciageometriára vonatkoznak, míg a legmagasabb szintű számítást CCSDTQP tech-

nikával valósítják meg. A modell diagonális Born–Oppenheimer-korrekción (DBOC) és a relativisztikus effektusok becslésére alkalmas Douglas–Kroll–Hess (DKH) módszerrel<sup>79–81</sup> nyert hozzájárulást is tartalmaz. Habár a Weizmann- $n$  modellek rendkívül pontos eredményeket szolgáltatnak, azok nagy számítási kapacitást és jelentős gépidőt igényelnek, így legfeljebb néhány (4–5) nehéz atomot tartalmazó rendszerekre alkalmazhatók.

A HEAT eljárás<sup>82,83</sup> hasonló lépésekből épül fel, azonban fontos felhívunk a figyelmet arra, hogy amíg a Weizmann- $n$  modellek a korrelációs járulékokat a befagyasztott mag közelítés keretein belül határozták meg, addig a HEAT protokoll az összes vegyérték- és törzselektront figyelembe veszi a korrelációs energiatagok becslése során. Ez általában kis mértékű (kevesebb, mint  $0.5 \text{ kJ mol}^{-1}$ ) hibát okoz, ugyanakkor a szub-kJ  $\text{mol}^{-1}$  pontosság eléréséhez nagymértékben hozzájárulhat.

### 2.2.1.3. További modellek

Az eddigi kompozit módszereken kívül számos más termokémiai modellt is kidolgoztak, melyek közül néhányról érdemes itt is szót ejtenünk. Az egyik megközelítést a DeYonker és munkatársai által megalkotott ccCA (*correlation consistent composite approach*) eljárások képviselik.<sup>84–90</sup> A ccCA modelleket a  $G_n$  protokollokhoz hasonlóan építették fel, ezek a modellek azonban nem tartalmaznak empirikus korrekciós tagokat. A ccCA módszerek esetén (a) a zérusponti rezgési energiát (ZPVE) B3LYP geometrián számítják a frekvenciák skálázásával, (b) a korrelációs tagokat eltérő bázison meghatározott  $MP_n(\text{FC})$  ( $n = 2, 4$ ) energiákból nyerik, illetve (c) olykor a skaláris relativisztikus korrekciót és a spin-pálya csatolásból származó hozzájárulást is tekintetbe veszik. Ezek a modellek viszonylag kis számításigényük ellenére a  $G_n$  modellekéhez hasonló teljesítőképességgel rendelkeznek, így azok közepes méretű molekulák termokémiai paramétereinek becslésére is sikerrel alkalmazhatók.

Egy másik fejlesztési irányt a Truhlar és munkatársai által fejlesztett MCCM (*multicoefficient correlation method*) módszerek képviselik.<sup>91–97</sup> Ezekben a protokollokban az eredményt az egyes hozzájárulásokra bevezetett empirikus skálázó faktorok segítségével igyekeznek javítani.

Feller és munkatársai szintén több kompozit eljárást közöltek,<sup>98–103</sup> melyek CC számításokra épülnek és a CBS határ elérésére különféle extrapolációs formulákat vetnek be. Ami a kisebb járulékokat illeti, a skaláris relativisztikus korrekciót DKH CC számításokból származtatják, emellett a Born–Oppenheimer-közelítés diagonális korrekcióját, a spin-pálya csatolást és a zérusponti energia anharmonikus korrekcióját is belefoglalják ezekbe a protokollokba.



### 2.2.2. Fókuszpont analízis

A fókuszpont analízis (*focal point analysis*, FPA)<sup>104,105</sup> – az eddigiektől eltérően – nem egy konkrét lépésekből felépülő protokollt ad meg, hanem egy olyan általános sémának felel meg, mely a számítási lépéssorozat kidolgozását a felhasználóra bízta. Mivel az FPA séma elemei a feladattól függően változhatnak, ez a módszer lehetővé teszi az eredmények pontosságának javítását a szükséges mértékben vagy a számítási kapacitásnak megfelelően. Természetesen ez azt is jelenti, hogy az FPA eljárás a korai, kereskedelmi szoftverekben implementált modellektől eltérően nem használható „fekete dobozként”. Mindazonáltal, ha nagy pontosságú adatokra van szükségünk, a nagyobb felhasználói erőfeszítés mindenképpen megtérül, amit a különböző rendszerekre szolgáltatott referenciaértékű becslések is jól mutatnak.<sup>106–108</sup> Ezen felül az FPA megközelítés nagy előnye, hogy az egyes járulékokat azok konvergenciájának mértéke alapján jól meghatározott bizonytalanságokkal látja el, lehetővé téve a származtatott mennyiségek pontosságának becslését is.

Egy FPA analízis során a cél az eredmény szisztematikus javítása, melyet – a korábbiakhoz hasonlóan – additívnak tekintett hozzájárulások összegeként állítunk elő. Az FPA filozófia értelmében az egy- és többreszecske problémák konvergenciáját egy két-dimenziós extrapolációs rácson térképezzük fel, majd azok kezeléséhez extrapolációs formulákat veszünk igénybe. Ehhez szisztematikusán növekvő, pontos extrapolációt biztosító bázisokra van szükség, mely követelménynek a Dunning-féle korrelációkonzisztens bázisok<sup>109–115</sup> mindenképpen eleget tesznek. Az FPA modellezés során további kisebb korrekciók (Born–Oppenheimer közelítés diagonális korrekciója, skaláris relativisztikus korrekciók, a spin-pálya csatolás, illetve a harmonikus zérusponthoz anharmonikus korrekciója) figyelembevételére is lehetőség nyílik.

## 2.3. QSPR modellezés

A kvantitatív szerkezet-aktivitás és szerkezet-tulajdonság összefüggés vizsgálatok fontos szerepet töltenek be a modern kémiában, biokémiában és a farmakológiában. Egy jól megválasztott QSAR/QSPR modell tetszőleges számú vegyület szűrését lehetővé teszi, beleértve a még elő nem állított vegyületeket is, így a releváns molekulák megtalálása jelentősen felgyorsítható. Mivel a szintetizálni kívánt vegyületek száma nagymértékben csökkenthető, ez a módszer egy kutatás vegyszerigényét és költségét lényegesen visszaszoríthatja. A QSAR/QSPR modelleket egyre gyakrabban alkalmazzák például a gyógyszertervezésben, a toxikológiai vizsgálatok során vagy bizonyos vegyületek fizikai-kémiai tulajdonságainak becslésére.

Egy megfelelő empirikus modell felállítása azonban több nehézségbe is ütközik. Az egyik kritikus pontot az ún. deskriptorok kiválasztása jelenti. A deskriptorok olyan

elméleti vagy kísérleti paraméterek, melyek jól korrelálnak a modellezni kívánt vegyületek tulajdonságaival. Egyszerűbb esetben a deskriptorok a molekula egy jól ismert tulajdonságát tükrözik, mint pl. a víz- $n$ -oktanol megoszlási hányados is, melyet az aktivitás vizsgálatok során előszeretettel használnak fel. Az elméleti deskriptorok közül a kvantumkémiai deskriptorokat érdemes kiemelni, melyeket a kémiai informatika és a számítási kapacitás fejlődésének köszönhetően QSAR/QSPR modellek felállításakor, mind önmagukban, mind más, hagyományos deskriptorokkal kombinálva egyre gyakrabban alkalmaznak.

A kvantumkémiai mennyiségek nemcsak deskriptorként használhatók, hanem – az azokból származtatott termodinamikai paraméterekhez hasonlóan – a függő változó szerepét is betölthetik egy esetleges QSPR modellezés során.<sup>116–118</sup> Az ilyesfajta modellalkotás igen fontos szerepet kaphat azokban az esetekben, amikor a nagy pontosságú kvantumkémiai adatok rendkívül időigényes számításokból származnak, lehetővé téve egy kis számítású, megbízható modell felállítását. Gyakran tanulmányozott tulajdonság a kísérleti és elméleti módszerekkel egyaránt meghatározható képződési entalpia és entrópia is, melyek leírására számos deskriptor (elektrotopológiai és Kier–Hall indexek, a forgatható kötések száma,<sup>117</sup> atomi összetétel,<sup>118</sup> atom-kötés konnektivitási index (*atom-bond connectivity*, ABC)<sup>119</sup>) alkalmasnak bizonyult.

Bizonyos mennyiségek predikciójához azonban háromdimenziós deskriptorok alkalmazása is szükséges.<sup>120</sup> Az viszont nem nyilvánvaló, hogy ezek a deskriptorok a szerkezet milyen pontos ismeretét igénylik, mivel azok minősége, illetve információartalma a felhasznált geometriától nagymértékben függhet.<sup>120</sup>

### 2.3.1. Topológiai indexek

Bár az elmúlt évtizedekben több ezer deskriptort írtak fel, azok fejlesztése még napjainkban is folyik. Habár a kezdeti empirikus leíró paraméterek, mint például a szubsztituenstől függő Hammett-állandó, a víz-oktanol megoszlási hányados vagy az Ostwald-féle oldhatósági együttható jó néhány szerves reakciómechanizmus megértéséhez jelentős mértékben hozzájárultak,<sup>121</sup> ezek sokszor olyan bonyolult fizikai-kémiai kölcsönhatásokat reprezentálnak, melyek értelmezése esetenként eléggé körülményes. Tovább bonyolítja a modellezést az is, hogy az ilyen tapasztalati deskriptorok csak olyan vegyületek esetén használhatók fel, melyekre a szükséges kísérleti adatok már rendelkezésre állnak.

Ezek a problémák vezettek az elméleti molekuláris deskriptorok megszületéséhez, melyek a molekulák különböző reprezentációiból vezethetők le. Ebbe a csoportba tartoznak például az összetételre épülő deskriptorok (0D), a molekulafragmens listák (1D), a gráfelméleti leírók (2D), kvantumkémiai-, WHIM-,<sup>122,123</sup> és GETAWAY-deszkriptorok,<sup>124</sup> a molekulafelszín vagy -térfogat (3D), ahol  $nD$  a molekulareprezentá-

ció dimenzióját adja meg, de „magasabb dimenziós” jellemzők definiálása is lehetséges.

Az elméleti deszkriptorokkal szembeni legfontosabb követelményként az invarianciát támaszthatjuk. Legegyszerűbb esetben ez a molekulát alkotó atomok számozásától való függetlenséget jelenti, de a 3D deszkriptoroktól kezdődően a translációs és rotációs invariancia is lényeges feltétel. Ideális esetben a deszkriptor egyszerű, könnyen értelmezhető és jól korrelál legalább egy tulajdonsággal. Mindazonáltal fontos szempont a deszkriptor esetleges degeneráltsága is, melynek megjelenése esetén a szóban forgó molekula-jellemző bizonyos szerkezetek megkülönböztetésére alkalmatlanná válik.

A fenti követelményeknek az ún. gráfinvariánsok eleget tesznek, melyek az utóbbi időben egyre növekvő népszerűségnek örvendenek. Ezek a szerkezeti jellemzők a molekulagráfából vezethetők le, melyben a csúcsok az egyes atomokat, az élek pedig a kémiai kötésekét reprezentálják. Ha csak az atomok konnektivitását vesszük figyelembe, topológiai indexekről beszélünk, melyek közül a kezdeti, egyszerűbb gráfelméleti deszkriptorokat érdemes kiemelni (Randić-index,<sup>125</sup> a Hosoya-féle  $Z$  index,<sup>126</sup> a Balaban-féle  $J$  index<sup>127</sup> vagy a Wiener-index<sup>128</sup>). Ezek az indexek jó néhány vegyületcsoport (pl. alkánok, alkanolok, halogénezett szénhidrogének) fizikai tulajdonságainak modellezésében<sup>119,129–131</sup> igencsak ígéretes eredményeket szolgáltattak.

### 2.3.2. Topografikus indexek

A tisztán topológiai alapú indexek komoly hiányosságoktól szenvednek, ugyanis kizárólag a konnektivitásról szolgálnak információval, ezáltal csak olyan molekulák (például homológ sorok) tulajdonságainak leírására alkalmasak, melyek háromdimenziós szerkezete hasonló és a szerkezet topológiája az egyedüli meghatározó tényező. Amennyiben a molekula heteroatomokat is tartalmaz, úgy ezeket az indexeket oly módon kell kiterjeszteni, hogy képesek legyenek az egyes atomok és azok kémiai környezeté közötti különbségekről is számot adni. Bizonyos esetekben a térbeli szerkezetre utaló információk felhasználására is szükség lehet,<sup>132</sup> gondolhatunk itt pl. a molekuláris konformerek modellezésére, melyek egymástól csak a háromdimenziós szerkezetük tekintetében mutatnak különbséget.

A csúcsok (atomok) és az élek (kötések, geometriai paraméterek) megkülönböztetése a gráf súlyozásával valósítható meg, amely információkat a különféle gráfelméleti mátrixok (főként az adjacenciamátrix és a távolságmátrix) elemeinek módosításával a kérdéses mátrixokból származtatott deszkriptorokban is kifejezésre juttathatjuk. Míg az egyes atomokra vonatkozó fizikai-kémiai jellemzőket általában a diagonális elemekben jelenítik meg, addig a térszerkezeti hatásokat az élek súlyainak (általános esetben a megfelelő mátrixok off-diagonális elemeinek) paraméterezésével veszik figyelembe. Abban az esetben, ha a topológiai távolságokat valamilyen egyéb távolságokkal helyettesítjük, akkor az így nyert deszkriptorokat topografikus deszkriptoroknak nevezhetjük.

Ezeknek a deszkriptoroknak az előállítására általában valóban topografikus, euklideszi távolságokat alkalmaznak.

A gráfelméleti mátrixok paraméterezésére molekulaszervezeti adatok (kötéshosszak, kötésszögek, torziós szögek) és az egyes atomokra jellemző mennyiségek (rendszám, vegyértékelektronok száma) egyaránt felhasználhatók. Az utóbbiak segítségével az ún. elektrotopológiai indexek definiálhatók,<sup>133,134</sup> melyek topológiai és elektronszerkezeti sajátosságok figyelembevételével állíthatók elő. Ezek szigorú értelemben nem nevezhetők topografikus indexeknek, mivel nem az atomok euklideszi távolságát foglalják magukba, ennek ellenére jelentős előrelépést jelentenek a topológiai indexekhez képest, ugyanis az egyéb adatok felhasználása a molekuláris deszkriptorok teljesítőképességét nagymértékben növeli. Az elektrotopológiai indexeket a legkülönbélebb QSAR/QSPR modellekben alkalmazták: a modellezett tulajdonságok között szerepel a vízdoldhatóság,<sup>135,136</sup> a forráspont, a kritikus hőmérséklet<sup>137</sup> és a vér-agy gát permeabilitása,<sup>138</sup> de gyógyszertervezés segítésére irányuló tanulmányok is fellelhetők az irodalomban.<sup>139,140</sup>

A topografikus távolságok használatára először Randić közleményeiben figyelhetünk fel,<sup>141,142</sup> melyekben a szerző a távolságmátrixot egy grafitrácsba ágyazott szénhidrogén-molekulára – egységnyi C–C kötéshosszakat feltételezve – határozta meg. Egy további munkában az ekképp definiált távolságmátrix súlyozásával foglalkoztak.<sup>142</sup>

Bogdanov és munkatársai a Wiener-index három dimenzióba történő kiterjesztését valósították meg.<sup>26</sup> Ebben a tanulmányban azt is kimutatták, hogy a 3D Wiener-index a kis szénatomszámú alkánok képződési entalpiáját – az alkalmazott modellfüggvény alakjától függetlenül – pontosabban írja le.

Az Estrada és munkatársai által fejlesztett deszkriptorok<sup>143–145</sup> alkalmazásával, melyek a molekulagráf adjacenciamátrixának kötésrendek segítségével történő súlyozásán alapulnak, szintén kiemelkedő eredmények érhetők el. Ezeket az indexeket például alkánok forráspontjának, illetve moláris refraktivitásának modellezésében vezették be.

Különbéle megfontolások alapján az eddigieken kívül is számos további topografikus indexet definiáltak.<sup>146–149</sup> Nyilvánvaló, hogy a módosított topológiai indexek, különösen a háromdimenziós szerkezeti információkat és kémiai tulajdonságokat is magukba foglaló topografikus indexek, sokféle, meglehetősen változatos tulajdonsággal korrelálnak, ami az említett deszkriptorok rendkívül sokrétű felhasználását teszi lehetővé.

### 2.3.3. Szimmetria deszkriptorok

A szimmetria alapján definiálható mennyiségek szintén korrelálhatnak a vegyületek bizonyos fizikai-kémiai tulajdonságaival. A szimmetriát leggyakrabban az olvadásponttal hozták összefüggésbe,<sup>150–152</sup> de néhány egyéb sajátosság modellezése során is hasznosnak bizonyult.<sup>153–155</sup>

Egy tulajdonság szimmetriafüggése leírható például a WHIM-deszkriptorok, a Kier-

féle szimmetria index, a forgási szimmetriaszám, a szimmetria faktor, a főtehetetlenségi momentumok, valamint az ezekből származtatott mennyiségek segítségével.<sup>156</sup>

Az irodalomban továbbá egy ún. szimmetriára vonatkozó információs indexet ( $I_{\text{SYM}} = N \log_2 N - \sum_{g=1}^G [N_g \log_2 N_g]$ ) is számon tartanak, ez azonban azonos a Kier-féle szimmetria indexszel ( ${}^0\kappa = -N \sum_{g=1}^G [N_g/N \log_2 (N_g/N)]$ ), ahol  $N$  az atomok száma,  $G$  a topológiailag ekvivalens atomokból alkotott osztályok halmazának mérete,  $N_g$  pedig a  $g$ -ik osztályba tartozó atomok száma).

Érdemes szót ejteni arról is, hogy a Kier-féle szimmetria index és a szimmetriaszám különböző vegyületek gőznyomásának<sup>153</sup> és olvadáspontjának<sup>154,155</sup> modellezésére is jó eredményeket mutat fel.

Korábbi vizsgálatok arra engednek következtetni, hogy egyes termokémiai mennyiségek is korrelálhatnak a molekuláris szimmetriával. Ezt a feltevést az alkánok entalpiakülönbségeinek becslésével foglalkozó tanulmányok is alátámasztják.

### 2.3.4. Általánosan alkalmazott modellfüggvények

A QSAR/QSPR modellezés során különféle modellfüggvényeket<sup>157,158</sup> használnak fel, melyek különböző paraméterbecslési technikákra épülnek. Habár a mögöttes paraméterbecslési feladatok megoldására számos konvencionális módszer létezik, újabban egyre több modern megközelítés lát napvilágot.

A legtöbb QSAR/QSPR modell multilineáris célfüggvényre épül, melyek paramétereit multilineáris regresszióval (*multilinear regression*, MLR) határozzák meg. Az MLR eljárás népszerűsége egyszerűségének és könnyű értelmezhetőségének köszönhető. A multilineáris modellek hátrányaként a változók lehetséges multikollinearitása említendő, mely a becsült paraméterek hibájának növekedését okozhatja, illetve az egyes indexek egyedi hatásának vizsgálatát is értelmetlenné teheti.

A modellépítés során gyakran folyamodnak mesterséges neurális hálózatokhoz (*artificial neural network*, ANN) is. Ezek a hálózatok párhuzamos, elosztott működésre képes információfeldolgozó eszközök, melyek általában nagyszámú, egymással nagymértékben összekapcsolt elemek rendszeréből állnak. Ennek megfelelően természetüknél fogva képesek nemlineáris kapcsolatok leírására, viszont a kapott modell rendkívül nehézkesen értelmezhető: könnyen felléphet az adatok túlillesztése és igencsak problémás feladatnak számít a szignifikáns deskriptorok kiválasztása is.

A multilinearitás problémája a parciális legkisebb négyzetek módszerének (*partial least squares*, PLS) használatával szorítható vissza. Ennek során a függő és független változók közötti kapcsolatot úgy keressük, hogy a változókat először egy altérbe vetítjük, független változóként pedig az eredeti változók (általában kisebb számú) lineáris kombinációit mint látens változókat tekintjük.

### 3. Célkitűzés

Jelen doktori munka célja egy, az elérhető szoftvereknél hatékonyabb, molekulaszimmetria detektálására szolgáló program kifejlesztése, illetve a molekulaszimmetria felhasználása saját, nagy pontosságú *ab initio* kvantumkémiai számítási eredmények QSPR modellezésére.

Míg az előbbi célunk eléréséhez vektoralgebrai eszközökre hagyatkozunk, addig az utóbbi terveinkhez magas szintű kvantumkémiai számítások elvégzésére és komplex statisztikai eszköztár alkalmazására van szükségünk.

A szimmetria meghatározásához tehát egy vektoralgebrai algoritmust építünk fel, melynek előnye, hogy jelentősen egyszerűbb, mint a közkedvelt mátrix- vagy tenzoralgebrai eljárások. A megfelelő transzformációk implementálása után az összes lehetséges szimmetriaelem vizsgálatával egy tetszőleges molekula szimmetriaelemeit előállíthatjuk, melynek ismeretében a molekuláris pontcsoport, valamint számos szimmetriával összefüggő egyéb tulajdonság (pl. vázcsoport, spektroszkópiai tulajdonságok, a szimmetriának megfelelő egzakt magkonfiguráció) megadható. Szeretnénk bemutatni a program alkalmazhatóságát, melyhez az általunk írt kód működését az irodalomban létező szoftverek teljesítőképességével kívánjuk összevetni. A dolgozatban tárgyalt vektoralgebrai programot néhány standard kvantumkémiai programcsomagba integrálni is szándékozunk, igazolva ezzel az újonnan létrehozott program flexibilitását.

A molekulaszimmetriát ezután kvantumkémiai számítási eredmények modellezésére szeretnénk felhasználni, melyekre saját nagy pontosságú, pentánkonformerekre végzett számítási eredményeket kívánunk igénybe venni. A minél nagyobb pontosság elérése a további modellezés során fellépő hibaterjedés kiküszöbölése miatt is jelentős, így a referenciaadatokat egy csatolt klaszter módszeren alapuló fókuszpont analízis modelltől fogjuk származtatni. A megfelelő precizitás elérése érdekében néhány kisebb, de szignifikáns hatású korrekciós tagot (Born–Oppenheimer-közelítés diagonális korrekciója, skaláris relativisztikus korrekciók) is tekintetbe szándékozunk venni.

Végül a kapott termokémiai mennyiségekre szeretnénk egy olyan, nagy pontosságú QSPR modellt felállítani, melyet e mennyiségek hőmérsékletfüggésének leírására használhatunk. E modellezés során változóknak nemlineáris modellfüggvények felhasználását tervezzük, melyekben az irodalomból jól ismert deskriptorokat, valamint azok általunk módosított változatait szerepeltetjük. A létrehozott QSPR modell statisztikai analízise alapján következtetni lehet annak extrapolációs erejére is, így sikeres modellezés esetén a vizsgálatot egyéb molekuláris struktúrákra, például a konformerek izomerizációjának átmeneti állapotaira vagy a bután konformerekre is kiterjeszthetjük.

## 4. Elméleti összefoglaló

### 4.1. Szimmetriaanalízis

Egy molekula pontcsoportjának megállapítása a szimmetriaelemek meghatározásával veszi kezdetét. A molekulaszimmetria felismeréséhez szükséges szimmetriaelemek száma az alkalmazott módszertől függ: míg egy folyamatára végigkövetéséhez általában néhány kiválasztott szimmetriaelem ismerete is elegendő, addig más eljárások a molekula pontcsoportját a detektált szimmetriaelemek pontos száma, illetve azok típusa alapján találják meg.

Az egyes szimmetriaelemek előállítására bizonyos megfontolások alapján definiált műveletek végrehajtásából és a transzformált struktúra eredetivel történő összehasonlításából áll. A vizsgált műveletek hagyományos térbeli forgatások és tükrözések, vagy – gráfelméleti módszerek esetén – az azonos atomok permutációi lehetnek, mely utóbbi esetben az adott szimmetriaelem típusának azonosítása a kérdéses permutáció elemzésével lehetséges.

#### 4.1.1. Szimmetriaműveletek

Egy molekulát absztrakt módon egy  $\langle A, \zeta, \kappa \rangle$  rendezett hármasként adhatunk meg, ahol  $A$  a molekulát alkotó atomok halmaza, a  $\zeta : A \mapsto \mathbb{Z}^+$  leképezés, melynek értékkészlete a pozitív egész számok halmaza ( $\mathbb{Z}^+$ ), az atomok rendszámát adja meg, a  $\kappa : A \mapsto \mathbb{R}^3$  injektív leképezés pedig hozzárendeli az atomokhoz azok koordinátáit egy háromdimenziós euklidészi térben. Természetesen a  $\zeta$  leképezés tetszés szerint módosítható vagy egy harmadik leképezés is bevezethető, mely a molekula halmazelméleti fogalmát izotopológok megkülönböztetésére is alkalmassá teszi.

Az  $\langle A, \zeta, \kappa \rangle$  absztrakt molekula esetén egy, az atomok koordinátáira ható  $\tau$  transzformáció a következő leképezésként írható fel:  $\tau : \mathcal{R}(\kappa) \mapsto \mathbb{R}^3$ , ahol  $\mathcal{R}(\kappa)$  a  $\kappa$  leképezés értékkészletét jelöli. A  $\tau$  leképezést szimmetriaműveletnek nevezzük, ha az nem változtatja meg a molekula magkonfigurációját, azaz a transzformáció alatt az atomok koordinátáinak halmaza és az egyes koordinátákhoz tartozó atomok rendszáma változatlan marad. A definícióból adódóan a  $\tau$  műveletet pontosan akkor tekintjük szimmetriaműveletnek, ha az alábbi feltételek teljesülnek:

$$\mathcal{R}(\tau) = \mathcal{R}(\kappa), \tag{4.1}$$

$$\zeta \circ \kappa^{-1} \circ \tau \circ \kappa = \zeta, \tag{4.2}$$

ahol  $\circ$  a leképezések kompozícióját jelöli és  $\kappa^{-1}$  értelmezési tartománya a  $\kappa$  leképezés értékkészletének felel meg. Ezek a feltételek azt jelentik, hogy  $\tau : \mathcal{R}(\kappa) \leftrightarrow \mathcal{R}(\kappa)$  megfelel egy olyan permutációnak, mely kizárólag az azonos rendszámú atomok koordinátáit cserélheti fel. A 4.1 és a 4.2 egyenletek azt is lehetővé teszik, hogy a szimmetriaanalízis során a potenciális szimmetriaelemekhez rendelt  $\tau$  transzformációk szimmetriaművelet voltát igazolhassuk vagy éppen megcáfolhassuk.

#### 4.1.2. A transzformációra vonatkozó megszorítások

A potenciális szimmetriaműveletek típusa alapján a  $\tau$  transzformációkra további megkötések róhatók ki. Abban az esetben, ha statikus szimmetriát feltételezünk, csupán középpontos tükrözések, síkokra való tükrözések, forgatások, illetve tükrözéses forgatások jöhetnek szóba, melyek definiáló egyenletei és operátorai a  $\tau$  leképezés megszorításával annak egzakt alakját szolgáltatják.

Középpontos tükrözés esetén – amennyiben a koordinátarendszer origóját a molekula tömegközéppontjában vesszük fel – az összes atom helyvektorának additív inverzét kell venni, azaz a

$$\mathbf{p}' = -\mathbf{p} \quad (4.3)$$

egyenlethez jutunk, melyben  $\mathbf{p} \in \mathcal{R}(\kappa)$  egy tetszőleges atom helyvektorát jelöli,  $\mathbf{p}'$  pedig  $\mathbf{p}$  transzformált koordinátáit foglalja magába. A középpontos tükrözés operátorának ( $\hat{\mathbf{I}}$ ) mátrixreprezentációja ( $\mathbf{I}$ ) a következőképp írható fel:

$$\mathbf{I} = -\mathbf{E}_3, \quad (4.4)$$

ahol  $\mathbf{E}_3$  a  $3 \times 3$ -as egységmátrixot szimbolizálja.

Egy olyan tükrözés, melynek tükörsíkja az origót is tartalmazza, a következő összefüggéssel fejezhető ki:

$$\mathbf{p}' = \mathbf{p} - 2 \text{proj}_{\mathbf{n}}(\mathbf{p}), \quad (4.5)$$

melyben  $\mathbf{n}$  a tükörsík (egy konstans faktor erejéig határozatlan) normálvektora,

$$\text{proj}_{\mathbf{n}}(\mathbf{p}) = \frac{\mathbf{p} \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \mathbf{n} \quad (4.6)$$

pedig a  $\mathbf{p}$  vektor  $\mathbf{n}$ -re vett projekciója. A tükrözés mátrixa ( $\mathbf{T}$ ) a

$$\mathbf{T} = \left\{ \delta_{ij} - 2 \frac{n_i n_j}{\mathbf{n} \cdot \mathbf{n}} \right\} \quad (4.7)$$

egyenlettel írható fel, ahol  $\delta_{ij}$  a Kronecker deltát jelenti.

Euler forgatásra vonatkozó tétele alapján egy merev test minden olyan transzformációja, mely a test egy pontját érintetlenül hagyja, egyetlen forgatással is leírható,



ha a forgástengelyt úgy választjuk meg, hogy az a rögzített ponton áthaladjon. A háromdimenziós forgatások, melyek az Euler-tétel értelmében három szabadsági fokkal rendelkeznek, többféle formalizmus alapján is megadhatók, az egyes reprezentációk azonban, melyek tetszés szerint egymásba transzformálhatók, háromnál több paramétert is magukba foglalhatnak.

Egy merev test orientációját leggyakrabban az ún. forgatási mátrixszal ( $\mathbf{F}$ ) írják le:

$$\mathbf{F} = [\mathbf{u} \ \mathbf{v} \ \mathbf{w}], \quad (4.8)$$

ahol a testhez rögzített  $\mathbf{u}$ ,  $\mathbf{v}$  és  $\mathbf{w}$  egységvektorok egy jobbsodrású koordinátarendszert feszítenek ki. Az  $\mathbf{F}$  mátrix, melynek konkrét alakjától itt eltekintünk, az adott forgatásra jellemző három Euler-szög ismeretében egyértelműen előállítható.

A forgatások ún. kvaterniókkal is reprezentálhatók, melyek normált négydimenziós vektoroknak felelnek meg. Ezek a kvaterniók számos előnyös tulajdonsággal rendelkeznek a forgatási mátrixhoz képest (pl. a szukcesszív forgatások végrehajtását lényegesen egyszerűbbé teszik).

Sajnos az eddig ismertett megközelítések egy molekula szimmetriaanalízise során eléggé nehézkesen alkalmazhatók, ezért érdemes egy olyan reprezentációt keresnünk, mely pusztán vektoralgebrai eszközökre hagyatkozik. Az ún. tengely-szög formalizmus keretein belül a forgatást az

$$\mathbf{r} = \varphi \mathbf{d} \quad (4.9)$$

vektorral fejezzük ki, melyben  $\mathbf{d}$  a forgástengely normált (két szabadsági fokú) irányvektora és a forgatás szöge ( $\varphi$ ) szerepel. Fel kell hívnunk a figyelmet arra is, hogy ha a forgatás zérus szöggel történik, akkor a forgástengely nem választható meg egyértelműen. Habár ebben a reprezentációban az egymás utáni forgatások végrehajtása kétségkívül sokkal bonyolultabb feladat, erre egy molekula szimmetriájának detektálása során nincs szükség.

A  $\mathbf{p}$  helyvektorú atom forgatása a Rodrigues-formula<sup>159</sup> alapján vektoriális szorzás ( $\times$ ) segítségével kivitelezhető:

$$\mathbf{p}' = \mathbf{p} + \sin \varphi \mathbf{d} \times \mathbf{p} + (1 - \cos \varphi) \mathbf{d} \times (\mathbf{d} \times \mathbf{p}), \quad (4.10)$$

melyhez  $\mathbf{d}$  és  $\varphi$  a molekula magkonfigurációjából egyszerűen meghatározható (ld. később).

A tükrözéses forgatás egy tengely körüli forgatás (v.ö. 4.10 egyenlet) és az  $\mathbf{n} = \mathbf{d}$  normálvektorú (merőleges) síkra való tükrözés (ld. 4.5 egyenlet) egymás utáni elvégzésével valósítható meg. (Természetesen, mivel a tükrözés és a forgatás független műveletek, azok a tükrözéses forgatás generálásakor tetszőleges sorrendben végrehajthatók.)

Mindezek alapján megállapíthatjuk, hogy – a tengely-szög reprezentáció alkalma-

zása esetén – a szimmetriaelemek detektálásához csupán a potenciális szimmetriasíkok és szimmetriatengelyek  $\mathbf{n}$  és  $\mathbf{d}$  vektoraira van szükségünk. Miután tehát az irányvektorokat és a normálvektorokat sikerült előállítanunk, a 4.1 és a 4.2 egyenletekre támaszkodva egy tetszőleges molekula szimmetriaelemeit megadhatjuk.

Amennyiben a molekula nem planáris, úgy a szimmetriasíkok két olyan atomot transzformálnak egymásba, melyek a síkon kívül helyezkednek el, azaz a kérdéses sík normálvektora két azonos rendszámú atom helyvektorából származtatható. Ebből adódóan a lehetséges szimmetriasíkok a

$$\{\langle a, b \rangle \in A^2 \mid \zeta(a) = \zeta(b)\} \mapsto \sigma_p \subset \mathbb{R}^3 \quad (4.11)$$

leképezésnek felelnek meg, ahol a  $p$  alsó index a potenciális szimmetriaelemeket jelöli.

$C_2$  és  $S_4$  tengelyek esetén az előzőhöz hasonló helyzettel állunk szemben. E műveletek hatására két-két atom cserél helyet vagy a kiszemelt atom az adott tengelyen fekszik, ezért a potenciális  $C_2$  és  $S_4$  tengelyek irányvektorára a

$$\{\langle a, b \rangle \in A^2 \mid \zeta(a) = \zeta(b)\} \mapsto C_{2,p} \subset \mathbb{R}^3 \quad (4.12)$$

leképezés nyerhető, melynek értelmezési tartománya megegyezik a 4.11 leképezésével.

Tekintve, hogy magasabbrendű szimmetriatengelyek létezése esetén az egymásba transzformálódó atomok egy szabályos sokszög csúcsainak felelnek meg, a három- vagy többfogatú szimmetriatengelyekre vonatkozó leképezést a

$$\{\langle a, b, c \rangle \in A^3 \mid \zeta(a) = \zeta(b) = \zeta(c)\} \mapsto C_{3,p} \subset \mathbb{R}^3 \quad (4.13)$$

formában írhatjuk, vagyis az egyes tengelyek irányvektorait három atomi pozíció felhasználásával nyerjük.

Az összes potenciális  $C_n$  ( $S_{2n}$ ) forgástengely irányvektorait a 4.12 és 4.13 leképezések együttesen adják meg, másszóval a  $C_n$  ( $S_{2n}$ ) tengelyek általánosan a

$$C_p = C_{2,p} \cup C_{3,p} \quad (4.14)$$

halmazzal fejezhető ki, míg az összes potenciális szimmetriaelem irány-, illetve normálvektorát az

$$E_p = \sigma_p \cup C_p \quad (4.15)$$

halmaz tartalmazza.

