

Hasonlósági keresés molekulagráfokon

Fekete István, Kovács Péter

ELTE, ChemAxon

{fekete,kpeter}@inf.elte.hu

Budapest, 2018.11.06.

Kémiai informatika ?



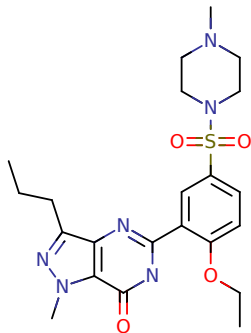
Nem a kémiai aspektussal
foglalkozunk...

Bevezetés



vegyületek/molekulák

kémiai problémák



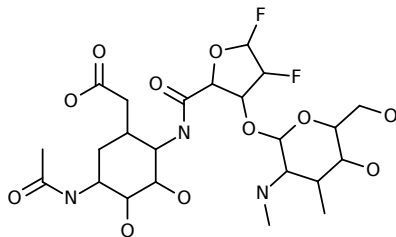
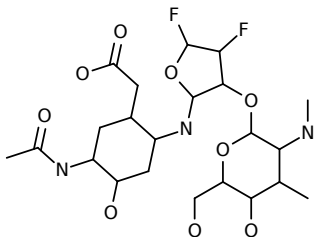
csúcs- és élcímkézett
irányítatlan gráfok



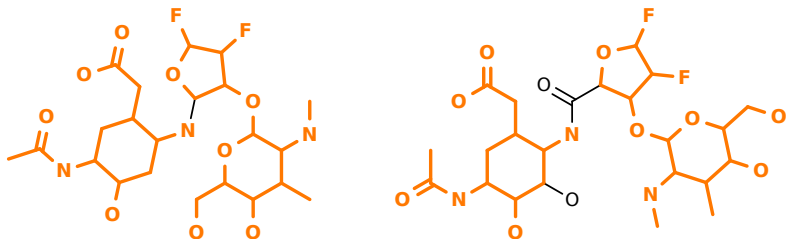
matematikai modellek
és problémák



Mennyire hasonló két molekula?



Mennyire hasonló két molekula?

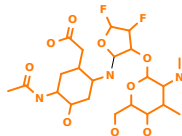


Tekintsük a legnagyobb közös részgráfjukat!

Legnagyobb közös részgráf meghatározása

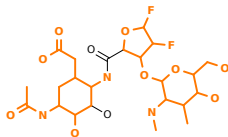
Előnyök

- Ismert matematikai probléma
- Intuitív metrikák definiálhatók vele
- Jól vizualizálható



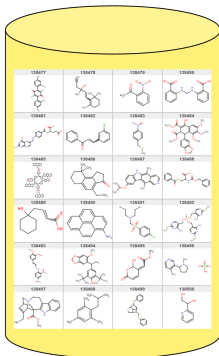
Hátrányok

- Számításigényes, lassú
- Nagy tárigény (sok molekula esetén)

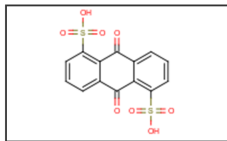


Nagy adatbázis esetén gyors keresés ?

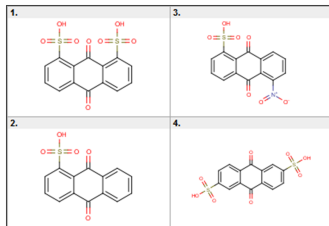
Molekula-adatbázis



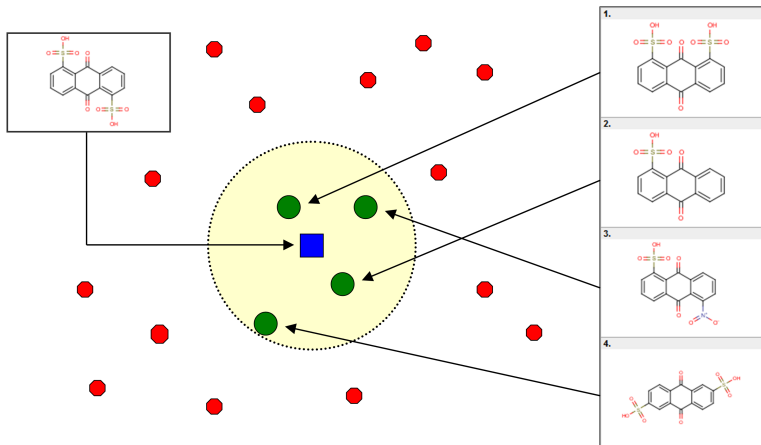
Lekérdezés



Eredmény

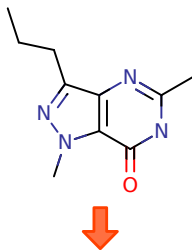


Molekulák leképezése egy metrikus térbe



Molekulák fingerprintjei

- Egy molekula reprezentálása egy 0-1 sorozattal (ún. fingerprinttel)
- Az 1-es bitek valamilyen tulajdonság vagy részstruktúra meglétét jelentik



01000101001100000101110001001101000110100...

Keresés molekulák fingerprintjein

- Nagy adatbázisok esetén
- Gyakran csak előszűréshez
- Nagyon hatékony: kis tárigény, gyors összehasonlítás

Legnagyobb közös részgráf keresése

- Lassabb, de pontosabb elemzés
- Hasonlóság vizualizálása
- Molekulák egymásra illesztése is lehetséges

Keresés molekulák fingerprintjein

- Nagy adatbázisok esetén
- Gyakran csak előszűréshez
- Nagyon hatékony: kis tárigény, gyors összehasonlítás

Néha még tovább akarjuk gyorsítani...

Az előadás első fele

Fekete István

Legnagyobb közös részgráf keresése

- Lassabb, de pontosabb elemzés
- Hasonlóság vizualizálása
- Molekulák egymásra illesztése is lehetséges

Nehéz probléma, hatékony algoritmust keresünk

Az előadás második fele

Kovács Péter