## 4.2. FPA modell entalpiakülönbségek becslésére

Amint arról a 2.2. fejezetben részletesen is beszámoltunk, a termokémiai mennyiségek lehető legpontosabb számítását az FPA séma teszi lehetővé. Ezt a megállapítást egy, az *n*-bután konformációs termodinamikájával foglalkozó tanulmány<sup>160</sup> is igazolja, melyben a szerzők egy magas szintű *ab initio* kvantumkémiai módszerekre épülő FPA modell segítségével rendkívüli precizitású ( $\pm 10$  cal mol<sup>-1</sup>) termokémiai adatokat javasoltak. Joggal várhatjuk tehát, hogy egy ilyesfajta kvantumkémiai protokoll az *n*-pentán konformációs energetikájának leírásában is sikerrel alkalmazható.

Az általunk felépített FPA számításorozatot, melynek szerkezete a szub-kJ mol<sup>-1</sup> bizonytalanság elérésére képes *Wn* és HEAT modellekéhez hasonló, CC számításokra alapoztuk, melyeket további korrekciókkal (adiabatikus energiatag, anharmonikus rezgési korrekció, gátolt belső rotációs hozzájárulás) egészítettünk ki. A nagyfokú pontosság érdekében a véges bázisú kezelésből adódó hibát a teljes bázisra történő extrapolációval igyekeztünk csökkenteni. Az alkalmazott FPA modellben szereplő csatolt klaszter energiákban egyszeres [S], kétszeres [D], háromszoros [(T)], illetve négyszeres [(Q)] gerjesztéseket vettük figyelembe, ahol a kerek zárójel az adott gerjesztés perturbatív kezelését jelöli. Fontos azt is hangsúlyoznunk, hogy a poszt-Hartree–Fock számításokat a CCSDT és a CCSDT(Q) energiák kivételével az összes elektron korrelációját figyelembe véve, azaz a befagyasztott mag közelítés mellőzésével végeztük el. A továbbiakban a kérdéses FPA modellben foglalt mennyiségek számításával fogunk foglalkozni.

### 4.2.1. Termokémiai mennyiségek számítása

Egy  $\alpha\beta \in \{tt, tg, gg, gx\}$  torziós szekvenciájú *n*-pentán konformer standard (moláris) entalpiája *T* hőmérsékleten a *tt* globális minimumra vonatkoztatva a következőképp definiálható:

$$\Delta H_T^\circ(\alpha\beta) = H_T^\circ(\alpha\beta) - H_T^\circ(tt), \quad (4.16)$$

ahol a  $H_T^\circ(\alpha\beta)$  entalpia – ideális gázok esetén – a standard belső energia ( $U_T^\circ = U_T^\circ(\alpha\beta)$ ) és az *RT* termikus energia összegeként írható fel, melyben *R* az egyetemes gázállandót jelöli.

A belső energia statisztikus termodinamikai értelemben az

$$U_T^\circ = U_0^\circ + U_T^{\text{kor}} \quad (4.17)$$

alakban fejezhető ki, ahol  $U_0^\circ$  az abszolút nulla fokon vett belső energia,  $U_T^{\text{kor}}$  pedig

a belső energia hőmérsékleti korrekciója. Míg az  $U_0^\circ$  mennyiséget a kvantumkémia-  
ilag számított teljes elektronenergia ( $E_{\text{TOT}}$ ) és a zérusponyi rezgési energia ( $E_{\text{ZPVE}}$ )  
összegével közelíthetjük, addig az  $U_T^{\text{kor}}$  tagot az

$$U_T^{\text{kor}} = RT^2 \frac{d \ln Q}{dT} \quad (4.18)$$

egyenlettel adhatjuk meg, melyben  $Q$  az állapotösszegnek felel meg.

Mindezek alapján a  $T$  hőmérsékleten vett standard moláris entalpiát az alábbi  
formulára hagyatkozva határozhatjuk meg:

$$H_T^\circ = E_{\text{TOT}} + E_{\text{ZPVE}} + RT^2 \frac{d \ln Q}{dT} + RT, \quad (4.19)$$

A teljes elektronenergiát az

$$E_{\text{TOT}} = E_{\text{CCSD(T)}} + \Delta E_{\text{kk}} \quad (4.20)$$

formában bonthatjuk fel, ahol  $E_{\text{CCSD(T)}}$  a végtelen bázisra kiterjesztett CCSD(T) ener-  
giát jelenti,  $\Delta E_{\text{kk}}$  pedig az ún. kis korrekciókat tartalmazza, melyekhez a referenciageo-  
metriát CCSD(T)/cc-pVTZ szinten érdemes optimalizálnunk. A teljes elektronenergia  
és a zérusponyi rezgési energia számításával kapcsolatos technikai részletekről a követ-  
kező két alfejezetben olvashatunk.

## 4.2.2. A teljes elektronenergia járulékeinak meghatározása

### 4.2.2.1. CCSD(T)/CBS energia

A 4.20 egyenletben előforduló, végtelen bázisra vonatkoztatott CCSD(T) energiát  
az

$$E_{\text{CCSD(T)}} = E_{\text{HF}} + \Delta E_{\text{MP2}} + \Delta E_{\text{CCSD}} + \Delta E_{\text{(T)}} \quad (4.21)$$

összegként vehetjük fel, ahol

- (a)  $E_{\text{HF}}$  a cc-pV6Z bázison számított Hartree–Fock (RHF/cc-pV6Z) energia,
- (b)  $\Delta E_{\text{MP2}}$  a másodrendű Møller–Plesset (MP2) korrelációs energiajárulék (MP2/cc-  
pV5Z és MP2/cc-pV6Z eredményekből extrapolálva), valamint
- (c)  $\Delta E_{\text{CCSD}} = E_{\text{CCSD}} - E_{\text{MP2}}$  és  $\Delta E_{\text{(T)}} = E_{\text{CCSD(T)}} - E_{\text{CCSD}}$  a CCSD és a CCSD(T)  
hozzájárulás a CCSD(T)/CBS energiához, melyeket cc-pVQZ és cc-pV5Z számí-  
tásokból történő extrapolációval becsültünk.

Mivel a korrelációs energiatajak extrapolációjára bevezetett formulák<sup>66,100,161,162</sup>  
teljesítőképességüket tekintve nem mutatnak jelentős különbségeket, erre a célra az

elméleti megalapozottsággal<sup>163</sup> rendelkező, Helgaker és munkatársai által javasolt modellt<sup>68</sup> vettük igénybe. Ez az összefüggés a cc-pVXZ báziskészlettel számított  $E_{\text{kor}}^X$  energia és az  $X$  kardinális szám között teremt kapcsolatot:

$$E_{\text{kor}}^X = E_{\text{kor}}^\infty + bX^{-3}, \quad (4.22)$$

ahol  $E_{\text{kor}}^\infty$  a végtelen bázisra kiterjesztett korrelációs energia,  $b$  pedig egy, az extrapoláció trendjére jellemző paraméter.

#### 4.2.2.2. Kis korrekciók

Egy adott  $n$ -pentán konformer – kicsiny abszolút értékű –  $\Delta E_{\text{kk}}$  járulékát a

$$\Delta E_{\text{kk}} = \Delta E_{\text{HO}} + \Delta E_{\text{DBOC}} + \Delta E_{\text{SR}} \quad (4.23)$$

kifejezéssel adhatjuk meg, melyben  $\Delta E_{\text{HO}}$  a magasabbrendű (*higher order*, HO) korrelációs energiatagot,  $\Delta E_{\text{DBOC}}$  a diagonális Born-Oppenheimer (DBOC) korrekciót,  $\Delta E_{\text{SR}}$  pedig a skaláris relativisztikus (SR) hozzájárulást jelöli.

A magasabbrendű korreláció hatását a különböző szimmetriájú szpécieszek esetén eltérő elméleti szinteken vettük figyelembe. Míg a nagyobb szimmetriával rendelkező  $C_{2v}$  és  $gg$  ( $C_2$ ) konformerekre CCSDT(Q)/cc-pVDZ szintű számításokat is el tudunk végezni, addig a  $C_1$  pontcsoportba tartozó konformerek esetén csupán a kisebb számítógépes kapacitást igénylő CCSDT/cc-pVDZ energiák meghatározására volt lehetőségünk.

Habár a Born–Oppenheimer (BO) közelítés, melyet az eddig tárgyalt energiatagok számítása során felhasználtunk, a becsülni kívánt termokémiai mennyiségek értékeiben szintén nem okoz jelentős hibát, a szub-kJ mol<sup>-1</sup>-es pontosság érdekében érdemes figyelembe vennünk azt a hatást is, melyet a magok kinetikus energia operátora fejt ki a teljes elektronenergiára. A BO közelítés legegyszerűbb korrekciójának a DBOC hozzájárulás felel meg, mely az alábbi várható értéket takarja:

$$\Delta E_{\text{DBOC}} = \langle \Psi_e(\mathbf{r}; \mathbf{R}) | \hat{\mathbf{T}}_n | \Psi_e(\mathbf{r}; \mathbf{R}) \rangle, \quad (4.24)$$

ahol  $\hat{\mathbf{T}}_n$  a magok kinetikus energia operátora,  $\Psi_e(\mathbf{r}; \mathbf{R})$  pedig az elektronokhoz rendelt hullámfüggvény. A szóban forgó energiajárulékot CCSD/cc-pCVDZ szintű számításokból nyertük.

Mivel a Schrödinger-egyenlet Lorentz-invarianciával nem rendelkezik, annak (közelítő) megoldása a relativisztikus jelenségeket nem képes leírni. Ha e tényezők hatásával is számolni szeretnénk, akkor ahhoz a Dirac-egyenletet (vagy általában egy közelítő relativisztikus Hamilton-operátor sajátértékegyenletét) kell megoldanunk. A relativisztikus effektusok kiértékelésére az ún. Douglas–Kroll–Hess Hamilton operátor terjedt

el széles körben, melynek várható értéke az SR járuléknak feleltethető meg. A jelen FPA protokollba a  $\Delta E_{\text{SR}}$  hozzájárulást CCSD(T)/aug-cc-pCVDZ-DK elméleti szinten foglaltuk bele.

### 4.2.3. A zérusponti rezgési energia számítása

A zérusponti rezgési energiát a következő alakban vettül fel:<sup>164</sup>

$$E_{\text{ZPVE}} = G_0 + \sum_i \frac{\omega_i}{2} + \sum_{i \geq j} \frac{x_{ij}}{4}, \quad (4.25)$$

ahol  $G_0$  egy rezgési állapottól független konstans,  $\omega_i$  az  $i$ -ik harmonikus rezgési frekvencia,  $x_{ij}$  pedig az  $i$ -ik és a  $j$ -ik normálrezgés anharmonikus csatolási paramétere. Az  $\omega_i$  frekvenciákat CCSD(T)/cc-pVTZ szintű geometriákból és analitikus második deriváltakból származtattuk,<sup>165</sup> szemben a  $G_0$  és az  $x_{ij}$  mennyiségekkel, melyeket MP2/cc-pVDZ szintű szemikvartikus erőterekből<sup>166</sup> nyertük.

Mivel egyes rezgési módusok tulajdonképpen belső rotációknak felelnek meg, azok rezgésként való kezelése a számítási eredményeket jelentős hibával terhelheti, amit egyszerűbben az egydimenziós gátolt rotor korrekció segítségével enyhíthetünk. Ezt a hatást úgy vehetjük figyelembe, hogy a ZPVE-ben és a termikus korrekciókhoz szükséges állapotösszegekben a metil és a C–C–C torziók diagonális anharmonicitási együtthatóit elhagyjuk, a megfelelő harmonikus frekvenciákat pedig a

$$\left( -\frac{\hbar^2}{2I_r} \frac{d^2}{d\vartheta^2} + V(\vartheta) \right) \psi = E\psi \quad (4.26)$$

egydimenziós Schrödinger-egyenlet legkisebb sajátértékével helyettesítjük. Ebben a kifejezésben  $I_r$  a redukált tehetetlenségi momentumot jelöli,  $V(\vartheta)$  pedig a vizsgált egydimenziós torziós mozgás (periodikus) PES-e, mely Fourier-sorként vehető fel:

$$V(\vartheta) = c + \sum_k (a_k \cos(k\vartheta) + b_k \sin(k\vartheta)) \quad (4.27)$$

Ami a gátolt rotor számítások technikai részleteit érinti, a 4.26 Schrödinger-egyenletet Marston és Balint-Kurti módszerével (*Fourier grid Hamiltonian*)<sup>167,168</sup> oldottuk meg, az  $a_k$ , a  $b_k$  és a  $c$  paramétereket MP2/cc-pVTZ szintű rácspont energiákra történő illesztésekből vettük, az  $I_r$  tényezőt pedig Pitzer módszerével<sup>169,170</sup> állapítottuk meg.

### 4.2.4. Hibaszámítás

A becsült kvantumkémiai és termokémiai mennyiségekhez jól definiált, megbízható hibahatárokat rendeltünk, melyek a szisztematikusan növelt bázisú számításaink konvergenciáját tükrözik. A rendelkezésre álló információk birtokában tehát

- (a) az  $X_{\max}$  és az  $X_{\max} - 1$  kardinális számú bázisokra vonatkozó adatok különbségének abszolút értékét (a HF és a ZPVE tagokra),
- (b) az  $(X_{\max} - 1, X_{\max})$  és az  $(X_{\max} - 2, X_{\max} - 1)$  extrapolált értékek előjel nélküli eltérését (MP2, CCSD és (T) energiajárulékok esetén),
- (c) a CC és a HF eredmények abszolút különbségét (DBOC és SR korrekcióknál), valamint
- (d) a kérdéses járulék abszolút értékét (egy bázisú számításokból nyert hozzájárulások esetében)

tüntettük fel a legjobb becslések bizonytalanságaként, ahol az  $X_{\max}$  kardinális szám a legnagyobb elérhető bázishoz tartozik. Természetesen a végső mennyiségek pontosságát is megadtuk, melyet minden esetben az egyedi járulékok hibáinak összegével közelítettünk.

#### 4.2.5. Alkalmazott programok

Ami a felhasznált kvantumkémiai programcsomagokat illeti, (a) a CCSDT(Q) számításokat a CFOUR<sup>171</sup> szoftverhez kapcsolt MRCC<sup>24,172</sup> programmal végeztük el, (b) a relativisztikus DKH komponenseket a MOLPRO<sup>173</sup> kóddal határoztuk meg, illetve (c) minden más eredmény a CFOUR program outputjaiból származik. Az elektronkorrelációs számításokhoz, melyek – mint azt korábban említettük – a CCSDT és a CCSDT(Q) energiák kivételével az összes elektron korrelációját magukban foglalják, RHF molekulapályákat vettünk igénybe.

### 4.3. Az entalpiakülönbség kísérleti meghatározása

Az  $n$ -pentán konformerek relatív entalpiájának pontos meghatározása az elmúlt évtizedekben számos tanulmány tárgyát képezte. A legtöbb mérést folyadékfázisban végezték,<sup>174–179</sup> azonban néhány gázfázisú mérési eredmény is található a szakirodalomban.<sup>180–182</sup> A kísérleti eredmények többségének alapját hőmérsékletfüggő infravörös- vagy Raman-rezgési spektrumok képezik.<sup>174,175,177,178,180,182</sup> A relatív entalpia spektrumok alapján történő becslésének alapja abban áll, hogy az egyes konformerekhez rendelhető rezgési sávok  $I$  intenzitása arányos az adott konformer  $n$  mennyiségével:  $I_{\alpha\beta} = f_{\alpha\beta}n_{\alpha\beta}$ , ahol  $f_{\alpha\beta}$  a kísérleti körülményektől, valamint az átmenet valószínűségétől függő együttható.

Az entalpiakülönbségek előállítására a különböző konformerekhez tartozó sávok in-

tenzítésének aránya alapján a következő egyenlet felhasználásával lehetséges:

$$\frac{I_{\alpha\beta}}{I_{tt}} = \frac{f_{\alpha\beta}n_{\alpha\beta}}{f_{tt}n_{tt}} = f'_{\alpha\beta}K, \quad (4.28)$$

ahol  $K$  a  $tt \rightleftharpoons \alpha\beta$  folyamat egyensúlyi állandója, melynek értéke a keresett entalpia-különbség segítségével kifejezhető. Az egyensúlyi állandó behelyettesítésével az

$$\frac{I_{\alpha\beta}}{I_{tt}} = f'_{\alpha\beta}e^{-\Delta G^\circ(\alpha\beta)/RT} = e^{-\Delta H^\circ(\alpha\beta)/RT + b_{\alpha\beta}} \quad (4.29)$$

egyenlet adódik, melyben  $b_{\alpha\beta} = \ln f'_{\alpha\beta} + \Delta S^\circ(\alpha\beta)/R$ , a  $\Delta G^\circ(\alpha\beta)$  paraméter az  $\alpha\beta$  konformer  $tt$  konformerre vonatkozó relatív szabadentalpiáját, a  $\Delta H^\circ(\alpha\beta)$  tényező annak relatív entalpiáját, míg  $\Delta S^\circ(\alpha\beta)$  a relatív entrópiáját jelöli. Az egyenletet linearizálva az intenzitásarányok hányadosának logaritmusára a hőmérséklet reciprokanak függvényében az

$$\ln \frac{I_{\alpha\beta}}{I_{tt}} = -\frac{\Delta H^\circ(\alpha\beta)}{R} \frac{1}{T} + b_{\alpha\beta} \quad (4.30)$$

egyenes illeszthető, melynek meredekségéből a keresett entalpiakülönbség kifejezhető. A 4.30 egyenlet alkalmazhatóságának szükséges feltétele a rezgési spektrumokon egyértelműen az egyes konformerekhez rendelhető, nem átfedő sávok megjelenése.

A transzformáció azonban hatással van a pontok bizonytalanságára, a korrekt kiértékeléshez ezért ennek a torzító hatásnak a figyelembevétele is elengedhetetlen. Amennyiben a kísérleti hibát homoszkedasztikusnak tételezzük fel, a linearizálás után az értékek varianciája különböző, melyet egy 4.29 egyenlethez hasonló nemlineáris  $f(x)$  függvény esetén, a függvényt az

$$f(x) \approx f(x_0) + \left. \frac{df(x)}{dx} \right|_{x=x_0} (x - x_0) \quad (4.31)$$

egyenlet szerint a lineáris tagig Taylor-sorba fejtvé, a lineáris függvényekre vonatkozó formula alapján egyszerűen közelíthetünk. Az egyetlen változó lineáris transzformációjára ( $f(x) = ax + b$ ) vonatkozó összefüggés a következőképp fejezhető ki:

$$\sigma_f^2 = a^2 \sigma_x^2 = \frac{df(x)}{dx} \sigma_x^2, \quad (4.32)$$

ahol  $\sigma_f^2$  a  $\sigma_x^2$  varianciájú  $x$  változó  $f$  függvény szerinti transzformációja után nyert szórásnégyzetét jelöli. A transzformált értékek varianciáinak ismeretében a súlyozott legkisebb négyzetek (WLLS) módszere szerint végzett regresszió már a kísérleti hibát is magába foglaló, korrekt eredményt szolgáltat.

A kísérleti adatok helyes kiértékelésének másik, egyszerűbb módját az eredeti nemlineáris függvény (4.29 egyenlet) empirikus szórásnégyzetek segítségével történő súlyozott illesztése jelenti.



## 4.4. QSPR modellezés

Korábbi eredmények arra engednek következtetni, hogy az  $n$ -alkánok konformereinek relatív entalpiái a konformerek szimmetriájával is korrelálhatnak. Az elméleti, valamint a kísérletileg meghatározott entalpiák a konformerek pontcsoportjának méretével mind az  $n$ -bután,<sup>160</sup> mind az  $n$ -pentán<sup>183</sup> esetén közel monoton változnak. Az  $n$ -butánnál a *transz* ( $t \approx 180^\circ$ ) konformer entalpiája alacsonyabb, pontcsoportjának ( $C_{2v}$ ) rendje pedig nagyobb, mint a *gauche* ( $g \approx \pm 60^\circ$ ) konformerének ( $C_2$ ). Hasonló helyzet áll fenn az  $n$ -pentán esetén is, ahol a konformerek entalpiái a  $tt$ ,  $tg$ ,  $gg$ ,  $gx$  ( $x \approx \pm 95^\circ$ ) sorrendben nőnek, míg szimmetriájuk általában csökken: pontcsoportjuk rendre  $C_{2v}$ ,  $C_1$ ,  $C_2$  és  $C_1$ .

Bizonyos 2D, illetve 3D deskriptorok a molekula szimmetriájára vonatkozó információt is kódolnak, ezért várhatóan a konformerek entalpiáival vagy egyéb termokémiai tulajdonságaival is korrelálhatnak, így azok kvantitatív leírására egy egyszerű lehetőséget nyújtanak.

A hőmérséklet mint változó alkalmazásával az entalpiakülönbségek hőmérsékletfüggésének jellemzésére is esély nyílik. A hőmérséklet természetesen a hagyományos értelemben nem nevezhető deskriptornak, mivel nem a szerkezetből vezethető le, ennek ellenére egy hasonló vizsgálat hőmérsékletfüggő mennyiségek ilyen jellegű modellezéséről értékes információkkal szolgálhat.

### 4.4.1. Módosított topológiai indexek

A topológiai indexek kiváló alkalmazhatóságát számos különféle molekulahalmaz tulajdonságainak modellezése során bizonyították, többek között alkánok standard képződéshőjének leírásakor<sup>119,131</sup> is. Egy tanulmányban kizárólag a Randić-index felhasználásával egy 48 alkánra végzett vizsgálat során  $12.26 \text{ kJ mol}^{-1}$  maradékszórású modellt állítottak fel.<sup>119</sup> Azonos szénatomszámú vegyületek között azonban az index értéke és a képződési entalpia ellentétesen változik. A probléma megoldására Randić a szénatomszám mint további deskriptor bevezetését javasolta.<sup>131</sup>

Egy ilyen jellegű modell felállítása érdekében egy további gráfinvariánst is kifejlesztettek, melyet ABC indexnek neveztek,<sup>119</sup> és szintén alkalmasnak bizonyult a képződéshő modellezésére. Az új deskriptorral végzett illesztés a 48 alkánra  $3.12 \text{ kJ mol}^{-1}$  maradékszórást eredményezett, mely jelentősen felülmúlja a Randić-indexszel kapott eredményt.

Az elektrotológiai állapot fogalmának bevezetésével és felhasználásával Kier és Hall új, az elektronállapotot is kódoló deskriptorokat definiáltak.<sup>184</sup> Ezek a leírók a tisztán topológiai alapú deskriptoroknál nyilvánvalóan több információt hordoznak, így különféle jellemzőkkel azoknál számottevően jobb korrelációt mutatnak. Az egyes

atomok kémiai környezetét is leíró változókat megalkotásuk óta ezért számos esetben használták QSAR/QSPR modellek felállítása során.<sup>135–140,184</sup> Elektrotopológiai indexeket többek között alifás ketonok képződési- és párolgási entalpiájának,<sup>185</sup> valamint heterociklusos vegyületek termokémiai paramétereinek leírására is sikerrel alkalmaztak.<sup>117</sup> Az indexek korrelációja különösen a képződési entalpiával bizonyult kiemelkedőnek, melyek segítségével 46 heterociklusos vegyületre négy deskriptor felhasználásával 16.75 kcal mol<sup>-1</sup> maradékszórású modell írható fel, a kiugró pontok eltávolításával ez az érték 8.45 kcal mol<sup>-1</sup>-re csökkenthető, ami vetekszik a PM3 kvantumkémiai számítások pontosságával. Ez a hiba ugyan nagyobb, mint ami az alkánok esetén megfigyelhető volt, azonban nem szabad elfelejtenünk, hogy sokkal változatosabb, heteroatomokat is tartalmazó vegyületekről van szó, melyek sokkal jelentősebb diverzitást mutatnak.

Az eddigi eredmények alapján a topológiai indexek és különösen az egyéb tulajdonságokat is figyelembe vevő elektrotopológiai deskriptorok vegyületek termokémiai tulajdonságainak, termodinamikai potenciálfüggvényeinek jellemzésére kiválóan alkalmazhatók. A háromdimenziós szerkezet ismeretének hiányában viszont a deskriptorok egy bizonyos molekula konformereinek megkülönböztetésére alkalmatlanok, mivel azok csak az atomi pozíciók tekintetében térnek el egymástól. Ilyen típusú modellezéshez bizonyos térszerkezeti jellemzőket is számításba kell vennünk. A topológiai indexek kiterjesztéseként korábban sikeresen alkalmaztak topografikus indexeket, melyek az atomok távolságával súlyozott gráfelméleti mátrixokból vezethetők le. Logikusnak tűnik ezért a termokémiai tulajdonságok modellezésére topografikus indexeket alkalmazni.

Egy egyszerű,  $V$  csúcshalmazzal és  $E$  élhalmazzal jellemezhető  $G(V, E)$  (molekula)gráf reprezentálható annak  $|V| \times |V|$ -s adjacenciamátrixa (szomszédsági mátrixa,  $\mathbf{A}$ ) segítségével ( $|V|$  a csúcshalmaz számossága):

$$A_{ij} = \begin{cases} 1, & \text{ha } i \in V \text{ és } j \in V \text{ szomszédos} \\ 0, & \text{különben} \end{cases} \quad (4.33)$$

Ha a  $G$  gráf irányítatlan, akkor  $\mathbf{A}$  szimmetrikus. Valós, szimmetrikus mátrixként egy egyszerű gráf adjacenciamátrixának  $\lambda_i$  sajátértékei valós számoknak adódnak, melyekből újabb molekuláris deskriptorok származtathatók.

A gráf csúcsainak fokszámát a fokszámvektor ( $\mathbf{v}$ ) elemei a következő formában adják meg:

$$v_i = \sum_{j=1}^{|V|} A_{ij} \quad (4.34)$$

A gráf távolságmátrixán ( $\mathbf{D}$ ) az egyes csúcsok közötti legrövidebb – metrikus vagy egyéb, tetszőleges – távolságokból felépített mátrixot értjük. Általában egy súlyozatlan gráf csúcsai közötti legrövidebb út hossza, melyen a legrövidebb útban szereplő élek számát értik, az első kategóriába sorolható.

A távolságösszeg-vektor ( $\mathbf{s}$ ) a távolságmatrix segítségével határozható meg a következőképp:

$$s_i = \sum_{j=1}^{|V|} D_{ij} \quad (4.35)$$

A gráfinvariánsok definiálása általában a szomszédsági- és a távolságmatrix, valamint a fokszám- és a távolságösszeg-vektorok eleminek vagy a szomszédsági matrix sajátértékeinek felhasználásával történik.

Ha az indexekbe a topológiai jellemzőkön kívül más tulajdonságokat is bele szeretnénk foglalni, az adjacenciamatrixot vagy a távolságmatrixot kell parametrizálnunk. A matrixok módosítása tetszőleges, azonban gyakran a legrövidebb utak hossza helyett az atomok euklideszi távolságát alkalmazzák. Az adjacenciamatrix helyett egy szabadon választott súlymatrix felépítése és felhasználása, melyet úgy kaphatunk, hogy a gráf egyes csúcsaihoz, valamint éleihez bizonyos megfontolások alapján valamilyen súlyokat rendelünk, további topografikus indexek definiálására is lehetőséget teremt. Ezek az indexek leginkább a súlyok megválasztásában különböznek.

A torziós szög periodicitása miatt konformerek leírásánál a definiált molekuláris deskriptor periodicitása alapvető követelményként nevezhető meg. Ez a feltétel könnyen teljesíthető például maguknak a torziós szögeknek a felhasználásával. Figyelmet kell azonban fordítani arra is, hogy az index értéke folytonos legyen. Egy rosszul definiált leíró esetén ugyanis előfordulhat, hogy annak értéke  $-180^\circ$  és  $+180^\circ$  (vagy  $0^\circ$  és  $360^\circ$ ) torziós szög értékeknél a szerkezetek egyezése ellenére eltérő. Az így definiált deskriptor a konformációs különbségek modellezésére a  $\pm 180^\circ$ -nál fellépő folytonossági hiány miatt nem alkalmas.

#### 4.4.2. Nemlineáris modellfüggvények

Egyes módszerek, például a mesterséges neurális hálózatok, erősen nemlineáris összefüggést alkalmaznak az eredmények bemenő adatokból történő előállítására. Bár egy ilyen hálózat modellként szolgálhat egy-egy tulajdonság leírására, a változók közötti összefüggés megértéséhez vagy a lényeges deskriptorok kiválasztásához nem megfelelő. Egy ilyen bonyolult rendszer esetén az adatok könnyen túlilleszthetők, ami a mérési hiba leírását is magával vonja.

Az adatok és a deskriptorok közötti összefüggés megállapítására néhány esetben nemlineáris modellfüggvényt is sikerrel alkalmaztak.<sup>26-28</sup> Az illesztett függvény hasonló az MLR analízis során használthoz, viszont a független változók között a deskriptorok (másodfokú) hatványai is szerepelnek. Ezek a modellek sokkal egyszerűbben értelmezhetők, mint az ANN-ek, ráadásul a függő és a független változók közötti összefüggés sokkal szemléletesebben fejezhető ki. Lényegesen egyszerűbb a felállított modell statisztikai analízisének végrehajtása is, mely egy QSAR/QSPR modell validálásában

alapvető szerepet játszik.

Az egyszerű nemlineáris modellfüggvények az analízisből jól ismert hatványsoron alapulnak, melynek alakja egy  $f(x)$  függvény egy adott  $x_0$  pontja körül a következő:

$$f(x) = \sum_{k=1}^{\infty} a_k (x - x_0)^k \quad (4.36)$$

A hatványsor konvergálhat a teljes  $\mathbb{R}$  (vagy  $\mathbb{C}$ ) halmazon vagy annak egy véges tartományában. Utóbbi esetben a konvergenciasugarat a Cauchy-Hadamard-képlettel számíthatjuk ki:

$$r = \frac{1}{\limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|}} \quad (4.37)$$

Bár véges tagszámú egyváltozós polinomok használatára találunk néhány példát az irodalomban, az ilyesfajta modellfüggvények széles körben nem terjedtek el, sőt a többváltozós polinomiális modellek meg sem jelentek a QSAR/QSPR modellezés területén. A konformációs entalpiakülönbségek modellezése során figyelembe vehető azok hőmérsékletfüggése is, így a teljes modellfüggvény többváltozós alakban írható fel. Az alkánkonformerek entalpiakülönbségeinek pontosabb modellezése érdekében tehát érdemes egy

$$f(\mathbf{x}) = \sum_{\epsilon \in \mathbb{N}_0^n} a_{\epsilon} (\mathbf{x} - \mathbf{x}_0)^{\epsilon} = \sum_{i_1, \dots, i_n=0}^{\infty} a_{i_1, \dots, i_n} \prod_{j=1}^n (x_j - x_{j,0})^{i_j} \quad (4.38)$$

egyenlettel jellemezhető többváltozós hatványsor alakú modellfüggvény alkalmazása, ahol  $n$  a változók száma,  $\epsilon$  pedig a kitevőkből alkotott vektor.

# 5. Eredmények és értékelésük

## 5.1. Szimmetriaanalízis

### 5.1.1. Algoritmus

Egy molekula geometriája atomjainak helyvektoraival reprezentálható, míg a szimmetriaműveletek vektoralgebrai leírása szintén lehetséges. A szimmetriacentrum annak helyvektorával jellemezhető, a szimmetriatengelyek egy pontjukkal és irányvektorokkal adhatók meg, a szimmetriasíkok pedig egy pontjukkal és a sík normálvektorával fejezhetők ki. Ez a szimmetria 4.3, 4.5 és 4.10 egyenletek alapján történő, tisztán vektoralgebrai meghatározására nyújt lehetőséget. A szimmetriaanalízis általunk alkalmazott folyamatának vázlata az 5.1. ábrán tekinthető meg.<sup>186,187</sup>

- 
- 1: atomok számának, rendszámának és koordinátáinak beolvasása
  - 2: atomok osztályozása rendszám alapján
  - 3: tömegközéppont eltolása az origóba
  - 4: inverziós centrum létezésének ellenőrzése
  - 5: linearitás ellenőrzése
  - 6: planaritás ellenőrzése
  - 7: szimmetriasíkok meghatározása (SYMMETRY PLANES eljárás)
  - 8: valódi szimmetriatengelyek meghatározása (PROPER AXES eljárás)
  - 9: nem valódi szimmetriatengelyek meghatározása
  - 10: pontcsoport, vázscsoport és egyéb tulajdonságok előállítása
  - 11: legnagyobb szimmetria detektálása
  - 12: választott részcsoporthoz optimalizálás (OPTIMIZE eljárás)
- 

5.1. ábra: A szimmetria és egyéb tulajdonságok meghatározásának lépései.

Mivel a szimmetriaműveletek azonos rendszámú atomokat cserélnek fel, a vizsgálat azokra a transzformációkra korlátozható, melyek azonos rendszámú atomokat permutálnak. Az egyező rendszám alapján előállított  $A_{\text{eq},i}$  ekvivalenciaosztályok (5.1. ábra 2. sor) halmazát jelöljük  $A_{\text{eq}}$ -val ( $A_{\text{eq}} = \{A_{\text{eq},i}\}$ ). A továbbiakban a 4.11, a 4.12 és a 4.13 egyenletekben szereplő  $A^2$  és  $A^3$  halmazok az  $\bigcup_i \{A_{\text{eq},i}^2\}$  és az  $\bigcup_i \{A_{\text{eq},i}^3\}$  halmazokkal helyettesíthetjük, mely – különösen nagyobb molekulák esetén – az ezekből előállított potenciális szimmetriaelemek számát jelentős mértékben csökkentheti.

A szimmetria detektálása a molekula tömegközéppontjának előállításával és annak a koordinátarendszer origójába történő eltolásával folytatódik (5.1. ábra 3. sor).

Ezt a pontot az összes szimmetriaelem tartalmazza, így minden szimmetriaműveletre nézve invariáns, ezért alkalmas ezeknek a műveleteknek a leírására. A továbbiakban az atomok helyvektorán tehát a molekula tömegközéppontjából az adott atomba mutató vektort értjük.

A szimmetriaelemek kijelöléséhez szükséges pont ismeretében a feladat a szimmetriacentrum létezésének ellenőrzése, valamint a szimmetriasíkok normálvektorának és a forgástengelyek irányvektorának, illetve rendjének generálása. A normálvektorok és az irányvektorok a 4.1. fejezetben kifejtett okból egy konstans faktor erejéig határozatlanok. Legegyszerűbben egységvektorként kezelhetők, így ezen vektorok alatt a továbbiakban mindig normált vektorokat értünk.

A szimmetriacentrum létének megállapítása (5.1. ábra 4. sor) egyszerűen az atomok  $\mathbf{p}_i$  helyvektorainak az 5.1 egyenlet szerinti additív invertálásával és az így előállított magkonfiguráció eredetivel történő összevetésével valósítható meg.

$$\mathbf{p}'_i = -\mathbf{p}_i \quad (5.1)$$

A kiindulási konfigurációval való összehasonlítás során a 4.1 és a 4.2 egyenletek érvényességét szükséges megállapítanunk. Az eredeti geometriával történő összehasonlítás a későbbiekben is ezzel azonos módon történik. Egy tetszőleges geometria kismértékű torzultságával minden esetben számolnunk kell, így a 4.1 egyenletet csak egy előre választott tolerancián belül tekinthetjük érvényesnek. Amennyiben az 5.1 transzformáció végrehajtása után a 4.1 és a 4.2 relációk fennállnak, a molekula rendelkezik szimmetriacentrummal.

A szimmetriaanalízis következő lépését (5.1. ábra 5. sor) a molekula linearitásának vizsgálata jelenti, mely az atomi helyvektorok párhuzamosságának ellenőrzésével az 5.2 egyenlet alapján valósítható meg.

$$|\mathbf{p}_i \cdot \mathbf{p}_j| = 1 \quad (5.2)$$

A molekula lineáris, amennyiben az 5.2 egyenlet minden  $i, j$  atompárra teljesül, mely esetben a  $D_{\infty h}$  vagy a  $C_{\infty v}$  pontcsoportba sorolható attól függően, hogy rendelkezik-e szimmetriacentrummal vagy sem.

Az előzőekben ismertetetthez hasonló módon a molekula planaritása is megállapítható (5.1. ábra 6. sor). Ha a molekula síkalkatú, tetszőleges  $i$ -re és  $j$ -re fennáll, hogy a  $\mathbf{p}_i \times \mathbf{p}_j = \mathbf{n}_p$  vektor a molekula síkjára merőleges. A vizsgált szerkezet planáris, ha az összes atomi helyvektor  $\mathbf{n}_p$ -re ortogonális:

$$\mathbf{p}_i \cdot (\mathbf{p}_1 \times \mathbf{p}_2) = 0, \quad (5.3)$$

ahol a sík normálvektorának előállítása az első két (tetszőlegesen rendezett) atom hely-

vektorának felhasználásával történt.

A vizsgálat következő lépéseként a szimmetriasíkok lokalizálása adható meg (5.1. ábra 7. sor). Az erre a célra szolgáló algoritmus az 5.2. ábrán, pszeudokód formában került bemutatásra, ahol  $\mathbf{P}$  és  $\mathbf{P}_{\text{tr}}$  az atomok kiindulási és a transzformáció utáni helyvektorainak mátrixát,  $\mathbf{O}$  az origót jelöli, a REFLECT( $\mathbf{O}, \mathbf{n}, \mathbf{P}$ ) eljárás a  $\mathbf{P}$  mátrix oszlopvektorait (atomi helyvektorokat) a 4.5 egyenlet szerint az  $\mathbf{O}$  pont és az  $\mathbf{n}$  normálvektor által kijelölt síkra tükrözi, az APPEND( $\sigma, \mathbf{n}$ ) rutin pedig az  $\mathbf{n}$  normálvektort a  $\sigma$  halmazhoz adja hozzá.

---

```

1: procedure SYMMETRY PLANES
2:   for all  $A_{\text{eq},i} \in A_{\text{eq}}$  do                                ▷  $A_{\text{eq},i}$  azonos rendszámú atomok halmaza
3:     for all  $i, j \in A_{\text{eq},i}$  do                                ▷ minden  $i, j$  atomra
4:        $\mathbf{h} \leftarrow (\mathbf{p}_i + \mathbf{p}_j)/2$                             ▷  $i, j$  atomok helyvektorának átlaga
5:        $\mathbf{n} \leftarrow (\mathbf{p}_i - \mathbf{p}_j)/|\mathbf{p}_i - \mathbf{p}_j|$                 ▷ a sík normálvektora
6:       if  $|\mathbf{h} \cdot \mathbf{n}| < \text{TOL} \wedge \mathbf{n} \notin \sigma$  then
7:          $\mathbf{P}_{\text{tr}} \leftarrow \text{REFLECT}(\mathbf{O}, \mathbf{n}, \mathbf{P})$           ▷  $\mathbf{P}$  helyvektorok tükrözése a síkra
8:         if  $\max_{i \in \mathbb{Z}_N^+} \min_{j \in \mathbb{Z}_N^+} |\mathbf{p}_{\text{tr},i} - \mathbf{p}_j| < \text{TOL}$  then
9:           APPEND( $\sigma, \mathbf{n}$ )                                ▷  $\mathbf{n}$  tárolása
10:        end if
11:      end if
12:    end for
13:  end for
14: end procedure

```

---

5.2. ábra: A szimmetriasíkok meghatározása.

Amennyiben a szerkezet planáris, az előző lépésben a molekula síkját mint szimmetriasíkot már meghatároztuk. Ellenkező esetben a keresett szimmetriasíkra történő tükrözés a síkon kívüli atomokat páronként egymásba transzformálja. A síknak az annak megfelelő szimmetriaművelet által permutált két atomot összekötő szakasz felezőpontját és a tömegközéppontot is tartalmaznia kell. Ennek a követelménynek a teljesülése az 5.2. ábra 6. sorában kerül ellenőrzésre (első feltétel). A kérdéses sík  $\mathbf{n}_{i,j}$  normálvektora az  $i$  és a  $j$  atomokat egymásba transzformáló művelet esetén

$$\mathbf{n}_{i,j} = \frac{\mathbf{p}_i - \mathbf{p}_j}{|\mathbf{p}_i - \mathbf{p}_j|}. \quad (5.4)$$

Az  $\mathbf{n}_p$  vektor és az összes  $\mathbf{n}_{i,j}$  halmaza a 4.11 egyenletben szereplő  $\sigma_p$  halmazt definiálja. Mivel az 5.4 egyenlettel meghatározott normálvektorok között előfordulhatnak azonosak, azok szigorú értelemben nem alkotnak halmazt, a lehetséges szimmetriasíkok  $\sigma_p$  halmazát ezért  $\mathbf{n}_p$  és a nem párhuzamos  $\mathbf{n}_{i,j}$ -k alkotják, míg a szimmetriasíkok

normálvektorainak  $\sigma$  halmazát ennek részhalmazaként keressük:

$$\sigma \subseteq \sigma_p = \{\mathbf{n}_p, \mathbf{n}_{i,j}\} \quad (5.5)$$

A  $\sigma_p$  halmaz elemeinek az 5.2. ábra 8. sorában végrehajtott vizsgálata és a  $\sigma$  halmaz elemeinek kiválasztása az inverziós centrum esetén ismertetettel azonos módon történik.

A szimmetriatengelyek előállítására (5.1. ábra 8. és 9. sor) felírható algoritmus pszeudokódja az 5.3. ábrán látható. A kódban felhasznált ROTATE( $\mathbf{O}, \mathbf{d}, \alpha, \mathbf{P}$ ) eljárás a  $\mathbf{P}$  mátrix oszlopvektorait a 4.10 egyenlet szerint az  $\mathbf{O}$  pont és a  $\mathbf{d}$  irányvektor segítségével definiált tengely körül  $\alpha$  szöggel forgatja el.

Lineáris molekulák esetén a végtelen rendű szimmetriatengelynek a molekula atomjai által meghatározott egyenes felel meg, a  $D_{\infty h}$  pontcsoportba sorolható szerkezeteknél pedig a végtelen számú  $C_2$  tengely erre az egyenesre merőleges. Minden további esetben egy tetszőleges szimmetriatengely körüli forgatás a tengely rendjének megfelelő számú atomot permutál. Ezek alapján a  $C_2$  tengelyek lokalizálása az 5.3. ábra 2 – 11. sora szerint két azonos rendszámú atom helyvektorának felhasználásával lehetséges, míg a magasabb rendű tengelyek irányvektorai három, egy szabályos sokszög csúcaiban elhelyezkedő atom helyvektoraival az 5.3. ábra 12 – 24. sorainak megfelelően írhatók fel.

A  $C_2$  tengelyek irányvektorai a tengely körüli forgatás során felcserélődő két atom helyvektorának összegével párhuzamosak (5.3. ábra 4. sor), és a következő egyenlet segítségével állíthatók elő:

$$\mathbf{d}_{i,j} = \frac{\mathbf{p}_i + \mathbf{p}_j}{|\mathbf{p}_i + \mathbf{p}_j|} \quad (5.6)$$

E vektorok által kijelölt tengelyek körüli forgatást elegendő abban az esetben végrehajtani, ha a kérdéses vektor nem nullvektor, valamint a tengely merőleges a két atomot összekötő szakaszra (5.3. ábra 5. sorának első és második feltétele).

A magasabb rendű tengelyek irányvektora három ( $i, j, k$ ) azonos rendszámú atom helyvektorából a következőképp állítható elő (5.3. ábra 13. sora):

$$\mathbf{d}_{i,j,k} = \frac{(\mathbf{p}_j - \mathbf{p}_i) \times (\mathbf{p}_k - \mathbf{p}_j)}{|(\mathbf{p}_j - \mathbf{p}_i) \times (\mathbf{p}_k - \mathbf{p}_j)|} \quad (5.7)$$

Az  $\mathbf{d}_{i,j,k}$  vektor ebben az esetben is csak akkor lehet egy szimmetriatengely irányvektora, ha merőleges az atomokat összekötő szakaszokra és nem nullvektor (5.3. ábra 15. sor első és 17. sor első két feltétele). Egy további feltételt ezekre a vektorokra nézve az jelent, hogy egy szabályos sokszög csúcaiban kell elhelyezkedniük (5.3. ábra 15. sorának második feltétel), ami jelentősen csökkenti azoknak a műveleteknek a számát, melyet a szimmetriatengelyek előállításához meg kell vizsgálnunk. A tengely  $n$  rendjét a három atom által bezárt szög segítségével kaphatjuk meg, ugyanis egy  $n$  oldalú szabályos sokszög három szomszédos csúcsa által bezárt szög értéke  $\alpha = (n-2)180^\circ/n$ . Az



---

```

1: procedure PROPER AXES
2:   for all  $A_{\text{eq},i} \in A_{\text{eq}}$  do                                 $\triangleright A_{\text{eq},i}$  azonos rendszámú atomok halmaza
   # $C_2$  tengelyek meghatározása
3:     for all  $i, j \in A_{\text{eq},i}$  do                                 $\triangleright$  minden  $i, j$  atomra
4:        $\mathbf{d} \leftarrow (\mathbf{p}_i + \mathbf{p}_j) / |\mathbf{p}_i + \mathbf{p}_j|$                  $\triangleright$  szimmetriatengely irányvektora
5:       if  $|\mathbf{d}| > \text{TOL} \wedge |\mathbf{d} \cdot (\mathbf{p}_i - \mathbf{p}_j)| < \text{TOL} \wedge \mathbf{d} \notin C$  then
    $\triangleright \mathbf{d}$  nem nullvektor és merőleges  $\mathbf{p}_i - \mathbf{p}_j$ -re
6:          $\mathbf{P}_{\text{tr}} \leftarrow \text{ROTATE}(\mathbf{O}, \mathbf{d}, 180^\circ, \mathbf{P})$          $\triangleright \mathbf{P}$  helyvektorok forgatása
7:         if  $\max_{i \in \mathbb{Z}_N^+} \min_{j \in \mathbb{Z}_N^+} |\mathbf{p}_{\text{tr},i} - \mathbf{p}_j| < \text{TOL}$  then
    $\triangleright$  az elforgatott helyvektorok azonosak az eredetiekkel
8:            $\text{APPEND}(C, \mathbf{d})$                                         $\triangleright \mathbf{d}$  tárolása
9:         end if
10:      end if
11:    end for

# magasabb rendű tengelyek meghatározása
12:   for all  $i, j, k \in A_{\text{eq},i}$  do                                 $\triangleright$  minden  $i, j, k$  atomra
13:      $\mathbf{d} \leftarrow (\mathbf{p}_j - \mathbf{p}_i) \times (\mathbf{p}_k - \mathbf{p}_j)$                  $\triangleright$  szimmetriatengely irányvektora
14:      $\alpha \leftarrow \cos^{-1} [(\mathbf{p}_j - \mathbf{p}_i) \cdot (\mathbf{p}_k - \mathbf{p}_j)] / (|\mathbf{p}_j - \mathbf{p}_i| |\mathbf{p}_k - \mathbf{p}_j|)$   $\triangleright$  forgatás szöge
15:     if  $|\mathbf{d}| > \text{TOL} \wedge \alpha = 360^\circ/k$  ( $k \in \mathbb{Z}^+$ ) then
    $\triangleright \mathbf{d}$  nem nullvektor és  $i, j, k$  egy szabályos sokszög csúcsai
16:        $m \leftarrow 360^\circ/\alpha$                                         $\triangleright$  szimmetriatengely rendje
17:       if  $|\mathbf{d} \cdot (\mathbf{p}_i - \mathbf{p}_j)| < \text{TOL} \wedge |\mathbf{d} \cdot (\mathbf{p}_j - \mathbf{p}_k)| < \text{TOL} \wedge \mathbf{d} \notin C$  then
    $\triangleright \mathbf{d}$  merőleges  $\mathbf{p}_i - \mathbf{p}_j$ -re és  $\mathbf{p}_j - \mathbf{p}_k$ -ra
18:          $\mathbf{P}_{\text{tr}} \leftarrow \text{ROTATE}(\mathbf{O}, \mathbf{d}, \alpha, \mathbf{P})$          $\triangleright \mathbf{P}$  helyvektorok forgatása
19:         if  $\max_{i \in \mathbb{Z}_N^+} \min_{j \in \mathbb{Z}_N^+} |\mathbf{p}_{\text{tr},i} - \mathbf{p}_j| < \text{TOL}$  then
    $\triangleright$  az elforgatott helyvektorok azonosak az eredetiekkel
20:            $\text{APPEND}(C, \mathbf{d})$                                         $\triangleright \mathbf{d}$  tárolása
21:         end if
22:       end if
23:     end if
24:   end for
25: end for
26: end procedure

```

---

5.3. ábra: A szimmetriatengelyek meghatározása.

összes atomhármásra elvégezve az elemzést a lehetséges legnagyobb rend minden esetben meghatározható, melyet a forgatás szögének szisztematikus változtatásával szintén megállapíthatunk. Ehhez a talált szimmetriatengely körüli forgatás szögét egy előre választott értékig szükséges csökkenteni, azaz a tengely rendjét növelni, majd az új szöggel újra végrehajtani a forgatást. A kérdéses tengely rendje ezután a legkisebb szögű szimmetriaműveletnek adódó forgatásból a korábban ismertetett módon adható meg.

A 4.14 egyenletben definiált lehetséges szimmetriatengelyek irányvektorait tartal-

mazó  $C_p$  halmazt tehát az egymástól eltérő  $\mathbf{d}_{i,j}$  és  $\mathbf{d}_{i,j,k}$  irányvektorok alkotják, a szimmetriatengelyek tényleges irányvektorait tartalmazó  $C$  halmaz pedig ennek rész-halmazaként áll elő:

$$C \subset C_p = \{\mathbf{d}_{i,j}, \mathbf{d}_{i,j,k}\} \quad (5.8)$$

A szimmetriasíkok tárgyalásánál említett okból az analízis során továbbra is csak az egyedi, nem párhuzamos  $\mathbf{d}_{i,j}$  és  $\mathbf{d}_{i,j,k}$  vektorokból felépített halmazt tekintjük.

A tükrözéses forgástengelyek irányvektorai szintén a  $C_p$  halmaz elemei közül kerülnek ki, meghatározásuk így a  $C_p$ -beli vektorok által kijelölt tengelyek körüli forgatás és az arra merőleges síkra történő tükrözés egymás utáni végrehajtásával valósítható meg.

Mivel a pontcsoportok eltérő számú szimmetriaelemmel rendelkeznek, azok számának és típusának ismeretében a molekula pontcsoportjának megadása az egyes csoportok elemeinek és a talált szimmetriaelemeknek az összevetésével egyszerűen kivitelezhető. A vázcsoporthoz ezen kívül mindössze az egyes atomok szimmetriaalterekbe való besorolását szükséges végrehajtani, melyet az ismert szimmetriaelemektől mérhető távolság kiszámításával könnyen elérhetünk: egy atom egy adott szimmetriaelemen fekszik, ha az attól mért távolsága egy adott toleranciánál kisebbnek adódik. A pontcsoport alapján számos egyéb tulajdonság is egyszerűen megállapítható, mint például az adott molekula polaritása, kiralitása vagy az infravörös-, illetve Raman-aktív normálrezgések száma, illetve szimmetriája. Utóbbihoz először a normálkoordináták bázisán felírt reducibilis reprezentációt szükséges előállítanunk, mely az  $\hat{\mathbf{R}}$  szimmetriaműveletekre a következőképp tehető meg:

$$\begin{aligned} C_n^m : \quad \chi^{\text{red}}(\hat{\mathbf{R}}) &= n_{\hat{\mathbf{R}}} \left( 1 + 2 \left( m \frac{2\pi}{n} \right) \right), \\ S_n^m : \quad \chi^{\text{red}}(\hat{\mathbf{R}}) &= n_{\hat{\mathbf{R}}} \left( -1 + 2 \left( m \frac{2\pi}{n} \right) \right), \end{aligned} \quad (5.9)$$

$$E = C_1, \quad i = S_2, \quad \sigma = S_1,$$

ahol  $n_{\hat{\mathbf{R}}}$  az  $\hat{\mathbf{R}}$  szimmetriaműveletre nézve invariáns atomok számát adja meg,  $\chi^{\text{red}}(\hat{\mathbf{R}})$  pedig a kérdéses reprezentációban a művelet karakterét jelöli. Az így előállított reprezentáció ezután a jól ismert

$$a_\mu = g^{-1} \sum_{\hat{\mathbf{R}}} \chi^{\text{red}}(\hat{\mathbf{R}}) \chi^\mu(\hat{\mathbf{R}}) \quad (5.10)$$

összefüggés segítségével redukálható a pontcsoport irreducibilis reprezentációinak direkt összegére, ahol  $a_\mu$  megadja, hogy a  $\mu$  irreducibilis reprezentáció a reducibilis reprezentációban hányszor szerepel,  $g$  a csoport rendjére vonatkozik,  $\chi^\mu(\hat{\mathbf{R}})$  pedig az  $\hat{\mathbf{R}}$  szimmetriaművelet  $\mu$  irreducibilis reprezentációbeli karakterének felel meg.

A pontcsoport ismeretében a geometria szimmetrizálása, tehát a pontcsoportnak vagy egy részcsoportjának megfelelő pontos koordináták beállítása is lehetséges. Először a szimmetriaelemek irány- és normálvektorainak optimalizálása szükséges, melyet az atomok szimmetriaelemekre történő projekciója és a szimmetriaelemek kívüli atomok helyvektorainak optimalizálása követ. Mivel a pontcsoport elemeinek helyzete (az egyes szimmetriaelemek által bezárt szög) minden esetben pontosan ismert, a szimmetriának megfelelő egzakt geometria generálására egy lehetőséget a szimmetriaelemek egyenként, a korábban megadott szimmetriaelemekhez viszonyítva történő meghatározása jelent. Ez az eljárás azonban minden pontcsoport esetén különbözőképp hajtható végre, melynek oka a csoportok eltérő szimmetriaelemeiben keresendő.

A szimmetrizálás általános, általunk implementált módszerére felírt pszeudokód az F1. ábrán tekinthető meg. Az eljárás során a főtengellyel párhuzamos vektorok helyett a főtengely irányvektorát alkalmazzuk (az F1. ábra 4. sor), míg az erre merőleges vektorok esetén – az ortogonalitás mellett – az egyes szimmetriaelemek által bezárt szögre is tekintettel kell lennünk. Egy tetszőleges vektor kiválasztása után (F1. ábra 7. sor) azt könnyen ortogonalizálhatjuk a főtengelyre (F1. ábra 8. sor). Az összes további főtengelyre merőleges vektor beállítása ezután ennek az ortogonalizált vektornak a segítségével történhet. Ezt a vektort az  $n_p$  rendű főtengely körül  $k \cdot 360^\circ / n_p$  ( $k \in \mathbb{Z}_{n_p}^+$ ) szöggel elforgatva (F1. ábra 9 – 10. sor) a főtengelyre merőleges, egymással megfelelő szöget bezáró vektorokat állíthatunk elő. Ha egy szimmetriaelem irány- vagy normálvektora egy ilyen elforgatott vektorral a választott tolerancián belül megegyezik, akkor a továbbiakban az eredetileg meghatározott vektor helyett ezt az új, optimalizált vektort használjuk (F1. ábra 11 – 15. sor). A pontos atomi helyvektorok előállításához az atomokat az azokat tartalmazó szimmetriaaltérbe vetítjük (F1. ábra 18 – 30. sor), majd ezen az új helyvektoron a pontcsoport minden elemét végrehajtjuk (a megfelelő műveletet a TRANSFORM eljárás végzi el) és a transzformált vektorokhoz legközelebb eső helyvektort erre a transzformált vektorra cseréljük (F1. ábra 31 – 35. sor).

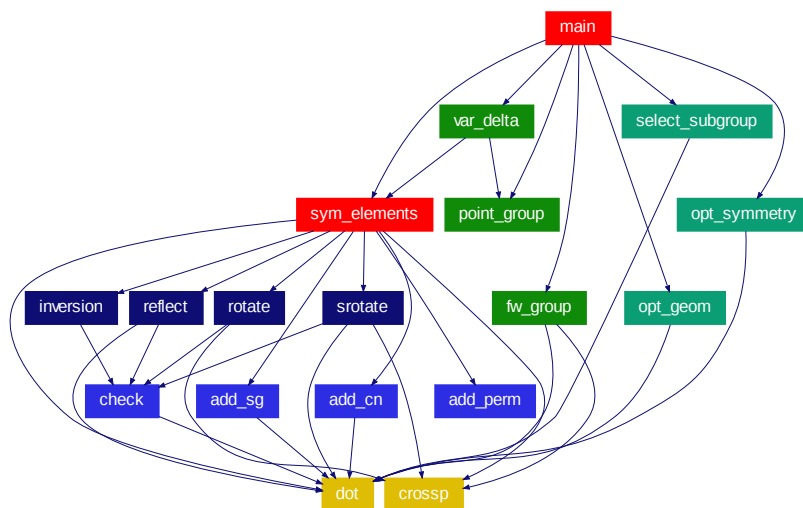
Az előzőekben ismertetett eljárás egyszerű ( $C_1$ ,  $C_s$ ,  $C_i$ ), lineáris ( $C_{\infty v}$ ,  $D_{\infty h}$ ) és főtengelyes ( $C_n$ ,  $C_{nv}$ ,  $C_{nh}$ ,  $D_n$ ,  $D_{nd}$ ,  $D_{nh}$ ,  $S_{2n}$ ), valamint oktaédes pontcsoportoknál ( $O$ ,  $O_h$ ) egyaránt alkalmazható. Mivel tetraédes ( $T$ ,  $T_d$ ,  $T_h$ ) és ikozaédes pontcsoportoknál ( $I$ ,  $I_h$ ) a szimmetriaelemek nem mind párhuzamosak egy további szimmetriaelemmel vagy merőlegesek egy másikra nézve, a korábban leírt módon, a szimmetriaelemeket az általuk bezárt szög ismeretében egyenként, egymáshoz viszonyítva adjuk meg.

### 5.1.2. Implementáció

Az 5.1.1. fejezetben ismertetett algoritmust Fortran nyelven implementáltuk.<sup>187</sup> A SYVA nevű program szerkezete annak hívási gráfiájával reprezentálva az 5.4. ábrán lát-

ható. Az alkalmazás különböző szintekre osztható, melyek az ábrán is megjelenítésre kerültek, továbbá a különböző funkciójú részeket eltérő szín jelöli.

A program három nagyobb egységre bontható, melyek a molekula szimmetriaelemeinek (`sym_elements` szubrutin), a különféle szimmetriával összefüggő tulajdonságainak (`var_delta`, `point_group`, `fw_group` szubrutinok) és a szimmetrizált geometriának (`select_subgroup`, `opt_symmetry`, `opt_geom` szubrutinok) az előállítását végzik.



5.4. ábra: A SYVA program hívási gráfja.

Az 5.1.1. fejezetben definiált lehetséges szimmetriaelemek (irány- és normálvektorok) előállítását a `sym_elements` szubrutin hajtja végre, mely ezután a megfelelő műveletet (`inversion`, `reflect`, `rotate`, `srotate` szubrutinok) is elvégzi. A transformált geometria eredetivel történő összevetését a `check` szubrutin valósítja meg, melyet azok egyezése után az adott szimmetriaművelet és az annak megfelelő permutáció `add_sg`, `add_cn`, és `add_perm` rutinok által véghezvitt tárolása követ.

A molekula pontcsoportját az egyes típusokba sorolható szimmetriaműveletek száma alapján a `point_group` szubrutin állapítja meg, mely ezután a szimmetriaműveletek pontos típusát ( $\sigma_d$  síkok, a főtengelyre merőleges  $C'_2$  és  $C''_2$  tengelyek) is meghatározza. A pontcsoport ismeretében a normálkoordináták által generált reprezentáció előállítását és redukcióját az ábrán fel nem tüntetett `mkrep` és `redrep` szubrutinok hajtják végre, így a normálrezgések szimmetriatulajdonságai is elérhetővé válnak. A szerkezet vázcsoportját az `fw_group` szubrutin állítja elő, míg a legmagasabb rendű pontcsoport meghatározását egy, az alapértelmezettnél lényegesen nagyobb tolerancia alkalmazásával a `var_delta` szubrutin végzi el. A pontcsoportnak a geometria optimalizálásához szükséges, felhasználó által választott részcsoportját a `select_subgroup` szubrutin generálja, majd a szimmetriaelemek és az atomi helyvektorok optimalizálását

az `opt_symmetry` és az `opt_geom` szubrutinok végzik. Megfigyelhető, hogy az alkalmazott vektoralgebrai módszernek megfelelően minden művelet a skaláris és vektori szorzatokat számító `dot` és `crossp` eljárásokra épül.

A kiindulási szerkezet torzultságát a program a szimmetriaelemek párhuzamostól, illetve merőlegetestől való eltéréseinek, azaz az irány- és normálvektorok egységnyi skaláris szorzatától vett különbségének vagy a skaláris szorzatuk abszolút értékének, valamint a szimmetriaelemeken fekvő atomok megfelelő szimmetriaelemtől mért távolságának maximumával jellemzi.

Az eljárás szükség szerint más szoftverekbe is beépíthető, melyet az MRCC<sup>24,172</sup> és a GAMESS(US)<sup>19,20</sup> programokba történő integrálással demonstráltunk. Ennek során a feladat alapvetően az egyes programok által használt adatok SYVA által alkalmazott formába való konvertálása, a legnagyobb szimmetria meghatározása és a pontcsoport, illetve a geometria keretprogram által használt alakba történő visszatranszformálása-ként definiálható. Az MRCC kód esetén ez a `pgroup` szubrutin rendszámok megadásával történő módosítását, majd a szimmetria meghatározása után a program által használt részcsoport kiválasztását igényli. A GAMESS szoftver esetén a a geometria korábbi beolvasása és a tömegközéppont origóba való eltolása is szükségessé válik, ezt a pontcsoport előállítását, majd a teljes geometriát generáló, szimmetria szempontjából egyedi atomok kiválasztása és azok a GAMESS által alkalmazott vonatkoztatási rendszerbe történő transzformációja követi.

A SYVA program forráskódja mint szabad szoftver, valamint az alkalmazott teszhalmaz korlátlanul hozzáférhető.<sup>188,189</sup>

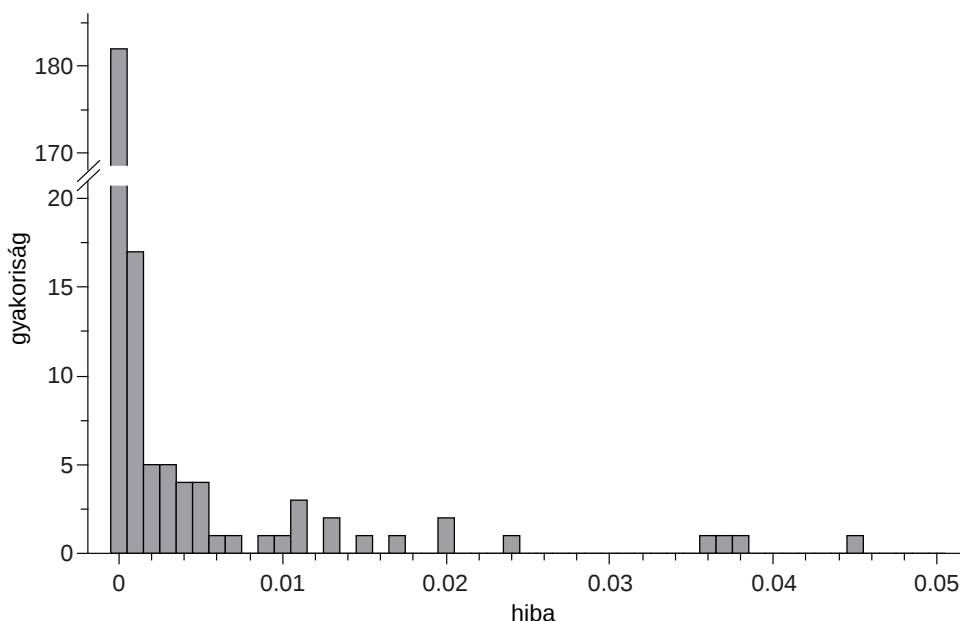
### 5.1.3. Jellemzés, összehasonlítás más szoftverekkel

A SYVA program teljesítőképeségének megállapításához egy 235 szerkezetből, főként létező molekulákból álló teszhalmazt hoztunk létre, mely 57 különböző pontcsoportot foglal magába. A halmaz összetétele mint az egyes pontcsoportokba sorolható struktúrák száma az F2. ábrán figyelhető meg. Általában elmondható, hogy a pontcsoportok reprezentáltsága a halmazban a csoportba sorolható molekulák számával arányos, például a főtengety rendjének növekedésével a szerkezetek gyakorisága a halmazban csökken. A ritkább (például T, T<sub>h</sub>, I, O) vagy egyszerűbb (C<sub>1</sub>, C<sub>s</sub>, C<sub>i</sub>) csoportokba szintén kevesebb elem tartozik. Egy-egy pontcsoportba átlagosan 4.1 struktúra sorolható, ennél kevesebb taggal a ritkábban előforduló pontcsoportok rendelkeznek: az egyelemű csoportok általában hét- vagy nyolcfogású szimmetriatengellyel rendelkeznek (kivéve C<sub>4h</sub>, C<sub>5h</sub> és D<sub>6d</sub>), a kételemű csoportok egy kivételtől eltekintve (D<sub>3</sub>) legalább negyedrendű tengelyt tartalmaznak. Bizonyos szerkezetek (például komplexek) esetén a molekulában egy, a teljes struktúrájánál magasabb szimmetriájú fragmens is található.

A teszhalmaz struktúráinak 5.1.2. fejezetben definiált hibája széles skálán mozog,

ennek az értéknek az eloszlása az 5.5. ábrán látható. Az ideális geometriától való átlagos eltérés  $1.77 \cdot 10^{-3}$  Å-nek adódik, míg a legnagyobb hiba értéke  $4.55 \cdot 10^{-2}$  Å, mindkét esetben a SYVA program  $10^{-3}$  Å nagyságú alapértelmezett toleranciájánál jelentősen nagyobb értékről beszélünk. Az alapértelmezett hibahatárnál nagyobb eltérés 31 esetben ( $\sim 13\%$ ) fordul elő, melyből a program ezzel a toleranciával 29-et nem tud meghatározni. Mivel a szerkezet hibája csak a szimmetriaelemek közötti eltérések és atom–szimmetriaelem távolságok maximumát adják meg, néhány esetben a pontcsoport meghatározása akkor is sikeres lehet, ha ez a hiba a megadott toleranciát meghaladja. Ilyen esetként említhető például a szimmetriaelemek ortogonalitásának hiánya vagy egy szimmetriaelem meghatározásához fel nem használt atomok adott szimmetriaelemtől vett nagyobb távolsága. Továbbá 15 esetben ( $\sim 6\%$ )  $10^{-2}$  Å mértékű, az alapértelmezett toleranciánál legalább egy nagyságrenddel nagyobb hiba is megfigyelhető.

Mivel a teszthalmaz megfelelő számban tartalmaz különféle pontcsoportba sorolható struktúrákat, emellett azok akár nagyobb mértékű hibával terheltek is lehetnek, elmondható, hogy alkalmas a szimmetria meghatározását célzó algoritmusok, szoftverek vizsgálatára.



5.5. ábra: A tesz geometriák hibájának eloszlása.

Az egyéb szoftverekkel történő összevetéshez tekintsünk három önálló, a szimmetria meghatározására szolgáló programot (SYMMOL,<sup>43–46</sup> SYMMETRY,<sup>190</sup> SYMMETRIZER<sup>47</sup>), továbbá három, a pontcsoportot is generáló kvantumkémiai alkalmazást (NWCHEM,<sup>22</sup> ORCA,<sup>23</sup> MRCC<sup>24,172</sup>).

5.1. táblázat: Egy molekulára vonatkozó átlagos futási idők a teljes teszhalmazra.

Program	SYVA	SYMMOL	SYMMETRY	SYMMETRIZER
Átlagos futási idő (ms)	8.8	9.3	51.7	311.8

Először a független programok átlagos futásidejét vizsgáltuk meg, a molekulamodellező szoftvereknél a számításokhoz szükséges egyéb műveletek elvégzése miatt erre nincs lehetőség. Ehhez a négy kódot azonos szinten optimalizálva újrafordítottuk és a teljes teszhalmazra lefuttattuk. Az átlagos gépidők az 5.1. táblázatban láthatók. Legrosszabbul 311.8 ms-mal a Java nyelven írt SYMMETRIZER teljesített, melyet 51.7 ms átlagos futási idővel a SYMMETRY nevű program követ. Utóbbi valószínűleg annak tudható be, hogy a szoftver minden esetben a szimmetriaelemek bizonyos optimalizálását is végrehajtja. A legjobb teljesítményt 8.8 és 9.3 ms átlagos gépidővel a SYVA és a SYMMOL programok nyújtották, bár a SYVA kismértékben (mindegy 5%-kal) gyorsabbnak bizonyult. A szimmetria gyorsabb detektálásának nagyszámú molekula analízise, szűrése esetén lehet jelentősége.

A továbbiakban a hét alkalmazás pontosságát vizsgáltuk, melyet a sikertelen meghatározások számával, valamint a helyesen megadott pontcsoportok sikeres szimmetriaanalízisekhez viszonyított százalékos arányával jellemeztünk. A szimmetria megállapítását akkor tekintettük sikertelennek, amikor az adott szoftver egyetlen pontcsoport detektálására sem volt képes, míg helyesnek a szerkezethez rendelhető legnagyobb pontcsoportot neveztük. Az eredmények az 5.2 táblázatban kerültek összefoglalásra.

5.2. táblázat: Az egyes programok pontossága és a sikertelen meghatározások száma.

Program	SYVA	SYMMETRY	SYMMETRIZER	SYMMOL	NWCHEM	ORCA	MRCC
Sikertelen <sup>a</sup>	0	0	0	15	35	52	0
Pontosság <sup>b</sup>	99.6%	97.5%	93.2%	95.0%	68.5%	96.2%	82.1%
Összesen <sup>c</sup>	99.6%	97.5%	93.2%	88.9%	58.3%	74.9%	82.1%

<sup>a</sup> A meg nem határozott pontcsoportok száma.

<sup>b</sup> A helyesen megadott pontcsoportok aránya a meghatározott pontcsoportok alapján.

<sup>c</sup> A helyesen megadott pontcsoportok aránya a teljes teszhalmazra nézve.

Mint látható, a független programok a pontcsoport előállítását az esetek nagy részében helyesen, 90% feletti pontossággal végezték. A SYMMOL-t kivéve, mely 15 esetben nem tudta megadni a szimmetriát, sikertelen analízis ezeknél a szoftvereknél nem fordult elő. A három molekulamodellező szoftver teljesítőképessége ebből a szempontból sokkal rosszabbnak mondható: az MRCC ugyan minden esetben szolgáltatott valamilyen pontcsoportot, ami viszont csak az esetek 82.1%-ában bizonyult helyesnek, az NWCHEM és az ORCA programok viszont 35 és 52 szerkezet esetén nem tudtak meghatározni semmilyen szimmetriát és a sikeres meghatározások is csupán 68.5%-ban, va-

lamint 96.2%-ban adódtak helyesnek. A teljes teszthalmazra vonatkoztatva a helyesen megadott pontcsoportok aránya a SYVA, a SYMMETRY, a SYMMETRIZER és a SYMMOL kódok esetén 99.6%, 97.5%, 93.2% és 88.9% körül mozog, valamint az NWCHEM, az ORCA és az MRCC programokat 58.3%, 74.9% és 82.1% pontossággal jellemezhetjük. Hibás és sikertelen meghatározásokra néhány példa az F1. táblázatban látható.

Megállapítható tehát, hogy az általunk felírt módszert megvalósító SYVA kód a vizsgált további hat szoftvert az alkalmazott teszthalmaz esetén mind futásidőben, mind pontosságban felülmúlja.

A SYMMETRIZER, a SYMMOL és a SYVA programok a talált szimmetriaelemeket is felsorolják, így azok megjeleníthetők vagy egyéb módon hasznosíthatók, továbbá a geometria szimmetrizálására is képesek, megkönnyítve ezzel a további számításokat. A SYVA szoftver ezenfelül a molekula vázcsoportjának megállapítására, a normálrezgések szimmetriaanalízisére és a legnagyobb szimmetria automatikus meghatározására is alkalmas, melyre a többi program nem képes.

Az 5.3. táblázatban a GAMESS használatával a HF/cc-pVDZ számításához néhány molekula esetén szükséges gépidő látható a szimmetria használata nélkül, illetve a SYVA integrálásával kapott pontcsoportot alkalmazva. Megfigyelhető, hogy a szimmetriát felhasználva a számítási idő már kis pontcsoportok esetén is (pentán,  $C_{2v}$ ) jelentősen lerövidíthető, mely a pontcsoport felhasználó általi definiálását sem igényli. A szimmetria egyik jelentősége tehát abban áll, hogy kihasználásával a kvantumkémiai számítások futásideje akár nagyságrendileg csökkenthető, még az általában alkalmazott legnagyobb Abel-részcsoport segítségével is.

5.3. táblázat: A HF/cc-pVDZ számítás futási ideje a GAMESS program használatával néhány molekulára.

Molekula	Számítási idő (s)		Pontcsoport <sup>c</sup>
	$C_1$ <sup>a</sup>	DET <sup>b</sup>	
Pentán	17.425	5.311	$C_{2v}$
Benzol	10.495	0.744	$D_{6h}$
$N_4S_4$	10.508	3.794	$D_{2d}$

<sup>a</sup> A szimmetria figyelmen kívül hagyásával ( $C_1$  pontcsoport alkalmazása).

<sup>b</sup> A SYVA által meghatározott pontcsoport alkalmazásával. A pontcsoportokat lásd a táblázat utolsó oszlopában.

<sup>c</sup> A SYVA által meghatározott pontcsoport.



## 5.2. A pentánkonformerek entalpiakülönbségei

### 5.2.1. Az entalpiakülönbségek kvantumkémiai meghatározása

#### 5.2.1.1. A referencia geometria előállítása

Egy korábbi tanulmány<sup>191</sup> igazolta, hogy a CCSD(T)/cc-pVQZ elméleti szinten optimalizált geometria megfelelő precizitással teszi lehetővé termokémiai számítások végrehajtását: a kötéstávolságokat 0.003 Å-nél, míg a kötésszögeket 0.5°-nál pontosabban adja vissza. Ezt a megállapítást azóta több munka<sup>160,192,193</sup> is megerősítette.

Az *n*-pentánkonformerek entalpiáinak meghatározására ezért CCSD(T) módszerrel optimalizált referencia geometriákat használtunk fel.<sup>183</sup> A geometriai optimalizálás és az energiajárulékok meghatározása során az összes elektron korrelációját figyelembe vettük. A báziskészlet szükséges méretének meghatározása érdekében a relatív teljes energiára legnagyobb hatást kifejtő energiajárulékokat a gg konformerre több, különböző bázison optimalizált geometria felhasználásával is előállítottuk, melynek eredménye az 5.4. táblázatban látható.

Megfigyelhető, hogy a kis bázison (cc-pVDZ) optimalizált geometrián számított hozzájárulások jelentősebb mértékben eltérnek a cc-pVQZ geometria használatával kapott eredményektől, egyes tagok esetén akár 150 cal mol<sup>-1</sup>-et megközelítő diszkrepancia is mutatkozik, míg e hozzájárulások összege is elérheti a 40 cal mol<sup>-1</sup> értéket.

A cc-pVTZ és a cc-pVQZ referencia geometrián nyert eredmények azonban sokkal közelebb esnek egymáshoz. Az ezeken a szerkezeteken számított eredmények legnagyobb hibája 20 – 25 cal mol<sup>-1</sup>-nek adódik, mely a  $\Delta E_{\text{HF}}$  és a  $\delta E_{\text{MP2}}$  értékek között figyelhető meg. A  $\delta E_{\text{CCSD}}$ , valamint a  $\delta E_{\text{(T)}}$  tagok a referencia cc-pVQZ értékeket még jobban megközelítik, a legnagyobb eltérés 6 cal mol<sup>-1</sup>. Ennek eredményeképp a cc-pVTZ, illetve a cc-pVQZ geometriákon számított tagok összege gyakorlatilag megegyezik, amennyiben a hozzájárulásokat legalább cc-pVTZ bázison számítjuk.

Következésképp elmondhatjuk, hogy a CCSD(T)/cc-pVDZ szinten számított egyensúlyi geometria alkalmazása a teljes energiában jelentős hibát eredményez, így termokémiai mennyiségek pontos meghatározására nem alkalmas. Ezzel ellentétben a CCSD(T)/cc-pVTZ geometria használatával kapott teljes energia a HF, az MP2, a CCSD és a CCSD(T) hozzájárulásokat figyelembe véve gyakorlatilag a referencia CCSD(T)/cc-pVQZ geometrián számítottal azonos. Ezért az egyes hozzájárulások számítását a CCSD(T)/cc-pVTZ szinten optimalizált geometrián végeztük. A felhasznált geometriák az F2. táblázatban láthatók.

5.4. táblázat: A geometria hatása a gg konformer  $\Delta E(\text{gg})$  relatív teljes energiájára.<sup>a</sup>

geometria	$X^b$	$\Delta E_{\text{HF}}$	$\delta E_{\text{MP2}}$	$\delta E_{\text{CCSD}}$	$\delta E_{(\text{T})}$	$\Sigma^c$
cc-pVDZ	2	2194	-1234	256	-114	1102
	3	2221	-1452	266	-159	876
	4	2242	-1404	283	-165	956
	5	2244	-1401	303	-167	966
	6	2246	-1412			969
cc-pVTZ	2	2340	-1381	285	-128	1117
	3	2338	-1600	301	-176	862
	4	2358	-1565	318	-183	928
	5	2361	-1565	324	-186	934
	6	2363	-1571			929
cc-pVQZ	2	2325	-1354	287	-123	1135
	3	2315	-1582	300	-170	862
	4	2335	-1546	317	-177	928
	5	2338	-1546	323	-180	934
	6	2339	-1552			929

<sup>a</sup> Minden érték cal mol<sup>-1</sup> egységekben szerepel.

<sup>b</sup> A hozzájárulások számításakor alkalmazott cc-pVXZ bázis kardinális száma.

<sup>c</sup> Az egyedi hozzájárulások összege.

### 5.2.1.2. Az eletronenergia számítása

A legjelentősebb hozzájárulások a konformerek teljes energiájához az F3. táblázatban tekinthetők meg. Elmondható, hogy a  $\Delta E_{\text{HF}}$  tag konvergenciája igen gyors, már cc-pVQZ bázis alkalmazásával is 5 cal mol<sup>-1</sup>-en belül konvergálnak tekinthető. A legjelentősebb különbségnek a legnagyobb bázisokkal számított eredmények között a gg konformernél megfigyelhető 2 cal mol<sup>-1</sup> számít. A gyors konvergencia miatt az ebből a tagból származó becsült hiba is igen alacsony, ennek nagysága rendre 1, 2 és 1 cal mol<sup>-1</sup> a tg, gg és gx konformerekre. A  $\Delta E_{\text{HF}}$  tag értéke a tg, gg, gx sorrendben  $1183 \pm 1$ ,  $2363 \pm 2$  és  $4279 \pm 1$  cal mol<sup>-1</sup>-ként adható meg.

A  $\delta E_{\text{MP2}}$  tag konvergenciája a tg konformernél a  $\Delta E_{\text{HF}}$  tag konvergenciájához hasonlóan gyors, már cc-pVTZ bázison számítva is mindössze 3 cal mol<sup>-1</sup>-gyel tér el a legjobb (5,6) extrapolált eredménytől. A többi konformer esetén viszont a konvergencia közel sem ilyen határozott, a cc-pV6Z bázissal kapott eredmények, valamint az (5,6) extrapolált érték különbsége 9, illetve 17 cal mol<sup>-1</sup> a gg és a gx konformerekre. A sokkal lassabb konvergencia miatt – a tg konformert leszámítva – a hozzájárulás hibája is komolyabb. A (4,5) és (5,6) extrapoláció különbségeként megadott becsült hiba rendre

1, 15 és 35 cal mol<sup>-1</sup> a tg, gg valamint gx konformereknél. A legjobb becslés a  $\delta E_{\text{MP2}}$  korrekcióra a tg, gg, gx sorrendben  $-615 \pm 1$ ,  $-1580 \pm 15$  és  $-1473 \pm 35$  cal mol<sup>-1</sup>.

A  $\delta E_{\text{CCSD}}$  és  $\delta E_{(\text{T})}$  tagok az előzőeknél egy nagyságrenddel kisebbek, továbbá értékük a bázis méretével monoton változik, a  $\delta E_{\text{CCSD}}$  hozzájárulás nagysága nő, míg a  $\delta E_{(\text{T})}$  korrekció csökken. Az extrapolált értékek a  $\delta E_{\text{CCSD}}$  korrekcióra 5, a  $\delta E_{(\text{T})}$  korrekció esetén pedig 2 cal mol<sup>-1</sup>-en belül konvergálnak tekinthetők. A hozzájárulások értéke a tg, a gg és a gx konformereknél a  $\delta E_{\text{CCSD}}$  tagra rendre  $117 \pm 5$ ,  $330 \pm 0$  és  $199 \pm 5$  cal mol<sup>-1</sup>, valamint  $-80 \pm 1$ ,  $-190 \pm 2$  és  $-204 \pm 1$  cal mol<sup>-1</sup> a  $\delta E_{(\text{T})}$  korrekcióra.

Érdemes megjegyezni, hogy míg a tg ég a gg konformereknél a  $\delta E_{\text{CCSD}}$  korrekció közelítőleg a  $\delta E_{(\text{T})}$  tag kétszeresének adódik, addig a gx konformernél a két érték közel azonos nagyságú, továbbá ellentétes előjelük miatt nagyjából kiejtik egymást.

A magasabbrendű hozzájárulások, a diagonális Born–Oppenheimer korrekció, illetve a skaláris relativisztikus korrekciók értékei az 5.5. táblázatban láthatók. Ezek a hozzájárulások a tg, a gg és a gx konformerek esetén rendre  $-1$ ,  $-7$  valamint  $-5$  cal mol<sup>-1</sup> értéket vesznek fel. Az egyes tagok abszolút értékeinek összegeként definiált hibák a tg, gg, gx sorrendben 13, 17, illetve 19 cal mol<sup>-1</sup>-ként határozhatók meg. A tg és a gx konformerekre alacsonyabb szimmetriájuk miatt a  $\delta E_{(\text{Q})}$  tag számítására nem volt lehetőségünk, a gg konformerre számított 9 cal mol<sup>-1</sup> eredmény alapján ezek értékét 10 cal mol<sup>-1</sup>-ként közelítettük, melyet csak a hiba becslésekor használtunk fel.

5.5. táblázat: Magasabbrendű hozzájárulások az  $n$ -pentánkonformerek relatív energiáikhoz.<sup>a</sup>

	$\delta E_{\text{T}}$	$\delta E_{(\text{Q})}$	$\delta E_{\text{DBOC}}$		$\delta E_{\text{SR}}$		$\sum$	$\sigma^{\text{b}}$
			HF	CCSD	HF	CCSD(T)		
tg	-1	- <sup>c</sup>	-1	0	1	0	-1	13
gg	2	-9	-3	0	3	0	-7	17
gx	-5	- <sup>c</sup>	-1	1	1	-1	-5	19

<sup>a</sup> Minden érték cal mol<sup>-1</sup> egységekben szerepel.

<sup>b</sup> Az egyedi hozzájárulások abszolút értékeinek összege.

<sup>c</sup> Nem került meghatározásra. A hiba számításakor a gg konformerhez tartozó érték alapján 10 cal mol<sup>-1</sup>-ként lett közelítve.

### 5.2.1.3. A zérus ponti rezgési energia és a hőmérsékleti korrekció

A zérus ponti energia meghatározásához számított hozzájárulások az 5.6. táblázatban láthatók.

A harmonikus rezgési hozzájárulás számítása CCSD(T)/cc-pVTZ szinten történt a belső rotációkként kezelt frekvenciák eltávolításával. A felhasznált rezgési frekvenciák az F4. táblázatban találhatóak.

A pontosabb leírás érdekében MP2/cc-pVDZ módszerrel anharmonikus korrekciót végeztünk, valamint a metil és a C–C–C–C torziókat gátolt belső rotációkként kezeltük. Az anharmonikus rezgési korrekció számítása nagyobb bázison (cc-pVTZ) csak a C<sub>2</sub> szimmetriával rendelkező gg konformerre volt kivitelezhető, a másik két konformerre a cc-pVTZ eredmények összegének számításakor a cc-pVDZ bázison becsült anharmonikus korrekciós értékeket használtuk fel. A gátolt rotorokra MP2/cc-pVTZ energiák illesztéséből származó potenciállal kapott sajátértékek az F5. és az F6. táblázatokban kerültek feltüntetésre.

5.6. táblázat: A harmonikus és az anharmonikus zérusponti energia, valamint a gátolt rotációra történő korrekció hozzájárulása az *n*-pentán konformereinek relatív teljes energiáihoz.<sup>a</sup>

<i>X</i> <sup>f</sup>	harmonikus <sup>b</sup>			anharmonikus <sup>c</sup>			gátolt rotáció <sup>d</sup>			$\Sigma$ <sup>e</sup>		
	tg	gg	gx	tg	gg	gx	tg	gg	gx	tg	gg	gx
2	71	216	208	−4	−21	−23	−39	−17	−45	28	178	139
3	87	241	234		−46		−44	−12	−32	39 <sup>g</sup>	183	179 <sup>g</sup>
$\sigma^h$	16	25	26	4	24	23	5	5	13	25	54	63

<sup>a</sup> Minden érték cal mol<sup>−1</sup> egységekben szerepel.

<sup>b</sup> CCSD(T) módszerrel számítva, a belső rotációk eltávolítása után.

<sup>c</sup> MP2 módszerrel számítva, az anharmonicitási mátrix belső rotációknak megfelelő diagonális elemeinek eltávolítása után.

<sup>d</sup> MP2 módszerrel meghatározva.

<sup>e</sup> Az egyedi hozzájárulások összege.

<sup>f</sup> Az alkalmazott cc-pVXZ bázis kardinális száma.

<sup>g</sup> Az anharmonikus korrekcióhoz a cc-pVDZ eredményeket felhasználva.

<sup>h</sup> A cc-pVTZ és a cc-pVDZ eredmények abszolút különbsége. Az anharmonikus korrekcióra a tg és a gx konformerek esetén a cc-pVDZ eredmény abszolút értéke.

Az utóbbi két hozzájárulás mértéke viszonylag nagy, mintegy 40 – 50 cal mol<sup>−1</sup>. A teljes ZPVE értéke akár a  $\delta E_{\text{CCSD}}$  vagy a  $\delta E_{(\text{T})}$  korrekciókkal összemérhető, továbbá jelentős hibával terhelt, különösen a harmonikus és az anharmonikus tagokat tekintve, a teljes energia bizonytalanságának mintegy fele ebből a hozzájárulásból származik. Az elvégzett számítások alapján legjobb becslésként a  $\Delta E_{\text{ZPVE}}$  tagra a tg, a gg és a gx konformerekre rendre  $39 \pm 25$ ,  $183 \pm 54$  és  $179 \pm 63$  cal mol<sup>−1</sup> határozhatók meg.

Néhány korábbi kísérleti vizsgálat során alkalmazott hőmérséklettartományra a 4.18 egyenlet alapján számított hőmérsékleti korrekció értéke az 5.7. táblázatban látható. A forgási állapotösszeg számításához használt CCSD(T)/cc-pVTZ geometriákból előállított forgási állandók az 5.8. táblázatban találhatók.

5.7. táblázat: A hőmérsékleti korrekció értéke az egyes  $n$ -pentánkonformerek relatív entalpiáihoz ( $\Delta H_T^{korr}$ ).<sup>a</sup>

geometria	T / K	$\Delta H_T^{korr}$ (tg)	$\Delta H_T^{korr}$ (gg)	$\Delta H_T^{korr}$ (gx)
CCSD(T)/cc-pVDZ	197 <sup>b</sup>	-3	-76	-99
	298	-9	-116	-226
	385 <sup>c</sup>	-17	-139	-366
	412 <sup>d</sup>	-20	-144	-413
CCSD(T)/cc-pVTZ	197 <sup>b</sup>	-5	-82	-110
	298	-15	-126	-242
	385 <sup>c</sup>	-27	-150	-382
	412 <sup>d</sup>	-30	-155	-428
hiba <sup>e</sup>	197 <sup>b</sup>	2	6	11
	298	6	10	16
	385 <sup>c</sup>	10	11	16
	412 <sup>d</sup>	10	11	15

<sup>a</sup> Minden érték cal mol<sup>-1</sup> egységekben szerepel.

<sup>b</sup> A 180 referenciában alkalmazott hőmérséklettartomány felezőpontja.

<sup>c</sup> A 181 referenciában alkalmazott hőmérséklettartomány felezőpontja.

<sup>d</sup> A 182 referenciában alkalmazott hőmérséklettartomány felezőpontja.

<sup>e</sup> A cc-pVTZ és a cc-pVDZ eredmények különbségének abszolút értéke.

5.8. táblázat: Az  $n$ -pentánkonformerek CCSD(T)/cc-pVTZ geometrián meghatározott forgási állandói.

konformer	forgási állandók / GHz		
tt	1.862	1.964	17.138
tg	2.729	3.017	7.630
gg	2.139	2.318	11.791
gx	2.551	3.230	7.315

#### 5.2.1.4. Összehasonlítás korábbi számítási eredményekkel

A korai *ab initio* számítási eredmények<sup>194–197</sup> a Hartree-Fock módszeren alapultak. Ezek a számítások értelemszerűen sokkal pontatlanabbak, mint a későbbi tanulmányok, ezért ezek részletes bemutatásától eltekintünk.

A további, korrelációs számítási eredmények az alkalmazott elméleti szinttől függően három csoportra oszthatók. Az első (A) csoportba a korai másod-, harmad-, illetve negyedrendű perturbációs módszereket alkalmazó vizsgálatok tartoznak. Ebbe

a csoportba sorolható a Wiberg és munkatársai<sup>198</sup> által 1988-ban közölt tanulmány. A szerzők az egyes konformerekre MP3/6-31G(d)//HF/6-31G(d) szinten  $\Delta E(\text{tg}) = 760 \text{ cal mol}^{-1}$ ,  $\Delta E(\text{gg}) = 1360 \text{ cal mol}^{-1}$ , valamint  $\Delta E(\text{gx}) = 3330 \text{ cal mol}^{-1}$  értékeket közöltek. Érdeemes megemlíteni, hogy a gx konformert a cikkben hibásan  $g^+g^-$ -ként jelölték meg, továbbá két átmeneti állapot is konformerként szerepel.

Tsuzuki és munkatársai<sup>199</sup> az  $n$ -pentánkonformerek relatív energiáit MP4(SDQ)/6-31G(d)//HF/6-31G(d) szinten határozták meg, melyekre a  $\Delta E(\text{tg}) = 740 \text{ cal mol}^{-1}$ , a  $\Delta E(\text{gg}) = 1302 \text{ cal mol}^{-1}$ , valamint a  $\Delta E(\text{gx}) = 3289 \text{ cal mol}^{-1}$  értékeket kapták. Az elektronenergiák különbsége jól egyezik a Wiberg és munkatársai által közölt eredményekkel, azoknál  $20 - 60 \text{ cal mol}^{-1}$ -gyel alacsonyabb. A HF/6-31G(d) szinten meghatározott ZPVE és hőmérsékleti korrekciós értékekkel együtt az entalpiakülönbségekre 0 K hőmérsékleten a  $\Delta H_0^\circ(\text{tg}) = 837 \text{ cal mol}^{-1}$ ,  $\Delta H_0^\circ(\text{gg}) = 1541 \text{ cal mol}^{-1}$  és  $\Delta H_0^\circ(\text{gx}) = 3496 \text{ cal mol}^{-1}$ , valamint 298 K-en  $\Delta H_{298}^\circ(\text{tg}) = 800 \text{ cal mol}^{-1}$ ,  $\Delta H_{298}^\circ(\text{gg}) = 1431 \text{ cal mol}^{-1}$  és  $\Delta H_{298}^\circ(\text{gx}) = 3424 \text{ cal mol}^{-1}$  értékek adódnak.

Mirkin és Krimm<sup>200</sup> a pentánkonformerek teljes energiáját MP2/6-31G(d) szinten számították. A tg, a gg és a gx konformerekre rendre 670, 1090 és 3190  $\text{cal mol}^{-1}$  értékeket közöltek. Ezek az eredmények a módszer nagyobb pontatlansága miatt már nagyobb mértékben eltérnek a korábbi értékektől. A legjelentősebb különbség ( $200 - 300 \text{ cal mol}^{-1}$ ) a gg konformernél mutatkozik.

A következő (B) csoportba sorolt eredmények kis, kétszeres  $\zeta$  minőségű bázissal végrehajtott csatolt klaszter számításokon alapulnak. Salam és Deleuze<sup>201</sup> egy fókuszpont analízis során B3LYP/6-311++G(d,p) geometriákon HF/cc-pVQZ, MP3/cc-pVTZ, valamint CCSD(T)/aug-cc-pVDZ energiaszámításokat hajtottak végre, és a  $\Delta E(\text{tg}) = 621 \text{ cal mol}^{-1}$ ,  $\Delta E(\text{gg}) = 1065 \text{ cal mol}^{-1}$  és  $\Delta E(\text{gx}) = 2917 \text{ cal mol}^{-1}$  értékeket közölték. Az  $x^+g^-$  konformert ebben az esetben is – hibásan –  $g^+g^-$ -ként tüntették fel. A B3LYP/6-311++G(d,p) rezgési analízis eredményei segítségével az entalpiakülönbségek is kiszámíthatók, melyekre  $\Delta H_0^\circ(\text{tg}) = 676 \text{ cal mol}^{-1}$ ,  $\Delta H_0^\circ(\text{gg}) = 1423 \text{ cal mol}^{-1}$  és  $\Delta H_0^\circ(\text{gx}) = 3111 \text{ cal mol}^{-1}$ , illetve  $\Delta H_{298}^\circ(\text{tg}) = 871 \text{ cal mol}^{-1}$ ,  $\Delta H_{298}^\circ(\text{gg}) = 1771 \text{ cal mol}^{-1}$  és  $\Delta H_{298}^\circ(\text{gx}) = 3476 \text{ cal mol}^{-1}$  adódik.

Klauda és munkatársai<sup>202</sup> MP2/cc-pVDZ geometrián végeztek CCSD(T)/cc-pVDZ és MP2/cc-pVQZ energiaszámításokat. Az MP2:CC névvel jelölt, CCSD(T)/cc-pVQZ eredményeket közelítő módszer a tg, gg, illetve gx konformerek energiáinak különbségére rendre 622, 985 és 2946  $\text{cal mol}^{-1}$  értékeket ad. A szerzők a tg konformerre legpontosabb eredményként a  $\Delta E(\text{tg}) = 618 \text{ cal mol}^{-1}$  értéket adták meg, mely CCSD(T)/cc-pVDZ és MP2/cc-pV5Z számításokon alapul.

Martin és munkatársai<sup>203</sup> egy W1h protokollhoz hasonló eljárás segítségével a tg, gg, valamint gx konformerek relatív energiáit CCSD/cc-pV(T,Q)Z és CCSD(T)/cc-pV(D,T)Z extrapolált eredmények alapján 614, 961 és 2813  $\text{cal mol}^{-1}$ -ként becsülték.

Az utolsó (C) csoportba nagyobb, háromszoros  $\zeta$  minőségű bázisok alkalmazásával

elvégzett csatolt klaszter számítások tartoznak.

A CCSD(T)/cc-pVTZ szintű  $\Delta E(\text{tg}) = 581 \text{ cal mol}^{-1}$ ,  $\Delta E(\text{gg}) = 912 \text{ cal mol}^{-1}$  és  $\Delta E(\text{gx}) = 2763 \text{ cal mol}^{-1}$  energiakülönbségek, továbbá a CCSD(T)-F12b/cc-pVTZ-F12//SCS-MP2/cc-pVTZ szinten meghatározott  $\Delta E(\text{tg}) = 582 \text{ cal mol}^{-1}$ ,  $\Delta E(\text{gg}) = 915 \text{ cal mol}^{-1}$  és  $\Delta E(\text{gx}) = 2767 \text{ cal mol}^{-1}$  értékek szintén Martin és munkatársai nevéhez köthetők. Látható, hogy az utóbbi két eredmény, az alkalmazott módszer különbsége és a geometria eltérése ellenére is rendkívül jó egyezést mutat. Az utóbbi eredményekhez az SCS-MP2/cc-pVTZ módszerrel előállított ZPVE-t és hőmérsékleti korrekciót hozzávéve a  $\Delta H_0^\circ(\text{tg}) = 682 \text{ cal mol}^{-1}$ ,  $\Delta H_0^\circ(\text{gg}) = 1205 \text{ cal mol}^{-1}$  és  $\Delta H_0^\circ(\text{gx}) = 3037 \text{ cal mol}^{-1}$ , illetve  $\Delta H_{298}^\circ(\text{tg}) = 640 \text{ cal mol}^{-1}$ ,  $\Delta H_{298}^\circ(\text{gg}) = 1052 \text{ cal mol}^{-1}$  és  $\Delta H_{298}^\circ(\text{gx}) = 2911 \text{ cal mol}^{-1}$  értékek adódnak.

Az eddigi elméleti eredmények az F7. táblázatban kerültek összefoglalásra. A magasabb relatív energiákból látható, hogy a perturbációs módszerek a tt konformer stabilitását a csatolt klaszter módszerhez képest túlbecsülik. A relatív entalpiák az A csoportban a B csoportbelieknél nagyjából  $50 - 150 \text{ cal mol}^{-1}$ -gyel magasabbak. Az MP2 eredmények valamivel alacsonyabbak a magasabb rendű módszerekkel elérteknél, így – valószínűleg mivel a hibák szerencsés módon részben kiejtik egymást – ezek az eredmények kissé közelebb esnek a pontosabb CC értékekhez. A gg konformer stabilitását az MP $n$  módszerek a B csoportba sorolt módszerekhez képest mintegy  $300 - 350 \text{ cal mol}^{-1}$ -gyel, a gx konformerét  $300 - 500 \text{ cal mol}^{-1}$ -gyel is alábecsülik. Az eredményeken az MP $n$ -sorozat jellemző oszcillációja is jól megfigyelhető. Látható továbbá, hogy a B és a C csoportba sorolt eredmények külön-külön jó egyezést mutatnak a Salam és Deleuze által közölt értékektől eltekintve, melyek valamivel magasabbak a többinél. Ezt az eltérést valószínűleg a termokémiai mennyiségek utóbbi tanulmányban való hibás számítása okozza.

Saját eredményeinkkel mint referenciával összevetve a korábbi értékeket megfigyelhető, hogy egyes módszerek bizonyos konformerek esetén az entalpiakülönbségek becslésére megfelelőbbek: Martin és munkatársai W1h-szerű modellje például a tg és gx konformerek stabilitását közelíti jól, azonban  $44 \text{ cal mol}^{-1}$ -gyel alábecsüli a gg konformerét, ellenben a CCSD(T)-F12b módszerrel kapott eredmények a gg konformerre az általunk elvégzett fókuszpont analízisből származó adatokkal jól egyeznek, viszont a tg és a gx konformerek stabilitása  $23$  és  $29 \text{ cal mol}^{-1}$ -gyel eltér.

Kijelenthető tehát, hogy az  $n$ -pentán konformereinek egységes kezelése egyáltalán nem egyszerű feladat. A tg konformer esetén az egyes hozzájárulások konvergenciája gyorsabban elérhető, mint a gg és a gx konformerek esetén, melynek oka feltételezhetőleg a tg és a tt konformer szerkezeti hasonlósága. Az  $n$ -pentánnál a zérusponti rezgési energia számítása is kihívást jelent, emiatt a végső értékekhez jelentős hibával járul hozzá. Mindezek ellenére elmondható, hogy a saját fókuszpont analízisen alapuló eredményeink a korábbi számítási eredményeket pontosságban felülmúlják, valamint

azokhoz megfelelő, realiztikus hibahatár is tartozik. Legpontosabb eredményeknek az 5.9. táblázatban összefoglalt értékeket tekinthetjük.

5.9. táblázat: Az  $n$ -pentán konformereinek kvantumkémiailag számított entalpiakülönbségei

konformer	$\Delta H_0^\circ/\text{cal mol}^{-1}$	$\Delta H_{298}^\circ/\text{cal mol}^{-1}$
tg	$644 \pm 46$	$628 \pm 52$
gg	$1099 \pm 90$	$974 \pm 100$
gx	$2975 \pm 124$	$2733 \pm 140$

### 5.2.2. Az entalpiakülönbségek meghatározása kísérleti adatok alapján

Az  $n$ -pentán konformerek hőmérsékletfüggő entalpiakülönbségeinek kísérleti meghatározása az elmúlt évtizedekben több tanulmány tárgyát képezte. A legtöbb esetben folyadékfázisú eredményeket<sup>174-179,204</sup> közöltek, míg a gázfázisú mérések<sup>180-182</sup> száma kisebb. Az eredmények nagy része<sup>174,175,177,178,180,182</sup> infravörös-vagy Raman-spektroszkópiás mérésekből származik, melyek kiértékelése általában a 4.30 egyenlet segítségével történik. A különböző hőmérsékleteken felvett spektrumok alapján az egyes konformerekhez tartozó sávok számított intenzitásainak arányait az abszolút hőmérséklet reciprokának függvényeként illetve az egyenes meredekségéből ( $a$ ) a megfelelő konformerek keresett entalpiakülönbsége ( $\Delta H^\circ = -aR$ ) kifejezhető.

5.10. táblázat: A kísérleti eredmények meghatározásához felhasznált hullámszámok

hivatkozás	tt	tg	gg
	$\text{cm}^{-1}$		
174	865	838	–
178	921	909, 764	–
181	401	336	–
176	401	336	384
182	399	331	380
180	399.0	328.9	267.1

Az ismertetett módszer alkalmazhatóságához az egyes konformerekhez rendelhető, jól elkülönülő vagy könnyen felbontható rezgési sávok megjelenése szükséges. Az átmenetek kiválasztása nem egyértelmű, az egyes szerzők az intenzitások számításakor eltérő hullámszámokat alkalmaznak, melyek értékei az 5.10. táblázatban szerepelnek. A régebbi tanulmányokban<sup>174,178</sup> a gg konformer relatív entalpiájának meghatározá-



sához nem találtak megfelelő átmenetet, míg a gx konformerre eddig nem közöltek eredményeket.

Az első gázfázisú eredményeket Maissara és munkatársai közölték<sup>181</sup> 1983-ban. A különböző hőmérsékleteken felvett Raman-spektrumokból meghatározott tt–tg entalpiakülönbségre  $\Delta H_T^\circ(\text{tg}) = 560 \pm 100 \text{ cal mol}^{-1}$  értéket kaptak. Kanésaka és munkatársai<sup>182</sup> ezt az entalpiakülönbséget  $465 \pm 30 \text{ cal mol}^{-1}$ -ben határozták meg. A szerzők által megadott hiba valószínűleg az entalpiakülönbség szórásának felel meg, melyet a többi tanulmányban megadott 95%-os konfidencia intervallumra átszámítva  $465 \pm 68 \text{ cal mol}^{-1}$  érték adódik. A két mérést hasonló hőmérséklettartományban (337 – 433 K és 316 – 508 K) végezték, így azok közvetlenül összehasonlíthatók. A két érték nagymértékben eltér, azonban a megadott bizonytalanságok átfednek, mely az első mérés jelentős hibájának ( $100 \text{ cal mol}^{-1}$ ) tudható be.

A második tanulmányban említésre kerül az is, hogy a nagyobb résszélességgel, alacsony hőmérsékleten meghatározott – bizonytalanabb – intenzitásokat is felhasználva a Maissara és munkatársai által közölt  $560 \text{ cal mol}^{-1}$  értékkel konzisztens eredmény nyerhető. További hibalehetőség a  $380 \text{ cm}^{-1}$ -nél megjelenő, gg konformerhez rendelhető átmenet, mely valószínűleg hatással van a mért  $I_{\text{tt}}$  értékekre. Ez a mért intenzitás növekedéséhez és ezáltal a számított entalpiakülönbség csökkenéséhez vezet.

A legpontosabb értékeket az *n*-pentán konformációs entalpiakülönbségeire Balabin közölte,<sup>180</sup> a tg konformer relatív entalpiáját  $\Delta H_T^\circ(\text{tg}) = 618 \pm 6 \text{ cal mol}^{-1}$ -ként határozta meg. Az egyetlen gázfázisú kísérleti eredmény a gg konformerre szintén ebben a tanulmányban olvasható, melyre a  $\Delta H_T^\circ(\text{gg}) = 940 \pm 20 \text{ cal mol}^{-1}$  érték adódik. Mivel ez a mérés alacsonyabb hőmérsékleten (143 – 250 K) történt, a fenti eredményekkel nem összehasonlítható. Megfelelő kezelés nélkül a  $380 \text{ cm}^{-1}$ -nél megjelenő átmenet azonban ebben az esetben is hibát okozhat.

5.11. táblázat: A kísérletileg meghatározott gázfázisú entalpiakülönbségek

hőmérséklet- tartomány / K	kísérleti eredmények		saját számítások <sup>a</sup>	
	$\Delta H^\circ(\text{tg})$	$\Delta H^\circ(\text{gg})$	$\Delta H^\circ(\text{tg})$	$\Delta H^\circ(\text{gg})$
	cal mol <sup>-1</sup>		cal mol <sup>-1</sup>	
337 – 433	$560 \pm 100$ <sup>b</sup>	–	$617 \pm 55$	–
316 – 508	$465 \pm 68$ <sup>c</sup>	–	$613 \pm 55$	–
143 – 250	$618 \pm 5$ <sup>d</sup>	$953 \pm 15$ <sup>d</sup>	$638 \pm 47$	$1017 \pm 97$

<sup>a</sup> A hőmérséklettartomány felezőpontjában számítva.

<sup>b</sup> 181 hivatkozás

<sup>c</sup> 182 hivatkozás, a 95%-os konfidencia intervallumra történő átszámítás után

<sup>d</sup> 180 hivatkozás, revízió a hibaterjedés figyelembevételével

A gázfázisú kísérleti entalpiakülönbségek a saját kvantumkémiailag meghatározott eredményeinkkel együtt az 5.11. táblázatban találhatóak.

Látható, hogy az elméleti értékek (613 és 617 cal mol<sup>-1</sup>) a korábbi tanulmányok közül inkább a Maissara és munkatársai által közölt 560 cal mol<sup>-1</sup> eredményt támasztják alá, míg a legpontosabb, Balabin által közölt entalpiakülönbségekkel jó egyezést mutatnak.

Meg kell jegyeznünk azonban, hogy Balabin az adatok kiértékelésekor szintén a 4.30 egyenletet alkalmazta az adatok súlyozása nélkül. A linearizálás viszont torzítja a mérési pontok szórását, ennek korrekciója nélkül pedig a végeredményre is hatással van. Ezért célszerű az adatok újbóli, korrekt kiértékelése, figyelembe véve a logaritmizálásból származó hibaterjedést is. Az experimentális hibát eredetileg homoszkedasztikusnak ( $\sigma_{I_{\alpha\beta}/I_{tt}}^2(T) = \sigma_{I_{\alpha\beta}/I_{tt}}^2$ ) feltételezve, a pontok transzformáció utáni  $\sigma_{\ln(I_{\alpha\beta}/I_{tt})}^2$  szórásnégyzete a 4.32 egyenlet szerint számítható:

$$\sigma_{\ln(I_{\alpha\beta}/I_{tt})}^2 = \frac{I_{tt}}{I_{\alpha\beta}} \sigma_{I_{\alpha\beta}/I_{tt}}^2 \quad (5.11)$$

Az illesztésekhez felhasznált adatok az F8. és az F9. táblázatokban láthatók. Az 5.11 egyenlet szerint számított szórásnégyzetek felhasználásával végzett súlyozott lineáris, illetve nemlineáris regresszió, valamint a közöséges lineáris regresszió eredményei az 5.12. táblázatban szerepelnek. A feltüntetett hibák a 95%-os konfidencia intervallumnak felelnek meg.

5.12. táblázat: Az *n*-pentánkonformerek relatív entalpiái kísérleti adatok alapján

konformer	módszer <sup>a</sup>	$\Delta H_T^\circ$ / cal mol <sup>-1</sup>
tg	OLLS	618 ± 6
	WLLS	618 ± 5
	ONLS	618 ± 5
gg	OLLS	943 ± 24
	WLLS	958 ± 13
	ONLS	953 ± 15

<sup>a</sup> OLLS: közöséges lineáris legkisebb négyzetek módszere, WLLS: súlyozott lineáris legkisebb négyzetek módszere, ONLS: közöséges nemlineáris legkisebb négyzetek módszere

Az OLLS illesztések eredménye (618 ± 6 és 943 ± 24 cal mol<sup>-1</sup>) gyakorlatilag megegyezik a Balabin által közölt értékekkel (618 ± 6 és 940 ± 20 cal mol<sup>-1</sup>). A tg konformer esetén a WLLS és az ONLS illesztések is a korábban meghatározott 618 cal mol<sup>-1</sup> értéket eredményezik, viszont azok 5 cal mol<sup>-1</sup> becsült hibája az eredetileg megadotthoz

képest kissé alacsonyabb. A gg konformernél azonban a súlyozott lineáris és a nemlineáris regressziók végeredménye a közölt értéktől nagyobb mértékben tér el. A WLLS regresszió 958, míg az ONLS módszerrel történő illesztés 953 cal mol<sup>-1</sup> értéket szolgáltat, melyek a Balabin által eredetileg közölt eredménytől 1.9%-ban, illetve 1.4%-ban különböznek. A WLLS módszerrel végzett kiértékelés esetén a becsült paraméter hibája kisebb, mint az eredeti értéktől való eltérés, csak a hibasávok fednek át. A mérések – becsült hibákból is látható – jelentős pontossága nem ad okot ekkora mértékű eltérésre, az csak a hibás kiértékelésből származhat. Elmondható tehát, hogy a mérés hőmérséklettartományában (143 – 250 K) a legjobb becslés az *n*-pentán konformereinek relatív entalpiáira  $\Delta H_T^\circ(\text{tg}) = 953 \pm 15$  cal mol<sup>-1</sup>, valamint  $\Delta H_T^\circ(\text{gg}) = 618 \pm 5$  cal mol<sup>-1</sup>, mely értékek az eredeti, nemlineáris függvény illesztésével (ONLS) határozhatók meg.

Az 5.11. táblázatból látható, hogy az általunk az adott hőmérséklettartományra meghatározott értékek a tg konformer esetén jól egyeznek a kísérleti adatokkal, a különbség mindössze 20 cal mol<sup>-1</sup>, ezzel szemben a gg konformer esetén az eltérés a 64 cal mol<sup>-1</sup> értéket is eléri. Az eltérés oka a korábban említett, gg konformerhez rendelhető, 380 cm<sup>-1</sup> hullámszámnál megjelenő sáv lehet. Bár Balabin a kiértékelés előtt a spektrumok dekonvolúcióját is elvégezte, ennek pontos módja (például a sávok pozíciói) nem ismertek. Ha a 380 cm<sup>-1</sup>-nél megjelenő sáv a kiértékelés során nem lett figyelembe véve, akkor a tt konformerhez tartozó intenzitások hibásan kerültek kiszámításra. Kanésaka és munkatársai<sup>182</sup> ennek az átmenetnek a hozzájárulását a tg konformer relatív entalpiájához 36 cal mol<sup>-1</sup>-nek becsülték. Érdeemes tehát ennek a sávnak a hozzájárulására korrekciót végezni.

Az említett átmenetre történő korrekció nélkül a gg konformer relatív entalpiájára 4.29 egyenlet alapján a

$$\ln \frac{I(\text{gg})}{I'(\text{tt})} = \ln \frac{I(\text{gg})}{I(\text{tt}) + I'(\text{gg})} = \ln \frac{I(\text{gg})}{I(\text{tt}) \left[ 1 + \exp \left( -\frac{\Delta H^\circ(\text{gg})}{RT} + b'_{\text{gg}} \right) \right]} \quad (5.12)$$

összefüggés írható fel, ahol  $I'(\text{tt})$  a tt konformerhez rendelt sáv  $I(\text{tt})$  intenzitásának és a 380 cm<sup>-1</sup>-es gg sáv  $I'(\text{gg})$  intenzitásának összege,  $\Delta H^\circ(\text{gg})$  pedig a gg konformer tényleges relatív entalpiája. Az 5.12 egyenletet átrendezve és a 4.30 egyenletet figyelembe véve a gg konformer relatív entalpiája a következő összefüggés segítségével határozható meg:

$$\ln \frac{I(\text{gg})}{I(\text{tt})} = -\frac{\Delta H^\circ(\text{gg})}{R} \frac{1}{T} + b_{\text{gg}} = \ln \frac{I(\text{gg})}{I'(\text{tt})} + \ln \left[ 1 + \exp \left( -\frac{\Delta H^\circ(\text{gg})}{RT} + b'_{\text{gg}} \right) \right] \quad (5.13)$$

Az 5.13 egyenletben szereplő  $b_{\text{gg}}$  és  $b'_{\text{gg}}$  tagok nem függetlenek. A közöttük fennálló

reláció a két, gg konformerhez tartozó sávra felírt

$$\ln \frac{I(\text{gg})}{I(\text{tt})} = -\frac{\Delta H^\circ(\text{gg})}{R} \frac{1}{T} + b_{\text{gg}} \quad (5.14)$$

$$\ln \frac{I'(\text{gg})}{I(\text{tt})} = -\frac{\Delta H^\circ(\text{gg})}{R} \frac{1}{T} + b'_{\text{gg}} \quad (5.15)$$

egyenletekből határozható meg. Az 5.14 reláció a 267.1 cm<sup>-1</sup>-nél, az 5.15 egyenlet pedig a 380 cm<sup>-1</sup>-nél megjelenő sávra vonatkozik. A két összefüggés különbségéből a következő összefüggés adódik a konstans tagokra:

$$b'_{\text{gg}} = b_{\text{gg}} + \ln \frac{I'(\text{gg})}{I(\text{gg})} \quad (5.16)$$

Ennek megfelelően az 5.13 egyenlet iteratíván oldható meg.  $b_{\text{gg}}$  kezdeti értékeként a

$$\ln \frac{I(\text{gg})}{I'(\text{tt})} = -\frac{\Delta H^m(\text{gg})}{R} \frac{1}{T} + b_{\text{gg}}^m \quad (5.17)$$

illesztésből származó konstans ( $b_{\text{gg}}^m = -0.69$ ) választottuk, míg a két átmenet intenzitásának arányát ( $I'(\text{gg})/I(\text{gg})$ ) a Gaussian programcsomag segítségével számított B3LYP/aug-cc-pVQZ Raman-aktivitások arányaként becsültük. A számított aktivitások az 5.13. táblázatban találhatók.

5.13. táblázat: A gg konformer B3LYP módszerrel meghatározott Raman aktivitásai.

	aug-cc-pVTZ		d-aug-cc-pVTZ		aug-cc-pVQZ	
frekvencia / cm <sup>-1</sup>	269	378	269	378	270	379
aktivitás / Å <sup>4</sup> amu <sup>-1</sup>	0.6723	0.1329	0.6741	0.1335	0.6767	0.1326

A gg konformerre kapott összefüggés megoldásával a  $b'_{\text{gg}}$  és a  $\Delta H^\circ(\text{gg})$  paraméterekhez jutunk, melyek ismeretében a tg konformer relatív entalpiája az 5.13 egyenlettel analóg összefüggés segítségével szintén korrigálható:

$$\ln \frac{I(\text{tg})}{I(\text{tt})} = -\frac{\Delta H^\circ(\text{tg})}{R} \frac{1}{T} + b_{\text{tg}} = \ln \frac{I(\text{tg})}{I'(\text{tt})} + \ln \left[ 1 + \exp \left( -\frac{\Delta H^\circ(\text{gg})}{RT} + b'_{\text{gg}} \right) \right] \quad (5.18)$$

Az 5.13 egyenlet megoldásaként a  $\Delta H^\circ(\text{gg}) = 961 \pm 15$  cal mol<sup>-1</sup> entalpiakülönbség áll elő, míg a  $b_{\text{gg}}$  paraméter értékére  $-0.66$  adódik. Ezeknek az értékeknek a segítségével a tg konformerre is végrehajtva a korrekciót a megfelelő entalpiakülönbség értéke  $\Delta H^\circ(\text{tg}) = 626 \pm 5$  cal mol<sup>-1</sup>. A kiértékelés pontos menetének ismerete nélkül teljes

bizonyossággal a kísérleti és elméleti értékek közötti eltérés oka nem állapítható meg, látható viszont, hogy a  $380 \text{ cm}^{-1}$ -es sávra történő korrekció a kísérleti és az általunk megadott elméleti eredmények közötti diszkrepanciát jelentősen csökkenti.

## 5.3. Az entalpiakülönbségek QSPR modellezése

### 5.3.1. Molekulaszimmetrián alapuló deskriptorok

Mivel az *n*-pentán konformereinek entalpiakülönbségei kvantumkémiailag a 4.2. fejezetben ismertetett módon csak jelentős számítási kapacitás igénybevételével határozhatók meg, ezért érdemes arra egy modellt felállítani, melynek segítségével ezek az értékek a szerkezet, esetleg a hőmérséklet függvényében kellő precizitással becsülhetők. Természetesen ilyen módon más termokémiai mennyiségek is leírhatók, azonban a QSPR-modellezés alapjául a tetszőleges hőmérsékleten rendelkezésre álló és kísérleti, valamint egyéb elméleti értékekkel összehasonlítható entalpiakülönbségeket választottuk. A hőmérsékletfüggés, valamint további vegyületekhez, illetve más rotációs izomerekhez (például nyeregpontokhoz) tartozó értékek modellezése ezeknek a mennyiségeknek a becsléséhez szükséges időt jelentős mértékben lerövidítené. A továbbiakban egy ilyen modell felállítását kíséreljük meg.

Az *n*-pentán konformereinek entalpiája az eddigi eredmények alapján azok szimmetriájával korrelálni látszik, a tt, tg, gg, gx sorrendben a szerkezetek pontcsoportja  $C_{2v}$ ,  $C_1$ ,  $C_2$  és  $C_1$ , míg a tt konformerre vonatkoztatott standard entalpiák 298 K hőmérsékleten 630, 940 és 2733 cal mol<sup>-1</sup> értéket vesznek fel. A szimmetria és az entalpiakülönbségek korrelációjának vizsgálatára általunk alkalmazott molekuláris deskriptorok<sup>156</sup> a tehetetlenségi alaktényező ( $S_I$ , *inertial shape factor*), a molekuláris excentricitás ( $\epsilon$ , *molecular eccentricity*), a molekuláris aszfericitás ( $\Omega_A$ , *molecular asphericity*), a molekuláris szfericitás ( $\Omega_S$ , *molecular sphericity*), a szimmetriatényező ( $\sigma_{\text{SYM}}$ , *symmetry factor*) és a molekuláris szimmetriára vonatkozó információs index ( $I_{\text{SYM}}$ , *information index on molecular symmetry*), melyek a következőképp definiálhatók:

$$\begin{aligned} \epsilon &= \frac{(I_C^2 - I_A^2)^{1/2}}{I_C} & S_I &= \frac{I_B}{I_A I_C} \\ \Omega_S &= 3 \frac{I_A}{I_A + I_B + I_C} & \Omega_A &= \frac{1}{2} \frac{(I_A - I_B)^2 + (I_A - I_C)^2 + (I_B - I_C)^2}{I_A^2 + I_B^2 + I_C^2} \\ \sigma_{\text{SYM}} &= \frac{\sigma_e \sigma_i}{I_R} & I_{\text{SYM}} &= N \log_2 N - \sum_{g=1}^G N_g \log_2 N_g \end{aligned} \quad (5.19)$$

ahol  $I_A \leq I_B \leq I_C$  a főtehetlenségi momentumok,  $N$  az atomszám,  $G$  a szimmetriaműveletek által egymásba transzformált atomok alkotta szimmetriaosztályok száma,  $N_g$  a  $g$ -ik szimmetriaosztály számossága,  $\sigma_e$  és  $\sigma_i$  a külső és a belső forgási szimmetriaszámok,  $I_R$  pedig az adott szerkezethez tartozó enantiomerek száma. A modellezés során felhasznált indexek értéke és a 298 K-re vonatkozó standard entalpiakülönbségek az 5.14. táblázatban láthatók. Az 5.19 egyenletben megadott deskriptorok vagy a molekula szimmetriáját meghatározó alakfüggő mennyiségekre (főtehetlenségi momentumok), vagy közvetlenül a szimmetriából származó tulajdonságokra (például szimmetriaszámok, szimmetriaosztályok rendje) épülnek.

5.14. táblázat: Az alkalmazott, molekulaszimmetrián alapuló deskriptorok és a modellezett entalpiakülönbségek értékei az egyes konformerekre.

Konformer	$S_I$	$\epsilon$	$\Omega_A$	$\Omega_S$	$\sigma_{SYM}$	$I_{SYM}$	$\Delta H_{298}^\circ$ <sup>a</sup>
tg	0.00602367	0.983410	0.325220	0.258645	4.5	69.4869	630
gg	0.00382238	0.933853	0.185197	0.474279	9.0	53.4869	940
gx	0.00319882	0.937217	0.189884	0.489251	4.5	69.4869	2733

<sup>a</sup> Entalpiakülönbség cal mol<sup>-1</sup> egységekben.

A modellépítést minden esetben az összes lehetséges regressziós egyenlet (BSS, *best subset selection*) módszerének alkalmazásával végeztük, mely során a modellezni kívánt mennyiséget a deskriptorok összes kombinációjával megillesztjük. A legjobb modell kiválasztása érdekében, ha másképp nincs megadva, a reziduális variancia ( $s_{r^2}$ ) és a PRESS (*predicted residual error sum of squares*) érték összegét tekintettük, így a legjobban illeszkedő görbe PRESS statisztika segítségével történő egyidejű belső validálására is lehetőség adódik.

5.15. táblázat: A szimmetriadeszkriptorok alkalmazásával kapott modellek jellemzői.

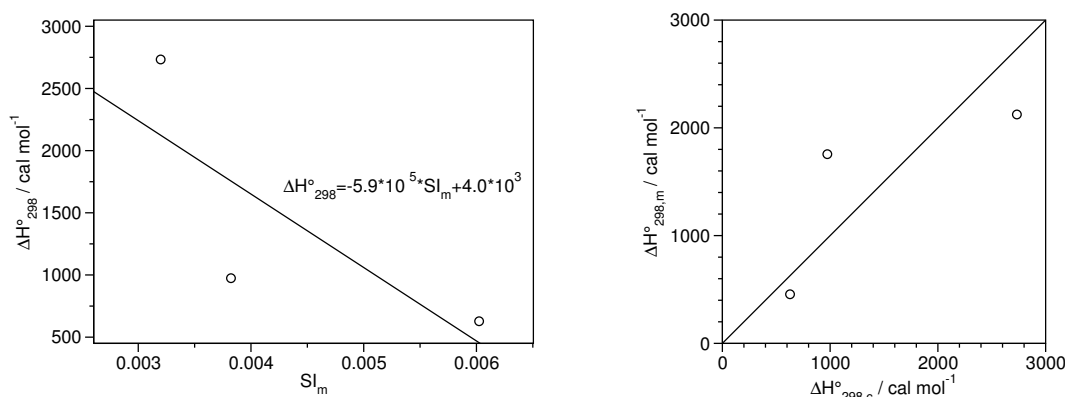
Alkalmazott deskriptor	$s_{r^2}$ <sup>a</sup> cal <sup>2</sup> mol <sup>-2</sup>	PRESS <sup>b</sup>	$s_r$ <sup>c</sup> cal mol <sup>-1</sup>
$\sigma_{SYM}$	$2.20 \cdot 10^6$	$8.90 \cdot 10^6$	1483
$I_{sym}$	$2.20 \cdot 10^6$	$9.10 \cdot 10^6$	1483
$SI$	$1.00 \cdot 10^6$	$3.90 \cdot 10^7$	1000
$\Omega_S$	$1.40 \cdot 10^6$	$6.30 \cdot 10^8$	1183
$\epsilon$	$1.70 \cdot 10^6$	$7.00 \cdot 10^8$	1304
$\Omega_A$	$1.60 \cdot 10^6$	$2.80 \cdot 10^9$	1265

<sup>a</sup> Reziduális variancia.

<sup>b</sup> PRESS (*predicted residual error sum of squares*) értékek.

<sup>c</sup> Reziduális szórás.

Az 5.14. táblázatban feltüntetett három pontra történő illesztés eredménye az 5.15. táblázatban látható, míg a legkisebb maradékszórású modell, valamint az e modell által jóslott entalpiakülönbségek a modellezett értékek függvényében az 5.6. ábrán tekinthetők meg. Az illesztés eredménye az  $S_I$  indexszel bizonyult a legjobbnak, ez ugyanis az entalpiakülönbségekkel monoton változik, valamint egyéb deskriptorokkal ellentétben, nem degenerált. Már ezekből az eredményekből is jól látható, hogy a konformerek entalpiakülönbségeivel egyetlen szimmetrián alapuló deskriptor csak csekély mértékű korrelációt mutat, a modellek  $1000 - 1500 \text{ cal mol}^{-1}$  maradékszórással jellemezhetők, mely az utolsó pont kivételével még a függő változó értékét is meghaladja.



(a) Az entalpiakülönbség mint az  $S_I$  deskriptor függvénye. (b) A számított értékek a modellezett adatok függvényében.

5.6. ábra: A legkisebb maradékszórású,  $S_I$  szimmetria deskriptorral történő illesztés eredménye.

Az 5.6. (a) ábrán jól látszik, hogy a modelfüggvény rosszul illeszkedik az entalpiakülönbségekre, valamint az 5.6. (b) ábráról is nyilvánvaló, hogy a modell által jóslott  $\Delta H_{298,m}^{\circ}$  értékek jelentősen különböznek a modellezett  $\Delta H_{298,c}^{\circ}$  értékektől, melyet az ábrán feltüntetett, ideális esetet reprezentáló zérus tengelymetszetű, egységnyi meredekségű egyenestől való eltérés is mutat. A főtehetetlenségi momentumokon, a szimmetriaszámokon és a szimmetriaosztályok rendjén alapuló deskriptorok tehát egyváltozós esetben az  $n$ -pentánkonformerek entalpiakülönbségeinek leírására nem alkalmasak.

### 5.3.2. Topografikus indexek

Korábbi tanulmányok alapján kijelenthető, hogy a módosított topológiai indexek (például elektrotológiai vagy topografikus indexek) különféle vegyületek termokémiai tulajdonságainak modellezése során sikerrel alkalmazhatók.<sup>117,185</sup> A továbbiakban ezért ilyen, a pentánkonformerek entalpiakülönbségeinek leírására alkalmas módosított topológiai indexeket igyekszünk definiálni.

A legegyszerűbb és a legkönnyebben kezelhető eljárást a konformerekben megfigyelhető C–C–C–C torziós szögek használata nyújtja, mellyel a szükséges periodikus, folytonos függvények könnyen előállíthatók. Módosított topológiai indexek a molekulagráf súly- és távolságmátrixának paraméterezésével hozhatók létre. Az új indexek definiálásakor a molekulagráfot egy  $N_c$  szénatomos  $n$ -alkán esetén ezért a két láncvégi szénatom (a molekulagráf levelei) elhagyásával egy  $N_c - 2$  csúcsú gráffal helyettesítettük, a gráf éleit pedig az adott élnek megfelelő kötés két pillératomja és az azokkal szomszédos egy-egy szénatom által meghatározott torziós szög abszolút értékével súlyoztuk, így a gráf  $(N_c - 2) \times (N_c - 2)$ -es  $\mathbf{W}$  súlymátrixa a következőképp írható fel:

$$W_{ij} = \begin{cases} |\psi_{k,i,j,l}|, & \text{ha } (k,i), (i,j), (j,l) \in E \\ 0, & \text{különben,} \end{cases} \quad (5.20)$$

ahol  $\psi_{k,i,j,l}$  a  $C_k$ – $C_i$ – $C_j$ – $C_l$  torziós szöget jelöli. Természetesen az így definiált súlymátrix pontosan akkor egyértelmű, ha minden  $\psi_{k,i,j,l}$  ( $k$  és  $l$  atomok) egyértelmű. Amennyiben  $k$  és  $l$  többféleképp is megadható, tehát a molekula elágazásokat tartalmaz, egyértelműen meg kell határozni, melyik torziót, esetleg milyen egyéb mennyiséget veszünk figyelembe a súlymátrix felírásakor. Mivel a jelen kutatás tárgyát kizárólag egyenes láncú alkánok képezik, e kérdések megválaszolásához további vizsgálatok szükségesek. A torziós szög abszolút értékének alkalmazásával, azokat a  $(-180^\circ, 180^\circ]$  tartományon értelmezve, az index  $\pm 180^\circ$ -ban való folytonosságára vonatkozó korábbi feltétel is teljesül.

A távolságmátrix elemeit a csúcsok közötti legrövidebb útban szereplő élekhez rendelt torziós szögek összegének abszolút értékeként definiáltuk, ahol a torziós szögek előjelét is figyelembe vettük. Mivel egyes, a gráfcúcsok távolságának meghatározására szolgáló algoritmusok (például a Bellman–Ford-algoritmus) nem alkalmazhatók abban az esetben, ha negatív összegű körök találhatók a gráfban, negatív torziós szögeket tartalmazó szerkezetek esetén a távolságmátrix meghatározására is gondot kell fordítani. Az abszolút érték alkalmazása ebben az esetben is biztosítja a deszkriptor folytonosságát.

A modellezés során alkalmazott topológiai indexek a Wiener-index ( $W$ ), a Balaban-index ( $J$ ), a Xu-index ( $Xu$ ), a geometriai-aritmetikai index ( $GA1$ ), a Fi-index ( $Fi$ ), a Randić-index ( ${}^1\chi$ ), az alkánok forráspontjának modellezése során sikeresen alkalmazott Fi-index és Randić-index összege ( $Fix$ ), a harmonikus index ( $HI$ ), a Schultz-index ( $MTI$ ), a módosított Schultz-index ( $\overline{MTI}$ ), az ISI-index ( $ISI$ ), az első ( $M_1$ ) és második ( $M_2$ ) Zágráb-indexek, a Narumi–Katayama-index ( $NK$ ), az atom-kötés konnektivitási index ( $ABC$ ) és az SCI-index ( $SCI$ ). A felsorolt deszkriptorok az 5.21 egyenlettel definiálhatók, ahol  $n_v$  a csúcsok,  $n_e$  az élek száma,  $w_{ij}$ ,  $v_i$ ,  $d_{ij}$  és  $s_i$  pedig a súlymátrix, a súlymátrix sorainak összegeként alkotott foksávvektor, a távolságmátrix és



a távolságösszeg-vektor elemeit jelentik. A továbbiakban az indexek  $m$  alsó indexe az előbb ismertetett módon felírt súly- és távolságmátrixokból előállított, módosított indexeket jelöli. Mivel ezek az indexek gráfinvariánsok (izomorf gráfokhoz azonos értékeket rendelnek), értékük független a gráf csúcsainak számozásától.

$$\begin{aligned}
W &= \frac{1}{2} \sum_{i=1}^{n_v} \sum_{j=1}^{n_v} d_{ij} & J &= \frac{n_e}{n_e - n_v + 1} \sum_{i=1}^{n_v} (s_i s_j)^{1/2} \\
Xu &= n_v^{1/2} \log \frac{\sum_{i=1}^{n_v} v_i s_i^2}{\sum_{i=1}^{n_v} v_i s_i} & GA1 &= \sum_{(i,j) \in E} \frac{(v_i v_j)^{1/2}}{(v_i + v_j)/2} \\
{}^1\chi &= \sum_{(i,j) \in E} (v_i v_j)^{-1/2} & HI &= 2 \sum_{(i,j) \in E} (v_i + v_j)^{-1} \\
MTI &= \sum_{i=1}^{n_v} \sum_{j=1}^{n_v} (w_{ji} + d_{ij}) v_j & \overline{MTI} &= \sum_{i=1}^{n_v} \sum_{j=1}^{i-1} d_{ij} v_i v_j \\
ISI &= \sum_{(i,j) \in E} \frac{v_i v_j}{v_i + v_j} & M_1 &= \sum_{i=1}^{n_v} v_i^2 \\
M_2 &= \sum_{(i,j) \in E} v_i v_j & NK &= \prod_{i=1}^{n_v} v_i \\
ABC &= \sum_{(i,j) \in E} \left( \frac{v_i + v_j - 2}{v_i v_j} \right)^{1/2} & SCI &= \sum_{(i,j) \in E} \frac{1}{(v_i + v_j)^{1/2}} \\
Fi &= \sum_{i=1}^{n_v} (4 - v_i) + \sum_{(i,j) \in E} [(v_i - 1)(v_j - 1)]^{1/2}
\end{aligned} \tag{5.21}$$

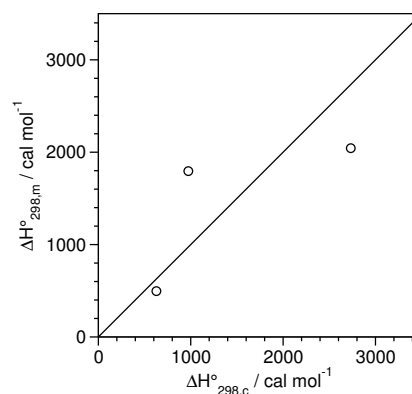
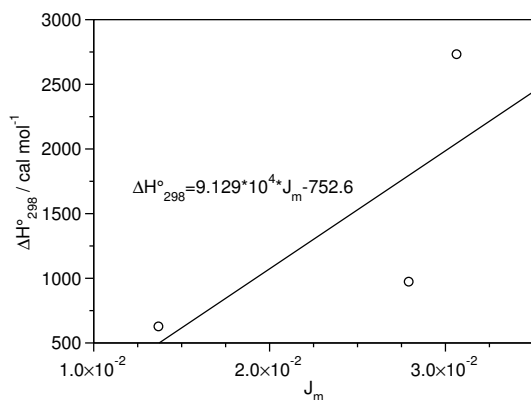
A módosított indexek értéke az  $n$ -pentán konformereire az 5.16. táblázatban látható. A súlyozáshoz az F2. táblázatban található, CCSD(T)/cc-pVTZ szinten optimalizált geometriákból származó torziós szöveget használtuk fel, melyek értéke a  $tt$  konformer esetén  $t = 180^\circ$ , a  $tg$  konformernél  $t = 176.452^\circ$  és  $g = 64.545^\circ$ , a  $gg$  konformernél  $g = 58.505^\circ$ , valamint a  $gx$  konformerre  $g = 59.993^\circ$  és  $x = 94.814^\circ$ .

A módosított indexekkel történő illesztések eredménye az 5.17. táblázatban tekinthető meg, melyek a szimmetrián alapuló deskriptorokkal elértékhez hasonlóak. A reziduális szórások ebben az esetben is nagyok ( $1100 - 1600 \text{ cal mol}^{-1}$ ). A legkisebb maradékszórású, Balaban-index segítségével felállított modell, valamint a modelfüggvénnyel számított értékek a modellezett entalpiakülönbségek függvényében az 5.7. ábrán láthatók. Megfigyelhető, hogy az illesztett függvény ebben az esetben sem illeszkedik jól a

modellezett adatokra.

5.16. táblázat: A módosított topológiai indexek értékei az egyes konformerekre.

Index	tg	gg	gx
$W_m$	482.0	234.0	189.6
$J_m$	0.01368	0.02791	0.03063
$Xu_m$	4.37399	3.77962	3.71341
$GA1_m$	1.804	1.886	1.868
$Fi_m$	18.51	13.32	15.15
${}^1\chi_m$	0.01287	0.02417	0.01863
$Fix_m$	18.52	13.34	15.16
$HI_m$	0.01134	0.02279	0.01732
$MTI_m$	1928.0	936.1	998.5
$\overline{MTI}_m$	1004007.2	400505.9	1391667.7
$ISI_m$	152.8	78.0	102.0
$M_{1,m}$	93380.9	20537.0	36554.1
$M_{2,m}$	330.9	165.5	217.5
$NK_m$	2744737.6	400505.9	880569.5
$ABC_m$	0.2385	0.3184	0.2813
$SCI_m$	0.00567	0.01140	0.00866



(a) Az entalpiakülönbségek a módosított Balaban-index függvényében.

(b) A számított értékek a modellezett mennyiségek függvényében.

5.7. ábra: A legkisebb maradékszórású, módosított topológiai indexszel történő illesztés eredménye.

5.17. táblázat: A módosított topológiai indexek alkalmazásával kapott modellek jellemzői.

Alkalmazott deszkriptor	$s_{r^2}$ <sup>a</sup>	PRESS <sup>b</sup>	$s_r$ <sup>c</sup>
	cal <sup>2</sup> mol <sup>-2</sup>		cal mol <sup>-1</sup>
$\overline{MTI}_m$	$1.30 \cdot 10^6$	$2.10 \cdot 10^7$	1140
$SCI_m$	$2.50 \cdot 10^6$	$3.40 \cdot 10^7$	1581
$HI_m$	$2.50 \cdot 10^6$	$3.40 \cdot 10^7$	1581
${}^1\chi_m$	$2.50 \cdot 10^6$	$3.40 \cdot 10^7$	1581
$ABC_m$	$2.50 \cdot 10^6$	$3.40 \cdot 10^7$	1581
$Fi_m$	$2.30 \cdot 10^6$	$4.10 \cdot 10^7$	1517
$Fix_m$	$2.30 \cdot 10^6$	$4.10 \cdot 10^7$	1517
$ISI_m$	$2.20 \cdot 10^6$	$4.50 \cdot 10^7$	1483
$M_{2,m}$	$2.20 \cdot 10^6$	$4.60 \cdot 10^7$	1483
$M_{1,m}$	$2.00 \cdot 10^6$	$7.90 \cdot 10^7$	1414
$GA1_m$	$2.00 \cdot 10^6$	$8.00 \cdot 10^7$	1414
$J_m$	$1.20 \cdot 10^6$	$8.30 \cdot 10^7$	1095
$NK_m$	$2.00 \cdot 10^6$	$8.80 \cdot 10^7$	1414
$W_m$	$1.20 \cdot 10^6$	$9.50 \cdot 10^7$	1095
$Xu_m$	$1.30 \cdot 10^6$	$2.40 \cdot 10^8$	1140
$MTI_m$	$1.70 \cdot 10^6$	$8.10 \cdot 10^8$	1304

<sup>a</sup> Reziduális variancia.

<sup>b</sup> PRESS (*predicted residual error sum of squares*) érték.

<sup>c</sup> Reziduális szórás.

Az összes illesztés, a szimmetria deszkritorokhoz hasonlóan, három pontra történt, mely a *leave-one-out* típusú PRESS statisztikára jelentős hatást gyakorol, egy nagyobb tanítóhalmazzal végzett modellezés esetén jobb értékekre lehet számítani. Mivel a konformerek egy hőmérsékleten számított három entalpiakülönbségére történő illesztésekből egyértelműen nem dönthető el, melyik deszkriptorok alkalmasak inkább ezen mennyiségek leírására, érdemes a modellezést egy nagyobb adathalmazon is végrehajtani.

### 5.3.3. Hőmérsékletfüggés modellezése

A modellezéshez nagyobb tanítóhalmaz az entalpiakülönbségek hőmérsékletfüggésének figyelembevételével állítható elő, melyhez mindössze a 4.18 egyenlet alapján számított hőmérsékleti korrekció értékének különböző hőmérsékleteken történő meghatározása szükséges. A 25–412 K tartományban előállított entalpiakülönbségek az 5.18. táblázatban láthatók. E hőmérséklettartományon az entalpiakülönbségek becsült hibája

a szerkezettől függően 45 – 140 cal mol<sup>-1</sup>.

A szerkezet hatásának figyelembevételére a korábban is alkalmazott szimmetrián alapuló és a módosított topológiai indexek szolgálnak, melyek a másik leíróként szereplő hőmérséklettel nem korrelálnak, így a multikollinearitás problémája nem lép fel. Érdeemes megemlíteni továbbá, hogy az 5.18. táblázat adatainak illesztése során kapott legjobb modellekben nem szerepel a hőmérséklet mint változó, a legkisebb reziduális szórású modellek két szimmetria deszkriptort vagy módosított topológiai indexet tartalmaznak. A kétváltozós, tehát háromparaméteres felületek a hőmérsékletfüggés hiánya miatt csak a szerkezet entalpiakülönbségekre gyakorolt hatását képesek leírni. Mivel minden hőmérsékleten három pont áll rendelkezésre (tg, gg és gx konformerek relatív entalpiái) a hőmérsékletet nem tartalmazó modellekre kapott egyenletrendszerek meghatározottak, ezért ugyanazt az eredményt (maradékokat) szolgáltatják. A hőmérsékletfüggés leírásához viszont az egyik modellváltozónak mindenképp a hőmérsékletet szükséges választanunk, az 5.19. táblázatban ezért csak a leíróként a hőmérsékletet tartalmazó legjobb modellek kerültek feltüntetésre.

5.18. táblázat: Az *n*-pentánkonformerek relatív entalpiáinak hőmérsékletfüggése.

T / K	$\Delta H_T^\circ(\text{tg})$	$\Delta H_T^\circ(\text{gg})$	$\Delta H_T^\circ(\text{gx})$
	cal mol <sup>-1</sup>		
25	645	1100	2976
50	649	1097	2978
75	650	1086	2972
100	649	1073	2958
125	647	1058	2937
150	644	1044	2914
175	641	1030	2889
196.5	639	1019	2866
225	636	1005	2833
250	634	994	2801
275	631	984	2767
298.15	629	976	2733
325	626	968	2692
350	623	962	2651
385	619	956	2593
412	617	954	2548

Megfigyelhető, hogy az 5.19. táblázatban szereplő modellek maradékszórása az entalpiakülönbségek hibájához képest még mindig magas (a 655 cal mol<sup>-1</sup>-et is elérheti), azonban a hőmérséklet változóként történő szerepeltetése a modellezett adatok leírá-

sát jelentős mértékben javítja. A legjobb modell a Balaban-index és a hőmérséklet függvényeként állítható fel, az ezzel a modellel számított értékek a modellezett entalpiakülönbségek függvényében az 5.8. ábrán láthatók. Az 5.7. (b) ábrán feltüntetett modellhez képest jelentős javulás nem figyelhető meg, a szórás csökkenése valószínűleg a modellparaméterek tanítóhalmaz méretéhez viszonyított kisebb számának tulajdonítható.

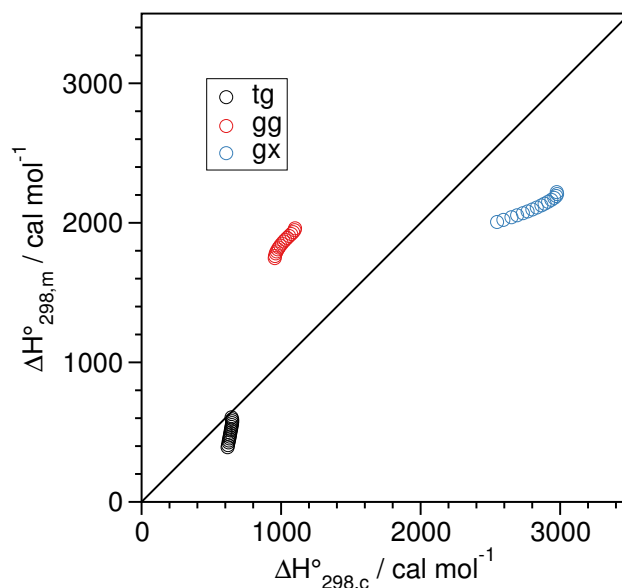
5.19. táblázat: A módosított topológiai indexek és a hőmérséklet alkalmazásával kapott modellek jellemzői.

Alkalmazott deszkriptorok	$s_r^2$ <sup>a</sup> cal <sup>2</sup> mol <sup>-2</sup>	PRESS <sup>b</sup>	$s_r$ <sup>c</sup> cal mol <sup>-1</sup>
$J_m, T$	$4.30 \cdot 10^5$	$4.80 \cdot 10^5$	655.7
$W_m, T$	$4.40 \cdot 10^5$	$4.90 \cdot 10^5$	663.3
$Xu_m, T$	$4.90 \cdot 10^5$	$5.50 \cdot 10^5$	700.0
$\overline{MTI}_m, T$	$5.10 \cdot 10^5$	$5.70 \cdot 10^5$	714.1
$\Omega_S, T$	$5.70 \cdot 10^5$	$6.60 \cdot 10^5$	755.0

<sup>a</sup> Reziduális variancia.

<sup>b</sup> PRESS (*predicted residual error sum of squares*) érték.

<sup>c</sup> Reziduális szórás.



5.8. ábra: A módosított Balaban-index és a hőmérséklet függvényében történő illesztés által jósolt entalpiakülönbségek a modellezett értékek függvényében.

### 5.3.4. Nemlineáris modellfüggvény

Bár QSPR modellezés során néhány esetben<sup>26,116</sup> nemlineáris modellfüggvényt is alkalmaztak, ezek a modellek mindig egyetlen változóra szorítkoztak. Amennyiben a modellezett tulajdonságot nem lehet egy változó segítségével leírni, a korábbi tanulmányok mintájára többváltozós nemlineáris modellt is választhatunk. Ez a megközelítés jelentősen javíthatja a leírást, míg nem növeli a szükséges változók, csupán a paraméterek számát.

A hőmérsékletfüggő entalpiakülönbségek esetén is az előbb említett probléma áll fenn, a modellépítés csak több egymástól független, a szerkezettől és a hőmérséklettől függő változó használatával lehetséges. A 4.19 egyenlettel megadott entalpiák, valamint ezek különbsége a hőmérséklettel alapvetően nemlineáris függést mutat, így egy lineáris modellfüggvénytől nem várható el, hogy ezeket a mennyiségeket tetszőleges pontossággal megadja. Bár az 5.3.3. fejezetben említett okból a szerkezet leírása két-tónél több változó alkalmazásával nem lesz pontosabb, a hőmérsékletfüggés jellemzése javul.

5.20. táblázat: A többváltozós nemlineáris illesztéssel kapott modellek jellemzői.

Deszkriptorok	$s_{r,2}$ <sup>a</sup>	PRESS <sup>b</sup>	$s_{r,2}$ +PRESS	$s_r$ <sup>c</sup>
		cal <sup>2</sup> mol <sup>-2</sup>		cal mol <sup>-1</sup>
$J_m, J_m^2, T, J_m \cdot T$	$1.20 \cdot 10^3$	$1.60 \cdot 10^3$	$2.80 \cdot 10^3$	34.64
$W_m, W_m^2, T, W_m \cdot T$	$1.20 \cdot 10^3$	$1.60 \cdot 10^3$	$2.80 \cdot 10^3$	34.64
$Xu_m, Xu_m^2, T, Xu_m \cdot T$	$1.40 \cdot 10^3$	$1.80 \cdot 10^3$	$3.20 \cdot 10^3$	37.42
$J_m, J_m^2, T^2, J_m \cdot T$	$1.60 \cdot 10^3$	$2.10 \cdot 10^3$	$3.70 \cdot 10^3$	40.00
$J_m, J_m^2, J_m \cdot T$	$1.70 \cdot 10^3$	$2.20 \cdot 10^3$	$3.90 \cdot 10^3$	41.23
$MTI_m, MTI_m^2, T, MTI_m \cdot T$	$1.80 \cdot 10^3$	$2.40 \cdot 10^3$	$4.20 \cdot 10^3$	42.43
$W_m, W_m^2, T^2, W_m \cdot T$	$2.10 \cdot 10^3$	$2.80 \cdot 10^3$	$4.90 \cdot 10^3$	45.83
$\overline{MTI}, \overline{MTI}^2, \overline{MTI} \cdot T$	$2.30 \cdot 10^3$	$2.90 \cdot 10^3$	$5.10 \cdot 10^3$	47.96
$NK_m, NK_m^2, T, NK_m \cdot T$	$2.20 \cdot 10^3$	$2.90 \cdot 10^3$	$5.10 \cdot 10^3$	46.90
$GAI_m, GAI_m^2, T, GAI_m \cdot T$	$2.30 \cdot 10^3$	$2.90 \cdot 10^3$	$5.20 \cdot 10^3$	47.96

<sup>a</sup> Reziduális variancia.

<sup>b</sup> PRESS (*predicted residual error sum of squares*) értékek.

<sup>c</sup> Reziduális szórás.

A 4.38 egyenlettel megadott hatványfüggvény alkalmazásával végzett modellépítés során kapott legjobb eredmények néhány statisztikai mutatója az 5.20. táblázatban látható. Az alkalmazott modell minden esetben legfeljebb kétváltozós és kvadratikus volt. Megfigyelhető, hogy ezekben a modellekben egy módosított topológiai index mellett mindig megjelenik annak négyzete is, ami az entalpiakülönbségek szerkezetfüggő

részének lehető legjobb leírását adja, viszont a hőmérséklet vagy annak négyzete mellett a két változó szorzata is előfordul, mely a hőmérsékletfüggés pontosabb jellemzéséhez járul hozzá.

Látható, hogy a modellek maradékszórása nagymértékben csökkent a korábbiakhoz képest, a legkisebb reziduális szórás  $34.64 \text{ cal mol}^{-1}$ -nek adódik, mely a számított entalpiakülönbségek becsült hibájánál alacsonyabb. Továbbá a Student-féle t-próba végrehajtásával 95%-os szignifikancia szint mellett az összes felsorolt modell minden paramétere szignifikánsnak bizonyult. A várakozásoknak megfelelően a PRESS statisztika értékei is jelentősen csökkentek, azok nagyságrendileg a maradékszórásokkal egyeznek meg. Különösen a Balaban- és a Wiener-indexek teljesítettek jól a vizsgálat során, melyek több modellben is szerepelnek.

A legjobb modell szintén a Balaban-index módosított változatával és a hőmérséklettel állítható fel, és a következő egyenlettel jellemezhető:

$$\begin{aligned} \Delta H_T^\circ = & -(152 \pm 1) \cdot 10^4 \cdot J_m + (375 \pm 3) \cdot 10^5 \cdot J_m^2 + (0.643 \pm 0.145) \cdot T \\ & -(49.7 \pm 5.8) \cdot J_m \cdot T + (1440 \pm 126) \end{aligned} \quad (5.22)$$

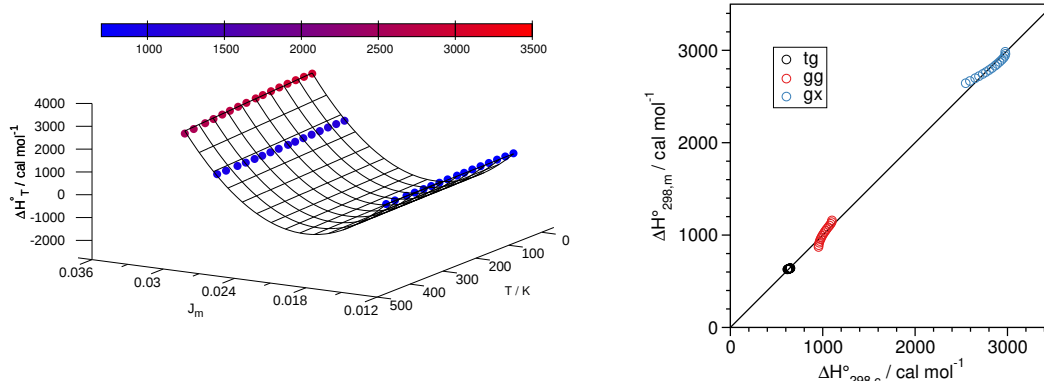
$$R = 0.9994 \quad s_r = 34.64 \text{ cal mol}^{-1} \quad F = 9039 \quad m = 48$$

Mint az az  $m = 48$  pontra végzett illesztés  $R$  regressziós együtthatójából látszik, az illeszkedés megfelelő, továbbá a Fischer-féle  $F$  érték és a t-statisztika alapján is elmondható, hogy a modell és annak minden paramétere szignifikáns. A reziduális szórás is igen alacsony, mindössze  $34.64 \text{ cal mol}^{-1}$ . A belső validálásra vonatkozó „leave-one-out” szórás értéke  $39.74 \text{ cal mol}^{-1}$ , ami azt mutatja, hogy a modell interpolációra való alkalmazhatósága is kielégítő. A 48 adatból a studentizált reziduumokon végrehajtott kétoldali t-próba alapján 4 bizonyult kiugró pontnak.

Maga a modell az illesztéshez használt pontokkal az 5.9. (a) ábrán tekinthető meg. Ezen az ábrán is megfigyelhető a jó illeszkedés a modellezett adatokra. Látható továbbá az is, hogy egy bizonyos tartományon a modellfüggvény negatív értékeket vesz fel. Ez a valóságban nem teljesülhet, egy tetszőleges rotációs izomer entalpiája nem lehet alacsonyabb, mint a legstabilabb tt konformeré, melyből arra következtethetünk, hogy a modell a konformereket, esetleg az azokhoz hasonló szerkezeteket ugyan megfelelően leírja, a teljes CPES jellemzésére azonban nem alkalmas.

Az 5.9. (b) ábrán a legjobb modellel számított entalpiakülönbségek láthatók a kvantumkémiaileg becsült értékek függvényében. Látható, hogy a pontok jól illeszkednek az ideális viselkedést jelentő zérus tengelymetszetű, egységnyi meredekségű egyenesre, ami a modellezés sikerének további bizonyítéka.

Az 5.10. ábrán az illesztés reziduuma láthatók. Megfigyelhető, hogy a maradékok a hőmérséklet függvényében szabályosan változnak, ami a számított értékeket terhelő szisztematikus hiba jelenlétére utal. Ez az eltérés lehet modellhiba vagy a kvantum-

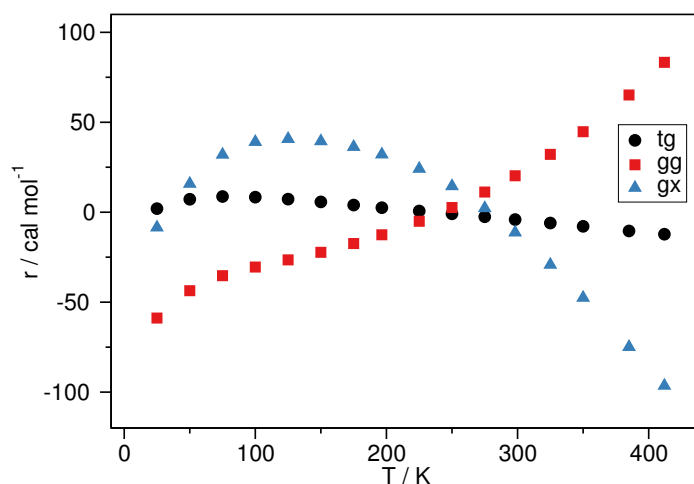


(a) Az entalpiakülönbségek a módosított Balaban-index és a hőmérséklet függvényében.

(b) A modell által jósolt entalpiakülönbségek a modellezett értékek függvényében.

5.9. ábra: A legkisebb maradékszórású, módosított topológiai indexszel és a hőmérséklettel történő nemlineáris modellfüggvény illesztésének eredménye.

kémiai számításokat terhelő szisztematikus hiba eredménye. Utóbbi kizárása kísérleti adatokra történő illesztéssel lenne megvalósítható, mely azonban azok csekély száma és inkonzisztenciája miatt – hasonló pontossággal legalábbis – jelenleg nem lehetséges.



5.10. ábra: A módosított Schultz-index és a hőmérséklet függvényében történő illesztés reziduumaival.

### 5.3.5. Az elsőfajú nyeregponatok modellezése

A modell interpolációra való alkalmazhatósága megadható az  $n$ -pentán konformációs potenciális energia felületén megjelenő nyeregponatok vizsgálatával. Az ehhez szükséges adatokat egy további fókuszpont analízis eredményeiből<sup>205</sup> számítottuk, mely az



$n$ -bután és az  $n$ -pentán konformációs entalpiakülönbségeinek számítása során alkalmazott eljárásokhoz hasonlít, bár a referencia geometria előállításában ebben az esetben egy sűrűségfunkcionál módszerrel (DSD-PBEP86-D2/cc-pVTZ) a befagyasztott mag közelítés alkalmazásával történt. A módszer az egyes hozzájárulások számításához kisebb bázist alkalmaz és a kis hozzájárulásokat nem veszi figyelembe, viszont az FC közelítés korrekciójára egy MP2 módszerrel becsült tagot is tartalmaz. Az említett tanulmány és annak eredményei nem képezik a jelen dolgozat részét, így azok részletes bemutatásától eltekintünk, azokat csak a modellezés során felhasznált entalpiakülönbségek előállításához használjuk fel.

Az  $n$ -pentán konformereinek izomerizációjához tartozó átmeneti állapotokra számított módosított topológiai indexek és a nyeregpontok tt konformerhez viszonyított relatív entalpiái az 5.22. és az F10. táblázatokban láthatók. A számított entalpiakülönbségek 4.2. fejezetben leírttal azonos módon becsült hibái konformertől függően 95 – 205 cal mol<sup>-1</sup> közé esnek. Az indexek előállításához felhasznált torziós szögek értékei az 5.21. táblázatban tekinthető meg.

5.21. táblázat: Az átmeneti állapotokban megjelenő torziós szögek értékei.

szerkezet	$\theta_1$	$\theta_2$
$\gamma^\pm\gamma^\mp$	$\pm 76.086^\circ$	$\mp 76.086^\circ$
$t\tau$	$\pm 180.000^\circ$	$0.000^\circ$
$x^\pm\tau$	$\pm 81.959^\circ$	$\mp 5.159^\circ$
$t\chi^\pm$	$\mp 178.972^\circ$	$\pm 119.792^\circ$
$g^\pm\chi^\pm$	$\pm 68.691^\circ$	$\pm 117.860^\circ$
$g^\pm\chi^\mp$	$\pm 66.951^\circ$	$\mp 122.512^\circ$

Az 5.22. táblázat adatait az 5.22 egyenletbe behelyettesítve a kapott entalpiakülönbségek általában negatív értékeket vesznek fel, kivéve a  $\gamma^\pm\gamma^\mp$  és az  $x^\pm\tau$  szerkezeteket ( $\gamma^\pm \approx \pm 75^\circ$ ,  $\tau \approx 0^\circ$ ), melyek rendre 9500 és 14500 cal mol<sup>-1</sup> körül helyezkednek el. Az 5.22 egyenlettel megadott modell tehát a konformációs izomerizáció átmeneti állapotainak leírására egyértelműen nem megfelelő.

Érdeemes ezután megvizsgálni, hogy a definiált deskriptorokkal lehetséges-e egy, a konformációs potenciális energia felület elsőfajú nyeregpontjainak jellemzésére is alkalmas modell felállítása. A legjobb modellek az F11. táblázatban tekinthetők meg. A modellezés során különösen a  $GA1$  és a  $Fi$  indexek teljesítenek jól, továbbá az is megfigyelhető, hogy a modellek a nemlineáris tagokkal együtt 6 – 7 leíró tartalmazzanak, ezek közül azonban legfeljebb kettő hőmérsékletfüggő. Ennek eredményeképp a hőmérsékletfüggés leírása a szerkezeténél kevésbé pontos.

5.22. táblázat: A módosított topológiai indexek értékei a konformációs interkonverzió átmeneti állapotaira.

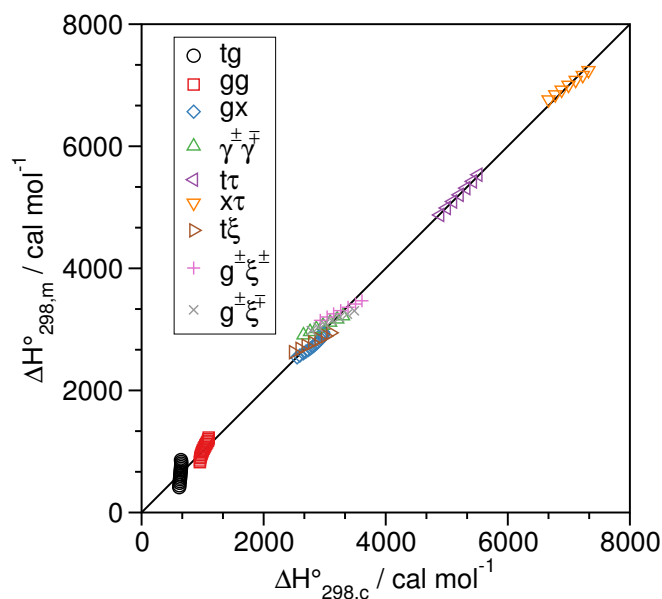
Index	$\gamma^\pm\gamma^\pm$	$t\tau$	$x^\pm\tau$	$t\xi^\pm$	$g^\pm\xi^\pm$	$g^\pm\xi^\mp$
$W_m$	152.2	360.0	163.9	357.9	373.1	245.0
$J_m$	0.03717	0.01897	0.04068	0.01615	0.01756	0.02402
$Xu_m$	3.643	4.291	3.669	4.198	4.148	3.895
$GA1_m$	1.886	1.000	1.459	1.872	1.861	1.855
$Fi_m$	15.07	14.37	10.79	20.79	16.53	16.64
$HI_m$	0.01752	0.01667	0.03350	0.00896	0.01441	0.01421
$MTI_m$	913.0	1440.0	676.3	1910.9	1492.4	1247.9
$\overline{MTI}_m$	880935.8	0.0	585195.8	4287300.2	2591376.1	2843686.2
$ISI_m$	101.4	90.0	47.1	197.4	122.4	123.9
$M_{1,m}$	34734.5	64800.0	14333.4	135641.0	53410.7	55387.9
$M_{2,m}$	215.2	180.0	105.7	420.4	261.5	265.0
$NK_m$	880935.8	0.0	36835.8	6405325.0	1510302.2	1554032.5
$SCI_m$	0.008760	0.008330	0.016750	0.004480	0.007200	0.007110
$SI$	0.003057	0.005302	0.003161	0.007737	0.005867	0.003960
$\epsilon$	0.9370	0.9811	0.9301	0.9902	0.9804	0.9574
$\Omega_A$	0.1903	0.3133	0.1767	0.3657	0.3125	0.2336
$\Omega_S$	0.4937	0.2814	0.5070	0.1972	0.2761	0.4033
$\sigma_{SYM}$	9.0	9.0	4.5	4.5	4.5	4.5
$I_{SYM}$	55.49	59.49	69.49	69.49	69.49	69.49

A legjobbnak tekintett modell és annak néhány fontosabb statisztikai jellemzője a következő egyenlettel fejezhető ki:

$$\begin{aligned}
\Delta H_T^o = & -(182 \pm 2) \cdot 10^3 \cdot GA1_m + (199 \pm 4) \cdot 10^2 \cdot GA1_m^2 - (108 \pm 1) \cdot 10^2 \cdot Fi_m \\
& - (139 \pm 2) \cdot Fi_m^2 - (3.50 \pm 0.79) \cdot T + (8450 \pm 80) \cdot GA1_m \cdot Fi_m \\
& + (1.31 \pm 0.44) \cdot GA1_m \cdot T + (231 \pm 2) \cdot 10^3 \quad (5.23) \\
R = & 0.9969 \quad s_r = 104.9 \text{ cal mol}^{-1} \quad F = 3821 \quad m = 90
\end{aligned}$$

A regressziós együttható értékéből látható, hogy az illeszkedés ebben az esetben is jónak mondható. Az  $F$  érték és a t-próba alapján a modell és a modellparaméterek szignifikánsak, a 90 adatpont közül mindössze 4 kiugró pont. A modell maradékszórása  $104.9 \text{ cal mol}^{-1}$ , mely ebben az esetben is alacsonyabb, mint a modellezett adatok becsült hibája, míg a belső validálásra vonatkozó PRESS érték gyöke („leave-one-out” szórás)  $113.4 \text{ cal mol}^{-1}$ , mely alapján a modell a bemenő adatok hibáján belül interpolációra is alkalmas.

Az 5.23 egyenlet által jósol entalpiakülönbségek a tanítóhalmaz adatainak függvényeként az 5.11. ábrán láthatók. Megfigyelhető, hogy a pontok továbbra is az egységnyi meredekségű egyenes körül helyezkednek el, azonban a hőmérsékletfüggés leírása inkább a nagyobb entalpiakülönbséggel jellemezhető átmeneti állapotok esetén korrekt. A kon-



5.11. ábra: Az átmeneti állapotokra kapott legjobb modell által jószolt entalpiakülönbségek a modellezett értékek függvényében.

formerekre vonatkozó maradékszórás ennek ellenére is  $92.97 \text{ cal mol}^{-1}$ , ami nem haladja meg azok becsült hibáját ( $45 - 140 \text{ cal mol}^{-1}$ ), míg ugyanez az érték az átmeneti állapotoknál  $105.63 \text{ cal mol}^{-1}$ -nek adódik, mely, figyelembe véve azok  $95 - 205 \text{ cal mol}^{-1}$  nagyságú bizonytalanságát, a konformerekre vonatkozó értéknél is jobbnak nevezhető.

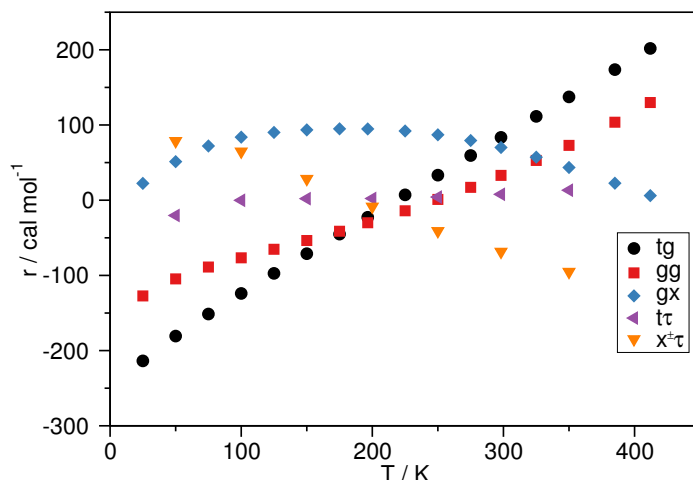
Az 5.12. ábrán látható, hogy a korábbiakhoz hasonlóan ez a modell is szisztematikus hibával terhelt. Az ábrán fel nem tüntetett szerkezetekhez tartozó maradékok hasonló hőmérsékletfüggést mutatnak.

Összességében elmondhatjuk, hogy egy, az entalpiakülönbségek becsült hibáinál nem nagyobb maradékszórású, statisztikailag megfelelő modell felállítása során sikerrel jártunk, bár ez a modell inkább az átmeneti állapotok jellemzésére használható.

### 5.3.6. Butánkonformerek

A modellek külső validálására az *n*-butánkonformerek entalpiakülönbségei<sup>160</sup> alkalmazhatók, azonban a megfelelő deskriptorok értékének korábbi, 5.23 modellbe történő behelyettesítésével ebben az esetben is irreális eredmények adódnak ( $65 - 66 \text{ kcal mol}^{-1}$ ). A tanítóhalmaz viszont tovább bővíthető ezen konformerek entalpiakülönbségeivel, mely egy több vegyület leírására alkalmas modell felállítására is lehetőséget teremt. Az *n*-bután *g* konformerére számított deskriptorok értéke, valamint a 160 hivatkozás adatainak felhasználásával előállított *g*-*t* entalpiakülönbségek az F12. és az F13. táblázatokban tekinthetők meg.

A korábbiakhoz hasonló minőségű eredmény két módosított topológiai index és



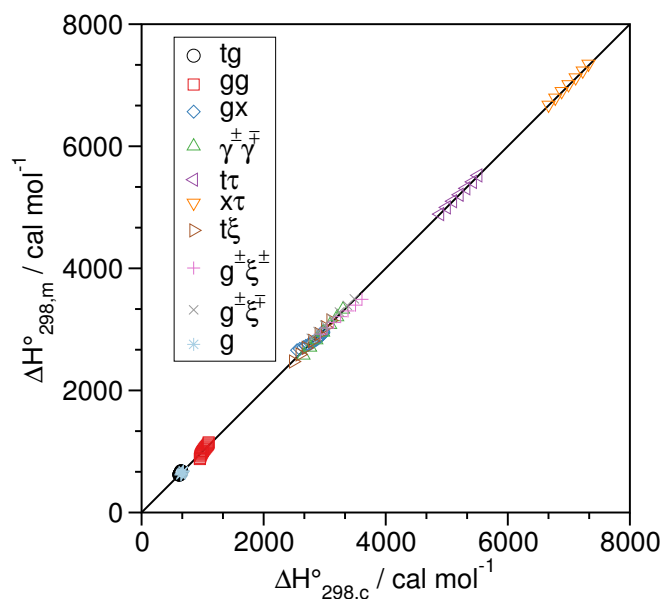
5.12. ábra: Az *n*-pentán konformereire és azok interkonverziós átmeneti állapotaira történő illesztés reziduális hibái néhány szerkezetre a hőmérséklet függvényében.

a hőmérséklet felhasználásával volt elérhető. A modellezés során a Balaban- és a harmonikus-indexek bizonyultak a legmegfelelőbbnek az entalpiakülönbségek leírására, a legjobb 19 modell e két deszkriptort tartalmazza, melyek bemutatásától a köztük fennálló csekély különbségek miatt eltekintünk, csak a legjobbnak mutatózó modellt ismertetjük.

$$\begin{aligned}
 \Delta H_T^\circ = & -(234 \pm 7) \cdot 10^4 \cdot J_m - (131 \pm 3) \cdot 10^6 \cdot J_m^2 + (234 \pm 4) \cdot 10^7 \cdot J_m^3 \\
 & -(68 \pm 1) \cdot HI_m^2 - (78 \pm 8) \cdot 10^{-5} \cdot HI_m^3 + (624 \pm 5) \cdot 10^2 \cdot J_m \cdot HI_m \\
 & -(297 \pm 3) \cdot 10^4 \cdot J_m^2 \cdot HI_m + (390 \pm 80) \cdot 10^2 \cdot J_m \cdot HI_m^2 \\
 & +(0.37 \pm 0.12) \cdot T - (39 \pm 5) \cdot J_m \cdot T - (96 \pm 5) \cdot 10^{-3} \cdot HI_m \cdot T \\
 & +(129 \pm 7) \cdot 10^{-7} \cdot HI_m^2 \cdot T - (128 \pm 5) \cdot 10^2 \\
 R = 1.000 \quad s_r = 40.37 \text{ cal mol}^{-1} \quad F = 16388 \quad m = 97
 \end{aligned} \tag{5.24}$$

A legjobb modell statisztikai mutatóival együtt az 5.24 egyenlettel foglalható össze. Az illesztés a butánkonformerek entalpiakülönbségeit is figyelembe véve 97 pontra történt. Továbbra is elmondható, hogy a modellparaméterek szignifikánsak a *t*-próba alapján és a modelfüggvény illeszkedése az *R* regressziós együttható szerint kiváló. A reziduális szórás értéke mindössze 40.37 cal mol<sup>-1</sup>, míg a 46.95 cal mol<sup>-1</sup> értékű „*leave-one-out*” szórás alapján a modell interpolációra is alkalmasnak bizonyul. A modellezés sikerességét az 5.13. ábrán látható, modell alapján becsült értékek modellezett adatok függvényében történő szemléltetése is remekül mutatja, a pontok gyakorlatilag az ideális viselkedést jelentő egyenesen fekszenek.

Elmondhatjuk tehát, hogy több *n*-alkán (bután és pentán) CPES-ének kritikus pontjaihoz rendelhető szerkezetek az adott vegyület globális minimumához viszonyí-



5.13. ábra: Az *n*-bután g konformerének t konformerre vonatkoztatott entalpiáit is leíró modell által jósolt értékek a modellezett entalpiakülönbségek függvényében.

tott relatív entalpiáinak hőmérsékletfüggését is pontosan leíró modell felállítása során sikerrel jártunk. A felállított modell használatával az entalpiakülönbségek becsléséhez mindössze az alkánmolekulákban megjelenő torziós szögek ismerete szükséges. A modellezés, illetve a definiált deskriptorok a továbbiakban egyéb szerkezetekre, valamint mennyiségekre is könnyen kiterjeszthetők. A topografikus indexek további módosításával elképzelhető, hogy a teljes konformációs potenciális energia felület leírására alkalmas deskriptorok is létrehozhatók, így azok akár dinamikai szimulációk során is felhasználhatóvá válnának.

## 6. Összefoglalás

Doktori munkám három részre osztható: az első rész molekulák statikus szimmetriájának meghatározásáról szól, a második szakaszban az  $n$ -pentánkonformerek entalpiakülönbségeinek számítását ismertettem kvantumkémiai úton és kísérleti adatok újbóli, korrekt kiértékelésével, végül a számított entalpiakülönbségekre a szimmetria és a geometria felhasználásával végzett modellezés eredményét mutattam be, melyet a pentánkonformereinek interkonverziós átmeneti állapotaira és az  $n$ -bután konformereire is kiterjesztettünk.

Az elméleti áttekintésben a szimmetriaanalízis általános halmazelméleti háttérét és a felhasznált vektoralgebrai műveleteket, az alkalmazott kvantumkémiai modellt, az  $n$ -pentán konformereire vonatkozó kísérleti adatok kiértékeléséhez szükséges egyenleteket, valamint a gráfelméleti deszkriptorokkal történő QSPR modellezés alapjait és ennek további lehetőségeit ismertettem.

A szimmetria az elméleti kémiában alapvető szerepet játszik, segítségével kvalitatív megállapítások fogalmazhatók meg, elengedhetetlen egyes kvantitatív számítások során, továbbá alkalmazásával bonyolult numerikus számítások lényegesen leegyszerűsíthetők. Ennek ellenére a különféle szoftverek a pontcsoport meghatározását nem mindig végzik helyesen, egyes esetekben nem is képesek a molekula szimmetriájának megadására, melyet a dolgozatban is ismertettem.

A célunk ezzel kapcsolatban az volt, hogy a statikus szimmetria detektálására egy eddigieknél hatékonyabb módszert dolgozzunk ki, valamint implementáljuk azt. Az általános elméleti háttér ismertetése után annak egyértelműsítéseként ismertettem az általunk definiált algoritmust és az erre épülő programot.

A szoftver hatékonyságát egy általunk összeállított teszhalmazon vizsgáltuk, mely legfeljebb nyolcadrendű szimmetriatengelyeket tartalmazó szerkezetekből áll és az egyes pontcsoportokba tartozó struktúrákat azok gyakoriságával közel arányosan reprezentálja. Az alkalmazott teszhalmaz a későbbiekben is használható például további szimmetriaanalízisre szolgáló programok vizsgálatára.

A szoftver teljesítőképességének megállapítására végzett vizsgálat egyértelműen bizonyítja, hogy az általunk definiált algoritmus és a megvalósítására írt kód több, önálló és molekulamodellező szoftvert a teszhalmazra nézve mind számítási idő, mind a meghatározott pontcsoport helyessége szempontjából felülmúl. A program képes továbbá egyéb szimmetriával összefüggő tulajdonságok meghatározására, valamint további számításokhoz használható, tetszőleges részcsoporthoz megfelelő, deformálatlan szerkezet előállítására.

Az általunk fejlesztett kód természetesen más programokban is felhasználható növelve azok hatékonyságát, melyet két szoftverbe történő integrációval mutattunk be. A vizsgálathoz használt teszthalmaz és a felírt algoritmus implementációja szabad szoftverként hozzáférhető.

Az  $n$ -pentán konformerekhez tartozó relatív entalpiák számítása során egy, az  $n$ -bután esetén már sikerrel alkalmazotthoz hasonló fókuszpont analízist használtunk, mely lehetőséget teremt a számított értékekhez történő hiba hozzárendelésére is. A modell CCSD(T)/cc-pVTZ referencia geometrián történő további (főként) CC számításokra épül, a bázist növelve a geometriai optimalizálás során – legalább háromszoros- $\zeta$  minőségű bázis alkalmazva a korrekciók számítására – az eredmény nem változik. Az egyes járulékok meghatározása során a CCSDT és CCSDT(Q) számításoktól eltekintve az összes elektront korreláltattuk.

A teljes energia a HF/cc-pV6Z szinten előállított elektronenergia, valamint további korrekciók összegeként áll elő. Az egyedi járulékokat MP2/cc-pV(5,6)Z, CCSD/cc-pV(4,5)Z és CCSD(T)/cc-pV(4,5)Z szinten végtelen bázisra történő extrapolációval állapítottuk meg. Magasabb rendű korrelációs effektusok számítása  $C_1$  szimmetriájú konformerekre CCSDT/cc-pVDZ, a többi konformerre CCSDT(Q)/cc-pVDZ szinten történt.

A pontosság további növelése érdekében végrehajtottuk a Born–Oppenheimer-közelítés diagonális korrekcióját és a skaláris relativisztikus effektusok hozzájárulásának becslését. A zérusponthoz tartozó rezgési energia számítása során a CCSD(T)/cc-pVTZ harmonikus rezgési analízis mellett anharmonikus korrekciót és a metil, valamint a C–C–C–C torziókra mint gátolt belső rotációkra történő korrekciót is meghatároztunk.

Megállapítottuk, hogy a HF energia, valamint a tg konformer esetén az MP2 hozzájárulás értéke gyorsan konvergál, a többi taggal ellentétben. A ZPVE számítása komoly problémát okoz, az ehhez számított korrekciók értéke viszonylag nagy, és a teljes energia hibájának mintegy fele ebből a tagból származik. Ellenben a magasabb rendű korrekciós tagok (DBOC és skaláris relativisztikus korrekció) nagyrészt kiejtik egymást, inkább a végeredmény hibájához járulnak hozzá.

Az entalpiák előállítása a merev rotor-harmonikus oszcillátor és az ideális gáz közelítések figyelembevételével a statisztikus termodinamika klasszikus formulái segítségével történt. Ezeket az értékeket korábbi számításokkal összevetve elmondható, hogy az  $n$ -pentán konformereinek egységes kezelése nehézkes, egyes módszerek az egyes szerkezetek esetén különbözőképp közelítik a fókuszpont analízis eredményét, mely azokat pontosságban egyértelműen felülmúlja.

A kísérleti eredmények nem egységesek, egyes esetekben a relatív entalpiák meghatározása során a végeredményt egy, a gg konformerhez rendelhető átmenet is zárhatja. Az erre az átmenetre történő korrekciót elvégezve a saját eredményeinkkel sokkal jobb egyezést mutató értékeket kaptunk.

A pentánkonformerek entalpiakülönbségei korrelálni látszanak azok szimmetriájával, ezért megpróbáltunk azokra egy modellt felállítani molekulaszimmetrián alapuló deskriptorok segítségével, kizárólag a 298.15 K hőmérsékleten számított három adatra, ez azonban nem vezetett sikerre. Korábbi tanulmányok alapján termokémiai mennyiségek jól modellezhetők módosított topológiai (topografikus) indexek segítségével, így a CCSD(T)/cc-pVTZ szinten optimalizált szerkezetekből nyert torziós szögek segítségével a topológiai indexek bázisán új indexeket definiáltunk, azonban a modellezés ezeknek az indexeknek a segítségével sem volt sikeres.

Egy megfelelő, a hőmérsékletfüggést is leíró modell felállítása érdekében az  $n$ -pentán konformációs entalpiáit a kvantumkémiai eredmények alapján különböző hőmérsékleteken előállítottuk. Az ezekre történő illesztéssel a korábbiaknál már egy szerkezeti deskriptor és a hőmérséklet mint változók segítségével is jelentősen jobb modell állítható fel. A modellezésre az általunk módosított topológiai indexek a szimmetria deskriptoroknál alkalmasabbnak bizonyultak.

A nemlineáris kapcsolatot a függő és a független változók között egy többváltozós hatványsor alakú modellfüggvénnyel írtuk le, az irodalomban található hasonló modellek mind egyetlen változóra szorítkoznak. Ezen modellezés során azt tapasztaltuk, hogy a modell már kvadratikus tagok figyelembevételével is számottevően javult.

A felállított modell nem bizonyult alkalmasnak az  $n$ -pentán konformereinek interkonverziós átmeneti állapotainak leírására, azonban egy új, a konformereket és az elsőfajú nyeregponthoz egyszerűen jellemző modell felállítása során sikerrel jártunk.

A modellezés további kiterjesztése az  $n$ -bután két konformerének entalpiakülönbségére a korábbiakhoz hasonló predikciós képességű modellt szolgáltatott. Egy, a hőmérséklettel együtt háromváltozós, köbös modellfüggvény segítségével a két molekula konformereinek és az  $n$ -pentán esetén a konformációs izomerizáció átmeneti állapotainak relatív entalpiáit rendkívül alacsony maradékszórással sikerült leírni.

Ezek alapján az újonnan definiált módosított topológiai indexek jól korrelálnak az  $n$ -pentán konformációs izomereinek entalpiakülönbségeivel, egy a hőmérséklet és két topografikus index felhasználásával felállított modell képes leírni a CPES összes fontos kritikus pontját. Valószínűsíthető, hogy ezek a deskriptorok egyéb termokémiai mennyiségekkel is jó korrelációt mutatnak, ennek bizonyítására azonban további vizsgálatok szükségesek.

Mivel az eredmények jól mutatják az általunk definiált leírók alkalmasságát termokémiai adatok modellezésére a konformációs potenciális energia felület fontos kritikus pontjaiban, elképzelhető, hogy ezek a deskriptorok – esetleg kisebb módosítások után – megfelelőek a teljes CPES kvantitatív jellemzésére, mely egy-egy adatpont számítását, vagy akár egy dinamikai szimuláció végrehajtását is jelentősen lerövidítheti. A felállított modellek azonban extrapolációra nem bizonyultak megfelelőnek, erre a célra mindenképp a deskriptorok bizonyos mértékű átdolgozása szükséges.



## 7. Summary

My doctoral work may be divided into three parts: the first is about the determination of the static symmetry of molecules; in the second part I reviewed the computation of the enthalpy differences for *n*-pentane conformers by means of quantum chemistry as well as proper reevaluation of experimental data; finally, I presented the modeling results of the enthalpy differences utilizing molecular symmetry and geometry which we also extended to the transition states of the conformational interconversion for *n*-pentane and the conformers for *n*-butane.

In the theoretical overview part I described the general set-theoretic background of symmetry analysis along with the applied vector algebraic operations, the employed quantum chemical model, the necessary equations to evaluate the experimental data for the conformers of *n*-pentane together with the foundations of QSPR modeling with graph theoretical descriptors and its further possibilities.

Symmetry plays an essential role in theoretical chemistry. Relying on symmetry analysis, qualitative results may be formulated, its use is also indispensable in certain quantitative calculations, furthermore, it allows complicated computations to be simplified significantly. Still, existing software does not always detect the correct point group, sometimes they even fail to determine any symmetry at all.

Regarding this problem, our goal was to develop and implement a method for static symmetry analysis that is more efficient and reliable than the previous ones. After explaining the general theory, I specified the involved sets and mappings, describing the new algorithm and the program written to carry out the point group detection.

The performance of the program was demonstrated on a test set we assembled for this purpose. The set contains structures with axes up to eighth order and represents the individual point groups approximately proportional to their occurrence. This set may also be used later to test methods and programs aiming to identify molecular symmetry.

The study on the efficiency of the code clearly pointed out that our method and its implementation is superior to the existing standalone and electronic structure software considering both the computational time and the correctness of the point groups found for the structures in our test set. Furthermore, the program is capable of the determining of other symmetry related properties and generating undistorted geometry resembling an arbitrary subgroup of the molecule for further calculations.

Our code can be integrated into other programs to increase their effectiveness what we have demonstrated by implementing it in two quantum chemical software. Our

well-constructed test set as well as our symmetry program called SYVA are both freely available.

The determination of the enthalpy differences of *n*-pentane conformers was carried out by means of a focal point analysis similar to the one previously used in case of *n*-butane. A great advantage of this approach is the possibility of assigning well defined uncertainties to the calculated values. The model is based on CC energy contributions computed using a CCSD(T)/cc-pVTZ reference geometry. We pointed out that the energy differences of the conformers change only slightly with the increase of the applied basis set if at least a triple- $\zeta$  basis is used to calculate the energy contributions. The various correction terms were calculated correlating all electrons, except the most time consuming CCSDT and CCSDT(Q) contributions where the frozen-core approximation was invoked.

The total energy was constructed as the sum of the electronic energy at the HF/cc-pV6Z level and various further corrections. The individual terms were determined at the MP2/cc-pV(5,6)Z, CCSD/cc-pV(4,5)Z and CCSD(T)/cc-pV(4,5)Z levels of theory by extrapolation to the complete basis. Higher order correlation effects were calculated using the CCSDT/cc-pVDZ method for structures with  $C_1$  symmetry and the CCSDT(Q)/cc-pVDZ method for other conformers.

To further improve the precision of the results, we performed the diagonal correction of the Born–Oppenheimer approximation and estimated the scalar relativistic contributions. The CCSD(T)/cc-pVTZ harmonic zero point vibrational energy was improved through anharmonic correction and treating the methyl- and C–C–C–C torsions as hindered internal rotations.

We concluded that the HF energy and, for the tg conformer, the MP2 energy converges quickly as opposed to other contributions. The calculation of both the harmonic and anharmonic parts of the ZPVE is challenging, therefore, this contribution introduces about half the uncertainty of the total energy. The small corrections like the diagonal Born–Oppenheimer correction and the relativistic correction, on the other hand, almost cancel out, contributing mainly to the error of the final results.

To generate the enthalpies the common formulas of statistical thermodynamics were invoked applying the usual rigid rotor-harmonic oscillator and ideal gas approximations. Comparing these values with the results of previous theoretical studies we can conclude that the uniform description of the conformers is complicated.

In general, the experimental result are not consistent. In some cases the outcome may be biased as a result of a vibrational band that can be assigned to the gg conformer. Performing a correction to this transition, much more consistent values with our results can be derived.

As the enthalpy differences of the *n*-pentane conformers seem to correlate with the symmetry of the corresponding structures, we attempted to establish a model

applying descriptors based on molecular symmetry, however, the modeling of the three data points at 298.15 K was unsuccessful. Since, based on previous investigations, thermochemical quantities correlate well with topological and, especially, topographical indices, we repeated the modeling with newly defined indices based on topological descriptors using the torsional angles appearing in the CCSD(T)/cc-pVTZ optimized structures, however this attempt also failed to produce an appropriate model.

With the help of the quantum chemical results, the relative enthalpies of the  $n$ -pentane conformers were calculated at various temperatures in order to build a suitable model that also describes the temperature dependence of this quantity. Compared to the earlier models, much better results could be achieved fitting these data even using only one structural descriptor and the temperature as independent variables. The modified topological indices we defined proved to be more appropriate for the description.

In some cases we also studied the applicability of a multivariate power series as a model function, contrarily to previously reported models which rely exclusively on a single variable. We may conclude that involving higher order terms greatly improves the results, even quadratic terms vastly increase the predictive capabilities of the model.

These models are unfortunately not able to describe the transition states of the rotational isomerization on  $n$ -pentane conformers. Nevertheless, we successfully derived a new model, simultaneously characterizing both the local minima and the first order saddle points of the CPES.

Further extensions to the above model to treat the conformers of  $n$ -butane can also be defined. A cubic model with three variables – including the temperature – was adequate to explain the enthalpy differences of  $n$ -pentane conformers and transition states as well as  $n$ -butane conformers at the same time.

Taking everything into consideration we may conclude that the modified topological indices we defined correlate remarkably well with the enthalpy differences of conformational isomers of  $n$ -pentane. All important critical points on the CPES can be described by a model with the temperature and two topographical indices as independent variables. Presumably, these descriptors show good correlation with other thermochemical properties as well, though this assumption can only be verified by further investigations.

Since the results obviously show the suitability of these modified topological indices to model thermochemical data in all important critical points of the CPES, it is also plausible that they are appropriate for the characterization of the full CPES. Such a model would considerably decrease the time needed to calculate either a single data point or to carry out a molecular dynamics simulation. The present models are not applicable for extrapolation, for this purpose further modification of the indices is necessary.

# Köszönetnyilvánítás

Köszönöm témavezetőmnek, Dr. Tasi Gyulának, hogy tanulmányaim során felkeltette és fenntartotta érdeklődésemet az elméleti kémia iránt. Hálás vagyok mind a szakmai, mind az emberi segítségéért és támogatásáért, melyet egyetemi éveim alatt nyújtott. Köszönöm, hogy a vezetése mellett lehetőségem nyílt nemcsak a dolgozatban szereplő témák, hanem az elméleti kémia számos más területének megismerésére is.

Köszönettel tartozom a Budapesti Műszaki és Gazdaságtudományi Egyetem Fizikai Kémia és Anyagtudományi Tanszékén működő Spektroszkópia Csoport vezetőjének, Dr. Kállay Mihálynak, valamint a csoport korábbi tagjainak, Dr. Csontos Józsefnek és Dr. Nagy Balázsnak a pentán konformereire végzett számítások végrehajtásában és a dolgozat alapjául szolgáló egyik cikk megírásában nyújtott segítségükért. Dr. Nagy Balázsnak külön köszönöm, hogy időt szakított a dolgozat átolvasására és annak javításához számos javaslattal járult hozzá.

Köszönöm Dr. Tóth Ágotának és Dr. Horváth Dezsőnek, hogy a náluk végzett munka mellett is időt biztosítottak a dolgozat megírására, továbbá azt számos jó tanáccsal segítették.

Köszönet illeti egykori évfolyamtársamat, Tóbiás Rolandot, töretlen segítőkészségéért és a dolgozattal kapcsolatos számlálhatatlan javaslataért. Hálás vagyok, amiért az évek során igaz barátként bármikor számíthattam rá.

Hálával tartozom Szüleimnek, akik önfeláldozó munkájukkal és szeretetükkel támogattak egyetemi és doktori tanulmányaim alatt egyaránt.

# Irodalomjegyzék

- <sup>1</sup> K. Piszczatowski, G. Łach, M. Przybytek, J. Komasa, K. Pachuki, J. Jeziorski, *J. Chem. Theory Comput.* **2009**, *5*, 3039.
- <sup>2</sup> K. Pachucki, J. Komasa, *Phys. Chem. Chem. Phys.* **2010**, *12*, 9188.
- <sup>3</sup> C. Gonzalez, H. Schlegel, *J. Chem. Phys.* **1989**, *90*, 2154.
- <sup>4</sup> C. Gonzalez, H. Schlegel, *J. Phys. Chem.* **1990**, *94*, 5523.
- <sup>5</sup> C. Gonzalez, H. Schlegel, *J. Chem. Phys.* **1991**, *95*, 5853.
- <sup>6</sup> P. Pulay, *Chem. Phys. Lett.* **1980**, *73*, 393 .
- <sup>7</sup> P. Pulay, *J. Comput. Chem.* **1982**, *3*, 556.
- <sup>8</sup> P. Császár, P. Pulay, *J. Mol. Struct.* **1984**, *114*, 31 .
- <sup>9</sup> K. N. Kudin, G. E. Scuseria, E. Cancès, *J. Chem. Phys.* **2002**, *116*, 8255.
- <sup>10</sup> X. Li, M. J. Frisch, *J. Chem. Theory Comput.* **2006**, *2*, 835.
- <sup>11</sup> D. Cook. *Handbook of Computational Quantum Chemistry*. Oxford University Press, Oxford **1998**.
- <sup>12</sup> U. Vonbarth, C. Gelatt, *Phys. Rev. B* **1980**, *21*, 2222.
- <sup>13</sup> F. Jensen. *Introduction to Computational Chemistry*. Wiley, Chichester **2013**.
- <sup>14</sup> A. Hinchliffe, P. Popelier, T. Simos, S. Wilson. *Chemical Modelling: Applications and Theory*. Royal Society of Chemistry, Cambridge **2000**.
- <sup>15</sup> C. Teichteil, L. Marona, V. Vallet. In P. Schwerdtfeger, editor, *Relativistic Electronic Structure Theory*, volume 14 of *Theoretical and Computational Chemistry*. Elsevier **2004**. pages 476 – 551.
- <sup>16</sup> J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496.
- <sup>17</sup> B. I. Dunlap, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113.
- <sup>18</sup> B. Nagy, F. Jensen. In A. L. Parrill, K. B. Lipkowitz, editors, *Reviews in Computational Chemistry, Volume 30*, chapter 3. John Wiley & Sons, Inc., Hoboken **2017**. pages 93–149.

- <sup>19</sup> M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347.
- <sup>20</sup> M. S. Gordon, M. W. Schmidt. In C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria, editors, *Theory and Applications of Computational Chemistry*. Elsevier, Amsterdam **2005**. pages 1167 – 1189.
- <sup>21</sup> M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.
- <sup>22</sup> M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong, *Comput. Phys. Commun.* **2010**, *181*, 1477.
- <sup>23</sup> F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73.
- <sup>24</sup> Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, M. Kállay, *J. Chem. Phys.* **2013**, *139*, 094105.
- <sup>25</sup> A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, D. A. Dobchev, *Chem. Rev.* **2010**, *110*, 5714.
- <sup>26</sup> B. Bogdanov, S. Nikolić, N. Trinajstić, *J. Math. Chem.* **1989**, *3*, 299.
- <sup>27</sup> M. Randić, *New J. Chem.* **2000**, *24*, 165.
- <sup>28</sup> M. Randić, N. Basak, D. Plav, M. Trunk Highway, *Croat. Chem. Acta* **2004**, *77*, 251.
- <sup>29</sup> E. Noether, *Transp. Theory Stat. Phys.* **1971**, *1*, 186.
- <sup>30</sup> F. Török, P. Pulay. *Elméleti Kémia I*. Nemzeti Tankönyvkiadó, Budapest **1995**.

- <sup>31</sup> J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay, *Phys. Rev. Lett.* **1964**, *13*, 138.
- <sup>32</sup> G. Luders, *Kong. Dan. Vid. Sel. Mat. Fys. Med.* **1954**, *28N5*, 1.
- <sup>33</sup> W. Pauli. *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of His Seventieth Birthday*. Pergamon Press, London **1955**.
- <sup>34</sup> J. Avery, S. Rettrup, J. Avery. *Symmetry-adapted Basis Sets: Automatic Generation for Problems in Chemistry and Physics*. World Scientific, Singapore **2011**.
- <sup>35</sup> M. Johansson, V. Veryazov, *J. Cheminform.* **2017**, *9*, 8.
- <sup>36</sup> J. Pople, *J. Am. Chem. Soc.* **1980**, *102*, 4615.
- <sup>37</sup> R. L. Flurry, *J. Am. Chem. Soc.* **1981**, *103*, 2901.
- <sup>38</sup> D. H. McDaniel, *J. Phys. Chem.* **1981**, *85*, 479.
- <sup>39</sup> E. Kryachko, *Int. J. Quantum Chem.* **1982**, *22*, 657.
- <sup>40</sup> V. Sokolov, *Fullerene Sci. Technol.* **1997**, *5*, 281.
- <sup>41</sup> K. Nedwed, K. Gatterer, H. Fritzer, *Comput. Chem. (Oxford)* **1994**, *18*, 371.
- <sup>42</sup> K. Balasubramanian, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 761.
- <sup>43</sup> T. Pilati, A. Forni, *J. Appl. Crystallogr.* **1998**, *31*, 503.
- <sup>44</sup> T. Pilati, A. Forni, *J. Appl. Crystallogr.* **2000**, *33*, 417.
- <sup>45</sup> T. Pilati, A. Forni. SYMMOL **2000**.  
<http://www.ccp14.ac.uk/ccp/web-mirrors/symmol/~pila/>, utolsó hozzáférés: 2018. 05. 10.
- <sup>46</sup> T. Pilati, A. Forni, A. Gilbert. SYMMOL **2011**.  
<https://github.com/nutjunkie/IQmol/blob/master/src/Main/symmol.f90>, utolsó hozzáférés: 2018. 05. 10.
- <sup>47</sup> R. J. Largent, W. F. Polik, J. R. Schmidt, *J. Comput. Chem.* **2012**, *33*, 1637.
- <sup>48</sup> M. Razinger, K. Balasubramanian, M. Perdih, M. E. Munk, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 812.
- <sup>49</sup> S. Bohanec, M. Perdih, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 719.
- <sup>50</sup> J. Ivanov, G. Schüürmann, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 728.

- <sup>51</sup> J. Ivanov, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 596.
- <sup>52</sup> K. Balasubramanian, *Int. J. Quantum Chem.* **1982**, *21*, 411.
- <sup>53</sup> L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, *J. Chem. Phys.* **1991**, *94*, 7221.
- <sup>54</sup> J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, L. A. Curtiss, *J. Chem. Phys.* **1989**, *90*, 5622.
- <sup>55</sup> L. A. Curtiss, K. Raghavachari, P. C. Redfern, A. G. Baboul, J. A. Pople, *Chem. Phys. Lett.* **1999**, *314*, 101 .
- <sup>56</sup> L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **2007**, *126*, 084108.
- <sup>57</sup> G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, J. Mantzaris, *J. Chem. Phys.* **1988**, *89*, 2193.
- <sup>58</sup> G. A. Petersson, M. A. Al-Laham, *J. Chem. Phys.* **1991**, *94*, 6081.
- <sup>59</sup> G. A. Petersson, T. G. Tensfeldt, J. A. Jr.Montgomery, *J. Chem. Phys.* **1991**, *94*, 6091.
- <sup>60</sup> J. A. Jr.Montgomery, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1994**, *101*, 5900.
- <sup>61</sup> J. W. Ochterski, G. A. Petersson, J. A. Jr.Montgomery, *J. Chem. Phys.* **1996**, *104*, 2598.
- <sup>62</sup> J. A. J. Montgomery, M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1999**, *110*, 2822.
- <sup>63</sup> J. A. Jr.Montgomery, M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **2000**, *112*, 6532.
- <sup>64</sup> J. P. Perdew, A. Ruzsinszky, L. A. Constantin, J. Sun, G. I. Csonka, *J. Chem. Theory Comput.* **2009**, *5*, 902.
- <sup>65</sup> D. Bakowies, *J. Phys. Chem. A* **2013**, *117*, 228.
- <sup>66</sup> D. Feller, K. A. Peterson, *J. Chem. Phys.* **1999**, *110*, 8384.
- <sup>67</sup> J. M. Martin, T. J. Lee, *Chem. Phys. Lett.* **1996**, *258*, 136 .
- <sup>68</sup> T. Helgaker, W. Klopper, H. Koch, J. Noga, *J. Chem. Phys.* **1997**, *106*, 9639.
- <sup>69</sup> D. G. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45 .



- <sup>70</sup> P. L. Fast, M. L. Sánchez, D. G. Truhlar, *J. Chem. Phys.* **1999**, *111*, 2921.
- <sup>71</sup> A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243 .
- <sup>72</sup> W. Klopper, *Mol. Phys.* **2001**, *99*, 481.
- <sup>73</sup> E. F. Laschuk, P. R. Livotto, *J. Chem. Phys.* **2004**, *121*, 12146.
- <sup>74</sup> D. W. Schwenke, *J. Chem. Phys.* **2005**, *122*, 014107.
- <sup>75</sup> T. Helgaker, W. Klopper, D. P. Tew, *Mol. Phys.* **2008**, *106*, 2107.
- <sup>76</sup> J. M. L. Martin, G. de Oliveira, *J. Chem. Phys.* **1999**, *111*, 1843.
- <sup>77</sup> A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, J. Gauss, *J. Chem. Phys.* **2004**, *120*, 4129.
- <sup>78</sup> A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, *J. Chem. Phys.* **2006**, *125*, 144108.
- <sup>79</sup> A. Wolf, M. Reiher, B. A. Hess, *J. Chem. Phys.* **2002**, *117*, 9215.
- <sup>80</sup> M. Reiher, A. Wolf, *J. Chem. Phys.* **2004**, *121*, 2037.
- <sup>81</sup> M. Reiher, A. Wolf, *J. Chem. Phys.* **2004**, *121*, 10945.
- <sup>82</sup> A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, J. F. Stanton, *J. Chem. Phys.* **2004**, *121*, 11599.
- <sup>83</sup> Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, J. F. Stanton, *J. Chem. Phys.* **2006**, *125*, 064108.
- <sup>84</sup> N. J. DeYonker, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *124*, 114104.
- <sup>85</sup> N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *125*, 104111.
- <sup>86</sup> D. S. Ho, N. J. DeYonker, A. K. Wilson, T. R. Cundari, *J. Phys. Chem. A* **2006**, *110*, 9767.
- <sup>87</sup> N. J. DeYonker, D. S. Ho, A. K. Wilson, T. R. Cundari, *J. Phys. Chem. A* **2007**, *111*, 10776.
- <sup>88</sup> N. J. DeYonker, K. A. Peterson, G. Steyl, A. K. Wilson, T. R. Cundari, *J. Phys. Chem. A* **2007**, *111*, 11269.

- <sup>89</sup> T. V. Grimes, A. K. Wilson, N. J. DeYonker, T. R. Cundari, *J. Chem. Phys.* **2007**, *127*, 154117.
- <sup>90</sup> N. J. DeYonker, B. Mintz, T. R. Cundari, A. K. Wilson, *J. Chem. Theory Comput.* **2008**, *4*, 328.
- <sup>91</sup> P. L. Fast, M. L. Sánchez, D. G. Truhlar, *Chem. Phys. Lett.* **1999**, *306*, 407 .
- <sup>92</sup> P. L. Fast, J. Corchado, M. L. Sánchez, D. G. Truhlar, *J. Phys. Chem. A* **1999**, *103*, 3139.
- <sup>93</sup> P. L. Fast, J. C. Corchado, M. L. Sánchez, D. G. Truhlar, *J. Phys. Chem. A* **1999**, *103*, 5129.
- <sup>94</sup> P. L. Fast, M. L. Sánchez, J. C. Corchado, D. G. Truhlar, *J. Chem. Phys.* **1999**, *110*, 11679.
- <sup>95</sup> Y. Zhao, B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 4786.
- <sup>96</sup> Y. Zhao, B. J. Lynch, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.
- <sup>97</sup> E. E. Dahlke, M. A. Orthmeyer, D. G. Truhlar, *J. Phys. Chem. B* **2008**, *112*, 2372.
- <sup>98</sup> D. A. Dixon, D. Feller, *J. Phys. Chem. A* **1998**, *102*, 8209.
- <sup>99</sup> D. Feller, D. A. Dixon, J. S. Francisco, *J. Phys. Chem. A* **2003**, *107*, 1604.
- <sup>100</sup> D. Feller, K. A. Peterson, W. A. de Jong, D. A. Dixon, *J. Chem. Phys.* **2003**, *118*, 3510.
- <sup>101</sup> D. Feller, N. C. Craig, *J. Phys. Chem. A* **2009**, *113*, 1601.
- <sup>102</sup> D. Feller, K. A. Peterson, D. A. Dixon, *J. Phys. Chem. A* **2011**, *115*, 1440.
- <sup>103</sup> D. Feller, J. M. Simmie, *J. Phys. Chem. A* **2012**, *116*, 11768.
- <sup>104</sup> J. Laane, M. Dakkouri, B. van der Veken, H. Oberhammer. *Structures and Conformations of Non-Rigid Molecules*. Springer Science & Business Media **2012**.
- <sup>105</sup> A. G. Császár, W. D. Allen, H. F. S. III, *J. Chem. Phys.* **1998**, *108*, 9751.
- <sup>106</sup> B. Ruscic, J. E. Boggs, A. Burcat, A. G. Császár, J. Demaison, R. Janoschek, J. M. L. Martin, M. L. Morton, M. J. Rossi, J. F. Stanton, P. G. Szalay, P. R. Westmoreland, F. Zabel, T. Bérces, *J. Phys. Chem. Ref. Data* **2005**, *34*, 573.
- <sup>107</sup> A. L. L. East, W. D. Allen, *J. Chem. Phys.* **1993**, *99*, 4638.

- <sup>108</sup> G. Czako, E. Mátyus, A. C. Simmonett, A. G. Császár, H. F. Schaefer III, W. D. Allen, *J. Chem. Theory Comput* **2008**, *4*, 1220.
- <sup>109</sup> T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.
- <sup>110</sup> S. S. Xantheas, T. H. Dunning, *J. Phys. Chem.* **1993**, *97*, 18.
- <sup>111</sup> T. H. Dunning, K. A. Peterson, A. K. Wilson, *J. Chem. Phys.* **2001**, *114*, 9244.
- <sup>112</sup> R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796.
- <sup>113</sup> D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1994**, *100*, 2975.
- <sup>114</sup> D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1995**, *103*, 4572.
- <sup>115</sup> K. A. Peterson, T. H. Dunning, *J. Chem. Phys.* **2002**, *117*, 10548.
- <sup>116</sup> P. Thanikaivelan, V. Subramanian, J. R. Rao, B. U. Nair, *Chem. Phys. Lett.* **2000**, *323*, 59 .
- <sup>117</sup> N. Adams, J. Clauss, M. Meunier, U. Schubert, *Mol. Simul.* **2006**, *32*, 125.
- <sup>118</sup> A. Vatani, M. Mehrpooya, F. Gharagheizi, *Int. J. Mol. Sci.* **2007**, *8*, 407.
- <sup>119</sup> E. Estrada, L. Torres, L. Rodríguez, I. Gutman, *Indian J. Chem., Sect A* **1998**, *37*, 849.
- <sup>120</sup> M. Hechinger, K. Leonhard, W. Marquardt, *J. Chem. Inf. Model.* **2012**, *52*, 1984.
- <sup>121</sup> L. P. Hammett, *J. Am. Chem. Soc.* **1937**, *59*, 96.
- <sup>122</sup> R. Todeschini, M. Lasagni, E. Marengo, *J. Chemom.* **1994**, *8*, 263 .
- <sup>123</sup> R. Todeschini, P. Gramatica, R. Provenzani, E. Marengo, *Chemom. Intell. Lab. Syst.* **1995**, *27*, 221 .
- <sup>124</sup> V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693.
- <sup>125</sup> M. Randic, *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- <sup>126</sup> H. Hosoya, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.
- <sup>127</sup> A. Balaban, *MATCH (Commun. Math. Chem.)* **1986**, *21*, 115.
- <sup>128</sup> H. Wiener, *J. Am. Chem. Soc.* **1947**, *69*, 17.
- <sup>129</sup> S. S. Godavarthy, R. L. Robinson, K. A. M. Gasem, *Ind. Eng. Chem. Res.* **2006**, *45*, 5117.

- <sup>130</sup> H. Haruo, K. Keiko, M. Kimiko, *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3415.
- <sup>131</sup> M. Randic, *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- <sup>132</sup> P. A. Bath, A. R. Poirrette, P. Willett, F. H. Allen, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 714.
- <sup>133</sup> L. Kier, L. Hall. *Molecular Connectivity in Structure-Activity Analysis*. John Wiley & Sons, Letchworth **1986**.
- <sup>134</sup> L. Kier, L. Hall. *Molecular Structure Description: The Electrotopological State*. Academic Press, San Diego **1999**.
- <sup>135</sup> I. V. Tetko, V. Y. Tanchuk, T. N. Kasheva, A. E. P. Villa, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488.
- <sup>136</sup> J. Huuskonen, *Environ. Toxicol. Chem.* **2001**, *20*, 491.
- <sup>137</sup> L. H. Hall, C. T. Story, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004.
- <sup>138</sup> K. Rose, L. H. Hall, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651.
- <sup>139</sup> R. R. A. Abou-Shaaban, H. A. Al-Khamees, H. S. Abou-Auda, A. P. Simonelli, *Pharm. Res.* **1996**, *13*, 129.
- <sup>140</sup> J. Buolamwini, K. Raghavan, M. R. Fesen, Y. Pommier, K. Kohn, J. N. Weinstein, *Pharm. Res.* **1997**, *13*, 1892.
- <sup>141</sup> M. Randć, *Int. J. Quantum Chem.* **1988**, *34*, 201.
- <sup>142</sup> M. Randić, B. Jerman-Blažić, N. Trinajstić, *Comput. Chem. (Oxford)* **1990**, *14*, 237 .
- <sup>143</sup> E. Estrada, L. A. Montero, *Molecular Engineering* **1993**, *2*, 363.
- <sup>144</sup> E. Estrada, A. Ramírez, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837.
- <sup>145</sup> E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 708.
- <sup>146</sup> M. Randić, M. Razinger, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 140.
- <sup>147</sup> M. V. Diudea, D. Horvath, A. Graovac **1995**, *35*, 129.
- <sup>148</sup> H. P. Schultz, E. B. Schultz, T. P. Schultz, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 864.
- <sup>149</sup> E. Estrada, *Chem. Phys. Lett.* **2000**, *319*, 713 .

- <sup>150</sup> Y. Ren, J. Qin, H. Liu, X. Yao, M. Liu, *QSAR Comb. Sci.* **2009**, *28*, 1237.
- <sup>151</sup> A. R. Katritzky, U. Maran, M. Karelson, V. S. Lobanov, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913.
- <sup>152</sup> S. S. Godavarthy, R. L. Robinson, K. A. M. Gasem, *Ind. Eng. Chem. Res.* **2006**, *45*, 5117.
- <sup>153</sup> G. Ding, J. Chen, X. Qiao, L. Huang, J. Lin, X. Chen, *Chemosphere* **2006**, *62*, 1057 .
- <sup>154</sup> R. Abramowitz, S. H. Yalkowsky, *Pharm. Res.* **1990**, *7*, 942.
- <sup>155</sup> A. R. Katritzky, U. Maran, M. Karelson, V. S. Lobanov, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913.
- <sup>156</sup> R. Todeschini, V. Consonni. *Handbook of molecular descriptors*. WileyVCH, Weinheim **2000**.
- <sup>157</sup> C. W. Yap, H. Li, Z.-L. Ji, Y. Chen, *Mini-Rev. Med. Chem.* **2007**, *7*, 1097.
- <sup>158</sup> P. Liu, W. Long, *Int. J. Mol. Sci.* **2009**, *10*, 1978.
- <sup>159</sup> S. Altmann. *Rotations, Quaternions, and Double Groups*. Oxford University Press, Oxford **1986**.
- <sup>160</sup> D. Barna, B. Nagy, J. Csontos, A. G. Császár, G. Tasi, *J. Chem. Theory Comput.* **2012**, *8*, 479.
- <sup>161</sup> D. Feller, K. A. Peterson, D. A. Dixon, *J. Chem. Phys.* **2008**, *129*, 204105.
- <sup>162</sup> D. Feller, K. A. Peterson, J. G. Hill, *J. Chem. Phys.* **2011**, *135*, 044102.
- <sup>163</sup> C. Schwartz, *Phys. Rev.* **1962**, *126*, 1015.
- <sup>164</sup> K. Rao, C. Mathews. *Molecular spectroscopy: modern research*. Academic Press **1972**.
- <sup>165</sup> J. Gauss, J. F. Stanton, *Chem. Phys. Lett.* **1997**, *276*, 70 .
- <sup>166</sup> J. F. Stanton, J. Gauss, *Int. Rev. Phys. Chem.* **2000**, *19*, 61.
- <sup>167</sup> C. C. Marston, G. G. Balint-Kurti, *J. Chem. Phys.* **1989**, *91*, 3571.
- <sup>168</sup> G. G. Balint-Kurti, R. N. Dixon, C. C. Marston, *Int. Rev. Phys. Chem.* **1992**, *11*, 317.
- <sup>169</sup> K. S. Pitzer, *J. Chem. Phys.* **1946**, *14*, 239.

- <sup>170</sup> J. E. Kilpatrick, K. S. Pitzer, *J. Chem. Phys.* **1949**, *17*, 1064.
- <sup>171</sup> J.F. Stanton, J. Gauss, L. Cheng, M.E. Harding, D.A. Matthews, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowitz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see <http://www.cfour.de>.
- <sup>172</sup> Mrcc, a quantum chemical program suite written by M. Kállay, Z. Rolik, J. Csontos, P. Nagy, G. Samu, D. Mester, J. Csóka, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, and B. Hégyel. See also Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, *J. Chem. Phys.* **139**, 094105 (2013), as well as: [www.mrcc.hu](http://www.mrcc.hu).
- <sup>173</sup> H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang. MOLPRO, version 2015.1, a package of ab initio programs **2015**. See <http://www.molpro.net>.
- <sup>174</sup> S. Mizushima, H. Okazaki, *J. Am. Chem. Soc.* **1949**, *71*, 3411.
- <sup>175</sup> N. Sheppard, G. J. Szasz, *J. Chem. Phys.* **1949**, *17*, 86.
- <sup>176</sup> R. G. Snyder, *J. Chem. Phys.* **1967**, *47*, 1316.
- <sup>177</sup> H. Issei, T. Hideo, S. Masaaki, M. Hiroatsu, S. Takehiko, *Bull. Chem. Soc. Jpn.* **1977**, *50*, 102.
- <sup>178</sup> L. Colombo, G. Zerbi, *J. Chem. Phys.* **1980**, *73*, 2013.
- <sup>179</sup> T. Tynkkynen, T. Hassinen, M. Tiainen, P. Soininen, R. Laatikainen, *Magn. Reson. Chem.* **2012**, *50*, 598.
- <sup>180</sup> R. M. Balabin, *J. Phys. Chem. A* **2009**, *113*, 1012.

- <sup>181</sup> M. Maissara, J. Cornut, J. Devaure, J. Lascombe, *Spectrosc. Int. J* **1983**, *2*, 104.
- <sup>182</sup> I. Kanesaka, R. G. Snyder, H. L. Strauss, *J. Chem. Phys.* **1986**, *84*, 395.
- <sup>183</sup> J. Csontos, B. Nagy, L. Gyevi-Nagy, M. Kállay, G. Tasi, *J. Chem. Theory Comput.* **2016**, *12*, 2679.
- <sup>184</sup> L. H. Hall, B. Mohny, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.
- <sup>185</sup> D. J. Marino, P. J. Peruzzo, G. Krenkel, E. A. Castro, *Chemical Physics Letters* **2003**, *369*, 325 .
- <sup>186</sup> G. Tasi, L. Gyevi-Nagy, R. Tóbiás, T. S. Tasi, *J. Math. Chem.* **2013**, *51*, 2187.
- <sup>187</sup> L. Gyevi-Nagy, G. Tasi, *Comput. Phys. Commun.* **2017**, *215*, 156 .
- <sup>188</sup> L. Gyevi-Nagy, G. Tasi. SYVA **2017**.  
<https://www.sciencedirect.com/science/article/pii/S0010465517300310>,  
utolsó hozzáférés: 2018. 05. 10.
- <sup>189</sup> L. Gyevi-Nagy, G. Tasi. SYVA test set **2017**.  
<https://data.mendeley.com/datasets/xd4g8hyjf2/1>, utolsó hozzáférés: 2018.  
05. 10.
- <sup>190</sup> S. Patchkovskii. symmetry **2003**.  
<https://github.com/nquesada/symmetry>, utolsó hozzáférés: 2018. 05. 10.
- <sup>191</sup> K. L. Bak, J. Gauss, P. Jørgensen, J. Olsen, T. Helgaker, J. F. Stanton, *J. Chem. Phys.* **2001**, *114*, 6548.
- <sup>192</sup> J. Csontos, Z. Rolik, S. Das, M. Kállay, *J. Phys. Chem. A* **2010**, *114*, 13093.
- <sup>193</sup> B. Nagy, J. Csontos, M. Kállay, G. Tasi, *J. Phys. Chem. A* **2010**, *114*, 13213.
- <sup>194</sup> J. A. Darsey, B. K. Rao, *Macromolecules* **1981**, *14*, 1575.
- <sup>195</sup> J. Scarsdale, C. V. Alsenoy, L. Schäfer, *THEOCHEM* **1982**, *86*, 277 .
- <sup>196</sup> L. Schäfer, K. Siam, J. D. Ewbank, E. Osawa, *THEOCHEM* **1986**, *139*, 125–144.
- <sup>197</sup> A. L. Aljibury, R. G. Snyder, H. L. Strauss, K. Raghavachari, *J. Chem. Phys.* **1986**, *84*, 6872.
- <sup>198</sup> K. B. Wiberg, M. A. Murcko, *J. Am. Chem. Soc.* **1988**, *110*, 8029.
- <sup>199</sup> S. Tsuzuki, L. Schafer, H. Goto, E. D. Jemmis, H. Hosoya, K. Siam, K. Tanabe, E. Osawa, *J. Am. Chem. Soc.* **1991**, *113*, 4665.

- <sup>200</sup> N. G. Mirkin, S. Krimm, *J. Phys. Chem.* **1993**, *97*, 13887.
- <sup>201</sup> A. Salam, M. S. Deleuze, *J. Chem. Phys.* **2002**, *116*, 1296.
- <sup>202</sup> J. B. Klauda, B. R. Brooks, A. D. MacKerell, R. M. Venable, R. W. Pastor, *J. Phys. Chem. B* **2005**, *109*, 5300.
- <sup>203</sup> D. Gruzman, A. Karton, J. M. L. Martin, *J. Phys. Chem. A* **2009**, *113*, 11974.
- <sup>204</sup> L. S. Bartell, D. A. Kohl, *J. Chem. Phys.* **1963**, *39*, 3097.
- <sup>205</sup> R. Tóbiás, A. G. Császár, L. Gyevi-Nagy, G. Tasi, *J. Comput. Chem.* **2018**, *39*, 424.
- <sup>206</sup> J. M. L. Martin, *J. Phys. Chem. A* **2013**, *117*, 3118.



# Függelék

---

```
1: procedure OPTIMIZE
  # főténgellyel párhuzamos vektorok optimalizálása
2:   for all  $\mathbf{v}_e \in E$  do                                     ▷ minden szimmetriaelemre
3:     if  $|\mathbf{v}_e \cdot \mathbf{v}_p - 1| < \text{TOL}$  then                   ▷  $\mathbf{v}_e$  merőleges a  $\mathbf{v}_p$  főténgelyre
4:        $\mathbf{v}_e \leftarrow \mathbf{v}_p$                                      ▷  $\mathbf{v}_e$  helyett a főténgely irányvektorát használjuk
5:     end if
6:   end for
  # főténgelyre merőleges vektorok optimalizálása
7:    $\mathbf{v} \leftarrow \arg \min_{\mathbf{v}_e \in E} (\mathbf{v}_e \cdot \mathbf{v}_p)$            ▷  $\mathbf{v}$  a főténgelyre merőleges vektor
8:    $\mathbf{v} \leftarrow (\mathbf{v} - (\mathbf{v} \cdot \mathbf{v}_p)\mathbf{v}_p) / |\mathbf{v} - (\mathbf{v} \cdot \mathbf{v}_p)\mathbf{v}_p|$    ▷  $\mathbf{v}$  ortogonalizálása a főténgelyre
9:   for  $k \leftarrow 1, n_p$  do                                     ▷  $n_p$  a főténgely rendje
10:     $\mathbf{v}_r \leftarrow \text{ROTATE}(\mathbf{O}, \mathbf{v}_p, k \cdot 360^\circ / n_p, \mathbf{v})$ 
11:    for all  $\mathbf{v}_e \in E$  do                                       ▷ minden szimmetriaelemre
12:      if  $\mathbf{v}_e \cdot \mathbf{v}_r < \text{TOL}$  then
13:         $\mathbf{v}_e \leftarrow \mathbf{v}_r$                                      ▷  $\mathbf{v}_e$  helyett az elforgatott vektort használjuk
14:      end if
15:    end for
16:  end for

  # helyvektorok optimalizálása
17:  for all  $i \in A$  do                                             ▷ minden atomra
18:    if  $|\mathbf{p}_i| < \text{TOL}$  then                                       ▷  $i$  atom a szimmetriacentrumban található
19:       $\mathbf{p}_i \leftarrow \mathbf{0}$ 
20:    end if

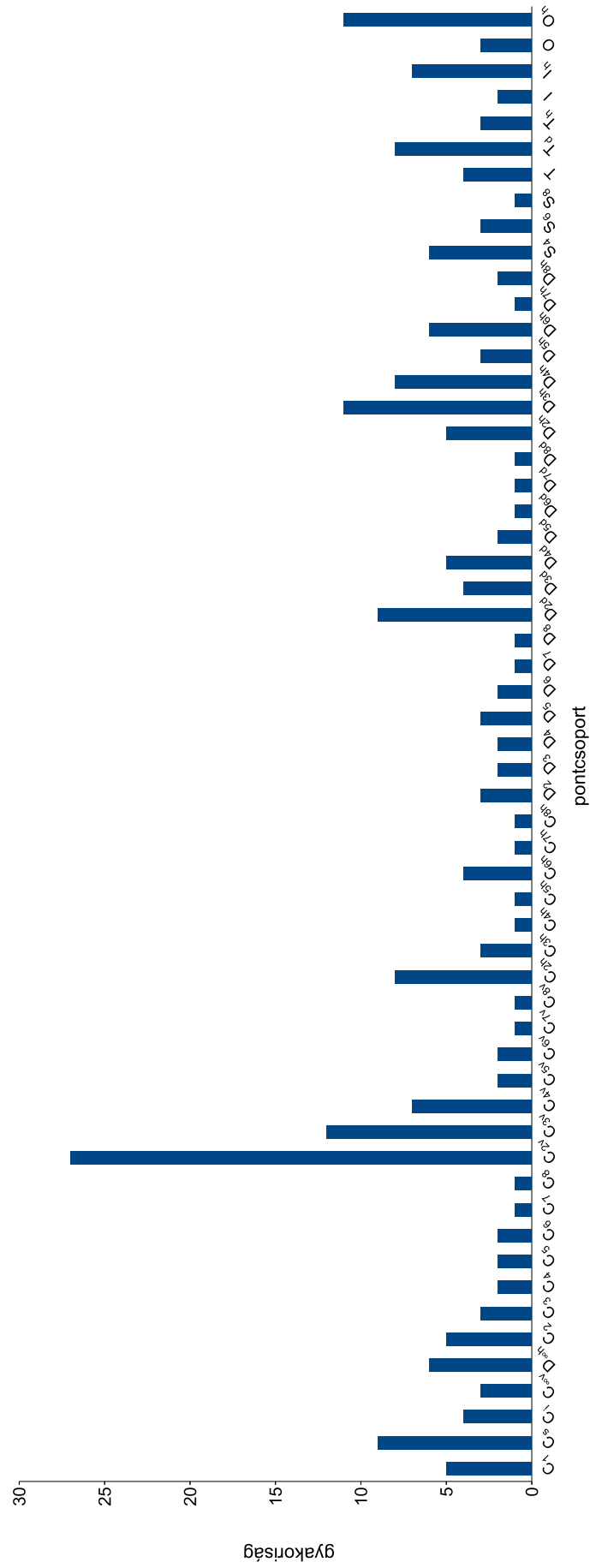
21:    for all  $\mathbf{v} \in C$  do
22:      if  $|\mathbf{p}_i - (\mathbf{p}_i \cdot \mathbf{v})\mathbf{v}| < \text{TOL}$  then                   ▷  $i$  atom egy szimmetriatengelyen található
23:         $\mathbf{p}_i \leftarrow (\mathbf{p}_i \cdot \mathbf{v})\mathbf{v}$ 
24:      end if
25:    end for

26:    for all  $\mathbf{n} \in \sigma$  do
27:      if  $\mathbf{p}_i \cdot \mathbf{n} < \text{TOL}$  then                                       ▷  $i$  atom egy szimmetriasíkon található
28:         $\mathbf{p}_i \leftarrow \mathbf{p}_i - (\mathbf{p}_i \cdot \mathbf{n})\mathbf{n}$ 
29:      end if
30:    end for

31:    for all  $\mathbf{v}_e \in E$  do                                             ▷ minden szimmetriaelemre
32:       $\mathbf{p}_{\text{tr}} \leftarrow \text{TRANSFORM}(\mathbf{v}_e, \mathbf{p}_i)$ 
33:       $\arg \min_{\mathbf{p}_j} (|\mathbf{p}_{\text{tr}} - \mathbf{p}_j|) \leftarrow \mathbf{p}_{\text{tr}}$            ▷ összes szimmetriaművelet végrehajtása a helyvektorokon
34:    end for
35:  end for
36: end procedure
```

---

F1. ábra: Az optimalizált geometria meghatározásának folyamata.



F2. ábra: Az egyes pontszámok gyakorisága a teszhalmazban.

F1. táblázat: Az egyes programok által néhány molekulához rendelt pontcsoport.

Molekula	SYVA	SYMMOL	SYMMETRY	SYMMETRIZER	NWCHEM	ORCA	MRCC
Brom-pentafluorid	C <sub>4v</sub>	C <sub>4v</sub>	C <sub>s</sub>	C <sub>4v</sub>	— <sup>a</sup>	C <sub>4v</sub>	C <sub>2</sub>
Diurea	C <sub>2h</sub>	C <sub>2h</sub>	C <sub>2</sub>	C <sub>2h</sub>	C <sub>2h</sub>	C <sub>2h</sub>	C <sub>2</sub>
1,4-dimetil- biciklo[2.2.2]oktán	D <sub>3h</sub>	C <sub>3h</sub>	D <sub>3h</sub>	D <sub>3h</sub>	C <sub>3h</sub>	D <sub>3h</sub>	D <sub>3h</sub>
Bis(cot)tórium(IV)	D <sub>8h</sub>	D <sub>8h</sub>	D <sub>8h</sub>	D <sub>4h</sub>	— <sup>a</sup>	— <sup>a</sup>	D <sub>8h</sub>
N <sub>4</sub> S <sub>4</sub>	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2</sub>	D <sub>2d</sub>
Tetrabromneopentán	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2d</sub>	D <sub>2d</sub>	S <sub>4</sub>	D <sub>2</sub>	D <sub>2d</sub>

<sup>a</sup> A pontcsoport meghatározása sikertelen.

F2. táblázat: Az  $n$ -pentán konformereinek CCSD(T)/cc-pVTZ szinten meghatározott egyensúlyi geometriái.<sup>a</sup>

tt konformer				tg konformer			
C	0.00000000	0.00000000	0.31811227	C	-2.09104202	0.69588070	0.17463855
C	0.00000000	1.27142882	-0.51767204	C	-1.38991042	-0.58958469	-0.24797422
C	0.00000000	-1.27142882	-0.51767204	C	0.04824330	-0.69207312	0.24923969
C	0.00000000	2.53347023	0.33332435	C	0.98464102	0.36149354	-0.32822994
C	0.00000000	-2.53347023	0.33332435	C	2.42377351	0.16709441	0.12854150
H	0.00000000	3.43159642	-0.28127344	H	-3.13073928	0.69937225	-0.14683103
H	-0.87960961	2.56578262	0.97569508	H	-1.61033200	1.57364271	-0.25264536
H	0.87960961	2.56578262	0.97569508	H	-2.07487196	0.80216026	1.25956055
H	0.87397475	1.26920245	-1.17113566	H	-1.95580012	-1.44255196	0.12711804
H	-0.87397475	1.26920245	-1.17113566	H	-1.39709943	-0.66767052	-1.33695542
H	0.87451375	0.00000000	0.97329685	H	0.05739264	-0.61929506	1.33993501
H	-0.87451375	0.00000000	0.97329685	H	0.44320748	-1.67941206	0.00265613
H	0.00000000	-3.43159642	-0.28127344	H	0.93594131	0.32006042	-1.41815807
H	0.87960961	-2.56578262	0.97569508	H	0.64655085	1.35637348	-0.03990897
H	-0.87960961	-2.56578262	0.97569508	H	3.08303194	0.92616092	-0.28817726
H	-0.87397475	-1.26920245	-1.17113566	H	2.79826562	-0.80914965	-0.17812521
H	0.87397475	-1.26920245	-1.17113566	H	2.49372481	0.22125070	1.21472837

gg konformer				gx konformer			
C	0.00000000	0.00000000	1.05412148	C	-1.56933971	1.02565183	-0.02765009
C	-0.62298163	-1.12542003	0.23122977	C	-1.39354277	-0.48633824	-0.09089432
C	0.62298163	1.12542003	0.23122977	C	-0.02810030	-0.99158287	0.36965391
C	0.36369240	-1.80458233	-0.71027902	C	1.15891900	-0.45151591	-0.43822050
C	-0.36369240	1.80458233	-0.71027902	C	1.81499540	0.78105149	0.17560958
H	-0.09621050	-2.65101235	-1.21657443	H	-2.58321509	1.30901872	-0.30423197
H	0.09621050	2.65101235	-1.21657443	H	-0.88854240	1.53453202	-0.70741116
H	1.23063569	-2.17189054	-0.16071830	H	-1.37851626	1.39844919	0.97868773
H	-1.23063569	2.17189054	-0.16071830	H	-2.16433202	-0.95962352	0.51807869
H	0.72327768	-1.11766425	-1.47467900	H	-1.56100391	-0.82210114	-1.11619018
H	-0.72327768	1.11766425	-1.47467900	H	0.11353922	-0.74093525	1.42412128
H	-1.03486386	-1.86694356	0.91562941	H	-0.03887448	-2.08003191	0.31273794
H	1.03486386	1.86694356	0.91562941	H	1.91606491	-1.22949123	-0.53081872
H	-1.46684958	-0.74033586	-0.34230555	H	0.82795725	-0.22759620	-1.45480529
H	1.46684958	0.74033586	-0.34230555	H	2.64221063	1.13438527	-0.43798793
H	0.76485410	-0.42534195	1.70698323	H	2.20939642	0.54523701	1.16367309
H	-0.76485410	0.42534195	1.70698323	H	1.10854607	1.59952622	0.29109182

<sup>a</sup> A koordináták Å egységekben szerepelnek.

F3. táblázat: A legjelentősebb hozzájárulások konvergenciája a bázis méretének növelésével az  $n$ -pentán konformereinek relatív teljes energiájához.<sup>a</sup>

$X^b$	$\Delta E(\text{tg})$					$\Delta E(\text{gg})$					$\Delta E(\text{gx})$				
	$\Delta E_{\text{HF}}$	$\delta E_{\text{MP2}}$	$\delta E_{\text{CCSD}}$	$\delta E_{\text{(T)}}$	$\sum^c$	$\Delta E_{\text{HF}}$	$\delta E_{\text{MP2}}$	$\delta E_{\text{CCSD}}$	$\delta E_{\text{(T)}}$	$\sum^c$	$\Delta E_{\text{HF}}$	$\delta E_{\text{MP2}}$	$\delta E_{\text{CCSD}}$	$\delta E_{\text{(T)}}$	$\sum^c$
2	1170	-567	98	-58	644	2340	-1381	285	-128	1117	4318	-1263	147	-134	3069
3	1171	-618	106	-76	584	2338	-1600	301	-176	862	4257	-1419	174	-193	2819
4	1182	-614	109	-77	600	2358	-1565	318	-183	928	4276	-1450	186	-200	2811
5	1182	-614	113	-79	603	2361	-1565	324	-186	934	4278	-1444	192	-202	2824
6	1183	-615			603	2363	-1571			929	4279	-1456			2813
(3, 4) <sup>d</sup>		-612	112	-79	604		-1540	330	-188	961		-1473	194	-205	2793
(4, 5) <sup>d</sup>		-614	117	-80	606		-1565	330	-190	938		-1438	199	-204	2834
(5, 6) <sup>d</sup>		-615			606		-1580			924		-1473			2801
$\sigma^e$	1	1	5	1	8	2	15	0	2	19	1	35	5	1	42

<sup>a</sup> Minden érték cal mol<sup>-1</sup> egységekben szerepel.

<sup>b</sup> A cc-pVXZ bázis kardinális száma.

<sup>c</sup> Az egyes hozzájárulások összege. Hiányzó értékeknl a legnagyobb bázison számított érték került felhasználásra.

<sup>d</sup> A cc-pV(X, X+1)Z extrapolációval kapott érték.

<sup>e</sup> A hozzájárulás hibája mint a két legpontosabb érték abszolút különbsége.

F4. táblázat: Az *n*-pentán konformereinek CCSD(T)/cc-pVTZ szinten meghatározott rezgési frekvenciái.<sup>a</sup>

tt konformer					tg konformer				
<b>106</b>	869	1276	1504	3044	<b>77</b>	860	1279	1507	3048
<b>113</b>	886	1295	1514	3048	<b>124</b>	876	1304	1512	3049
174	945	1337	1519	3048	216	934	1340	1518	3051
<b>254</b>	998	1341	1520	3052	<b>254</b>	1007	1348	1521	3062
<b>262</b>	1055	1379	1523	3068	<b>291</b>	1036	1389	1526	3072
393	1072	1420	1529	3101	329	1064	1413	1529	3102
402	1103	1422	3025	3103	485	1115	1423	3022	3104
738	1179	1425	3031	3109	736	1170	1426	3034	3109
767	1218	1499	3038	3110	771	1205	1497	3039	3114

gg konformer					gx konformer				
<b>86</b>	851	1279	1510	3049	<b>101</b>	837	1273	1506	3049
<b>157</b>	876	1294	1510	3052	<b>137</b>	884	1304	1509	3052
<b>244</b>	920	1342	1520	3057	<b>226</b>	919	1331	1519	3057
268	1017	1373	1522	3066	<b>277</b>	1026	1375	1522	3062
<b>302</b>	1031	1400	1525	3070	304	1040	1393	1527	3071
377	1035	1402	1536	3102	363	1050	1397	1533	3103
485	1124	1424	3028	3104	454	1130	1424	3022	3103
728	1168	1426	3039	3110	751	1157	1426	3034	3110
786	1187	1496	3048	3111	775	1194	1500	3035	3125

<sup>a</sup> Minden érték  $\text{cm}^{-1}$  egységekben szerepel. A félkövérrel szedett értékek kezelése gátolt belső rotációként történt

F5. táblázat: A metil-torziókra vonatkozó Schrödinger-egyenlet sajátértékei.<sup>a</sup>

tt	gg	tg(t-Me)	tg(g-Me)	gx(g-Me)	gx(x-Me)
109.47	107.28	109.91	104.19	110.37	102.18
109.47	107.28	109.91	104.26	110.40	102.31
109.49	107.38	109.91	104.26	110.40	102.31
323.48	316.26	324.88	308.63	316.55	300.40

Folytatás a következő oldalon.

Folytatás az előző oldalról.

323.48	316.26	324.88	308.63	316.55	300.40
323.48	316.27	324.89	308.63	316.56	300.41
526.22	512.61	528.90	502.87	494.22	483.89
526.27	512.64	528.94	502.93	494.46	484.06
526.28	512.67	528.94	502.94	494.47	484.09
714.94	695.28	718.86	681.82	645.73	646.09
714.95	695.29	718.86	681.83	645.73	646.11
715.76	695.91	719.57	682.75	648.91	648.68
881.00	856.68	886.80	836.29	762.53	770.60
887.85	862.10	892.93	844.20	780.46	787.63
887.85	862.10	892.93	844.20	780.46	787.64
1018.62	992.41	1025.90	959.56	863.44	869.43
1018.63	992.41	1025.90	959.56	863.44	869.43
1055.02	1022.04	1059.87	1001.12	929.08	936.89
1113.15	1087.12	1120.25	1042.45	943.89	946.30
1179.63	1145.38	1184.25	1112.05	1031.03	1035.63
1179.63	1145.38	1184.25	1112.05	1031.03	1035.64
1288.34	1247.75	1291.05	1216.37	1140.00	1144.69
1288.34	1247.75	1291.05	1216.37	1140.00	1144.69
1411.65	1362.10	1412.56	1335.99	1262.12	1266.66
1414.27	1366.91	1415.02	1336.01	1262.45	1267.67
1552.52	1496.07	1551.50	1469.26	1397.04	1402.26
1552.52	1496.08	1551.50	1469.26	1397.04	1402.26
1705.68	1640.74	1702.76	1615.26	1543.83	1549.49
1705.68	1640.74	1702.76	1615.26	1543.83	1549.49
1871.83	1797.78	1866.94	1773.49	1702.38	1708.60
1871.87	1797.93	1866.96	1773.52	1702.40	1708.60
2050.66	1967.05	2043.71	1943.71	1872.55	1879.39
2050.66	1967.05	2043.71	1943.71	1872.55	1879.39
	2148.07	2125.67	2054.19	2061.74	
	2148.07	2125.67	2054.19	2061.74	

<sup>a</sup> Minden érték  $\text{cm}^{-1}$  egységekben szerepel.

F6. táblázat: A C–C–C–C-torziókra vonatkozó Schrödinger-egyenlet sajátértékei.<sup>a</sup>

tt	gg	tg (t-torzió)	tg (g-torzió)	gx(g-torzió)	gx(x-torzió)
54.86	54.71	45.04	54.20	62.69	42.22
163.75	165.12	134.96	162.54	194.57	131.51
270.66	272.60	222.67	271.47	331.00	180.98
375.23	378.33	308.73	380.03	479.71	310.73
477.16	482.78	393.30	487.17	636.93	319.12
576.12	585.83	476.16	591.92	798.06	578.92
671.78	687.01	556.91	693.22	959.23	582.71
763.71	785.71	635.07	789.74	1118.13	955.75
851.32	881.14	710.17	879.31	1267.71	958.39
934.10	972.28	781.79	959.91	1423.50	1440.81
1008.89	1057.54	849.27	1036.34	1525.91	1442.86
1084.05	1134.90	912.94	1118.81	1731.65	2033.86
1121.44	1207.16	965.85	1206.84	1784.62	2035.53
1229.06	1284.12	1030.16	1302.14	2095.01	
1232.71	1369.32	1050.53	1392.13	2134.29	
1398.58	1461.06	1154.20	1500.50		
1398.72	1557.85	1159.06	1574.08		
1599.34	1658.75	1301.37	1713.72		
1599.37	1763.01	1303.63	1759.35		
1829.01	1869.97	1471.61	1949.23		
1829.02	1978.98	1473.19	1977.94		
2086.01	2089.37	1663.37	2212.41		
2086.02	2200.32	1664.60	2233.62		
	1970.03	1875.68			
	2151.14	1876.70			
	2167.30	2108.00			
	2366.36	2108.88			

<sup>a</sup> Minden érték  $\text{cm}^{-1}$  egységekben szerepel.



F7. táblázat:  $n$ -pentánkonformerek teljes energia- és entalpiakülönbségei elméleti munkák alapján.<sup>a</sup>

csoporth	$\Delta E(\text{tg})$	$\Delta E(\text{gg})$	$\Delta E(\text{gx})$	$\Delta H_0^\circ(\text{tg})$	$\Delta H_0^\circ(\text{gg})$	$\Delta H_0^\circ(\text{gx})$	$\Delta H_{298}^\circ(\text{tg})$	$\Delta H_{298}^\circ(\text{gg})$	$\Delta H_{298}^\circ(\text{gx})$	hivatkozás
	760	1360	3330							198 <sup>c</sup>
A	740	1302	3289	837	1541	3496	800	1431	3424	199 <sup>d</sup>
	670	1090	3190							200 <sup>e</sup>
B	621	1065	2917	676	1423	3111	871	1771	3476	201 <sup>f</sup>
	618	985	2846							202 <sup>g</sup>
	614	961	2813							203 <sup>h</sup>
C	581	912	2763							206 <sup>i</sup>
	582	915	2767	682	1205	3037	640	1052	2911	206 <sup>j</sup>
saját munka	605±5	917±15	2796±35	707±73	1235±67	3038±92	666±74	1070±71	2906±95	183

<sup>a</sup> Minden érték cal mol<sup>-1</sup> egységekben szerepel.

<sup>a</sup> Csoportosítás az elméleti szint alapján; A: perturbációszámítás, B: CCSD(T) kétszeres- $\zeta$  minőségű bázison, C: CCSD(T) háromszoros- $\zeta$  minőségű bázison

<sup>b</sup> MP3/6-31G(d)//HF/6-31G(d)

<sup>c</sup> MP4(SDQ)/6-31G(d)//HF/6-31G(d)

<sup>d</sup> MP2/6-31G(d)

<sup>e</sup> fókuszpont analízis

<sup>f</sup> MP2:CC

<sup>g</sup> W1h-hez hasonló protokoll

<sup>h</sup> CCSD(T)/cc-pVTZ

<sup>i</sup> CCSD(T)-F12b/cc-pVTZ-F12//SCS-MP2/cc-pVTZ

F8. táblázat: A tt–tg entalpiakülönbség meghatározásához felhasznált adatok.<sup>a</sup>

$\frac{1}{T}/\frac{1}{K}$	$\ln\left(\frac{I_{tg}}{I_{tt}}\right)$	$\frac{I_{tg}}{I_{tt}}$	$\sigma_{\ln(I_{tg}/I_{tt})}^2$ <sup>a</sup>
0.00400	−1.459	0.23236	0.43035
0.00415	−1.509	0.22106	0.45234
0.00430	−1.549	0.21231	0.47098
0.00445	−1.599	0.20199	0.49506
0.00460	−1.650	0.19203	0.52074
0.00475	−1.699	0.18269	0.54736
0.00490	−1.739	0.17552	0.56971
0.00505	−1.790	0.16693	0.59905
0.00520	−1.840	0.15881	0.62966
0.00535	−1.880	0.15258	0.65538
0.00550	−1.929	0.14516	0.68888
0.00565	−1.980	0.13800	0.72461
0.00580	−2.009	0.13399	0.74629
0.00595	−2.070	0.12609	0.79304
0.00610	−2.110	0.12119	0.82512
0.00625	−2.160	0.11525	0.86761
0.00640	−2.210	0.10961	0.91229
0.00655	−2.240	0.10639	0.93992
0.00670	−2.310	0.09921	1.00793
0.00685	−2.349	0.09542	1.04795
0.00700	−2.410	0.08979	1.11360

<sup>a</sup> Az 5.11 egyenlet alapján meghatározva

F9. táblázat: A tt–gg entalpiakülönbség meghatározásához felhasznált adatok.<sup>a</sup>

$\frac{1}{T}/\frac{1}{K}$	$\ln\left(\frac{I_{tg}}{I_{tt}}\right)$	$\frac{I_{tg}}{I_{tt}}$	$\sigma_{\ln(I_{tg}/I_{tt})}^2$ <sup>a</sup>
0.00700	−3.979	0.01870	1.78288
0.00685	−3.926	0.01972	1.69026
0.00670	−3.934	0.01957	1.70367
0.00655	−3.814	0.02206	1.51070
0.00640	−3.794	0.02250	1.48125
0.00625	−3.656	0.02583	1.29035
0.00610	−3.640	0.02626	1.26934
0.00595	−3.534	0.02917	1.14254
0.00580	−3.477	0.03089	1.07914
0.00565	−3.405	0.03321	1.00365
0.00550	−3.328	0.03588	0.92909
0.00535	−3.261	0.03836	0.86896
0.00520	−3.189	0.04123	0.80855
0.00505	−3.106	0.04479	0.74430
0.00490	−3.024	0.04863	0.68547
0.00475	−2.974	0.05110	0.65229
0.00460	−2.882	0.05601	0.59514
0.00445	−2.821	0.05955	0.55976
0.00430	−2.751	0.06385	0.52206
0.00415	−2.674	0.06897	0.48328
0.00400	−2.601	0.07423	0.44905

<sup>a</sup> Az 5.11 egyenlet alapján meghatározva

F10. táblázat: Az  $n$ -pentánkonformerek interkonverziós átmeneti állapotainak relatív entalpiái különböző hőmérsékleteken.

T / K	$\Delta H_T^\circ(\gamma^\pm\gamma^\mp)$	$\Delta H_T^\circ(t\tau)$	$\Delta H_T^\circ(x^\pm\tau)$	$\Delta H_T^\circ(t\xi^\pm)$	$\Delta H_T^\circ(g^\pm\xi^\pm)$	$\Delta H_T^\circ(g^\pm\xi^\mp)$
cal mol <sup>-1</sup>						
50	3300.4	5513.6	7321.4	3098.0	3610.7	3485.4
100	3205.4	5423.9	7227.5	3011.3	3505.1	3370.9
150	3088.7	5316.2	7111.2	2905.5	3382.2	3241.5
200	2973.5	5206.6	6994.5	2798.6	3262.1	3117.8
250	2863.1	5098.7	6882.1	2692.9	3146.8	3000.7
298.15	2760.5	4996.6	6777.7	2592.3	3039.6	2892.7
350	2652.7	4888.1	6668.1	2484.9	2927.6	2780.2

F11. táblázat: A többváltozós nemlineáris illesztéssel kapott modellek jellemzői.

Deszkriptorok	$s_{\gamma^2}^a$	PRESS <sup>b</sup>	$s_{\gamma^2}^c$
	cal <sup>2</sup> mol <sup>-2</sup>		
$GAI_m, F_{i_m}, T, GAI_m^2, GAI_m \cdot F_{i_m}, GAI_m \cdot T, F_{i_m}^2$	$1.10 \cdot 10^4$	$1.30 \cdot 10^4$	$2.40 \cdot 10^4$
$GAI_m, F_{i_m}, T, GAI_m^2, GAI_m \cdot F_{i_m}, F_{i_m}^2$	$1.20 \cdot 10^4$	$1.40 \cdot 10^4$	$2.60 \cdot 10^4$
$GAI_m, F_{i_m}, GAI_m^2, GAI_m \cdot F_{i_m}, F_{i_m}^2, F_{i_m} \cdot T$	$1.30 \cdot 10^4$	$1.60 \cdot 10^4$	$2.90 \cdot 10^4$
$NK_m, SCI_m, T, NK_m^2, NK_m \cdot SCI_m, SCI_m^2$	$1.40 \cdot 10^4$	$1.60 \cdot 10^4$	$3.00 \cdot 10^4$
$GAI_m, F_{i_m}, GAI_m^2, GAI_m \cdot F_{i_m}, GAI_m \cdot T, F_{i_m}^2$	$1.30 \cdot 10^4$	$1.60 \cdot 10^4$	$3.00 \cdot 10^4$
$GAI_m, M_{2,m}, T, GAI_m^2, GAI_m \cdot M_{2,m}, GAI_m \cdot T, M_{2,m}^2$	$1.40 \cdot 10^4$	$1.60 \cdot 10^4$	$3.00 \cdot 10^4$

<sup>a</sup> Reziduális variancia.

<sup>b</sup> PRESS (predicted residual error sum of squares) értékek.

<sup>c</sup> Reziduális szórás.

F12. táblázat: A deszkriptorok értékei az  $n$ -bután konformereire.

$SI$	$\epsilon$	$\Omega_A$	$\Omega_S$	$\sigma_{SYM}$	$I_{SYM}$	$W_m$	$J_m$	$Xu_m$	$GAl_m$	$F\dot{v}_m$
0.006271	0.9510	0.2179	0.4295	9.0	39.30	64.17	0.01558	2.556	1.0	8.670

${}^1\chi_m$	$F^i x_m$	$HI_m$	$MTI_m$	$\overline{MTI}_m$	$ISI_m$	$M_{1,m}$	$M_{2,m}$	$NK_m$	$ABC_m$	$SCI_m$
0.01558	8.686	0.01558	256.7	264238.5	264238.5	8235.6	64.17	4117.8	0.1752	0.007790

F13. táblázat: Az  $n$ -bután konformereinek relatív entalpiáinak hőmérsékletfüggése.

$T / K$	50	100	150	200	250	298.15	350
$\Delta H_T^\circ(g) / \text{cal mol}^{-1}$	684.4	682.6	671.4	661.2	652.7	647.5	644.0