

Podobnost je často vyjadřována jako obrácená míra vzdálenosti v popisovaném prostoru. Míry vzdálenosti jsou klasifikovány v euklidovských nebo neeuklidovských hodnotách, podle toho, platí-li trojúhelníková nerovnost.

Mezey považuje podobnost za fuzzy ekvivalenci.

PODOBNOST

PODOBNOST

Podobnost systémů jako princip

- shlukování
- podobnost v myšlenkovém prostoru
- sdružování
- chemická izomerie
- fyzikální mezomerie
- definice izomerie
- molekulární podobnost a teorie informací
- koncepce molekulární podobnosti

Analýza chemického prostoru

- molekulární deskriptory
- hledání maximální společné substruktury
- strukturní genetické otisky prstů
- virtuální prosívání
- kvantifikace podobnosti, indexy podobnosti
- princip molekulární podobnosti
- mapování chemického prostoru
- definice míry kvantové podobnosti
- princip hledání podobnosti
- teorie důvěryhodnosti
- fuzzy logika, strukturní podobnost a shlukování
- fuzzy podobnost
- geometrické myšlení a topologické prostory
- shluková analýza
- podobnost a chemotopologické studie
- kvantová atomová podobnost

Kvantově molekulární podobnost

- vztahy mezi strukturou a vlastnostmi látek
- QSRR, QSAR, QShAR, ekvivalent atomu TAE
- molekulární kvantová podobnost MQS, QSPR, QSAR
- Laplacian elektronové hustoty jako deskriptor reaktivity molekuly
- Tanimotův koeficient (index)

Podobnost v BCP prostoru

- vzdálenost $d(A,B)$ a suma BCP vzdáleností
- BCP v Baderově teorii atomu v molekule, AIM

Molekulární kvantová podobnost a deskriptory DFT

- molekulární kvantová podobnost MQSM
- Carbó-ův index podobnosti
- mapy a topologie Laplacianu

Fuzzy podobnost

- míra podobnosti režimu fuzzy množin a fuzzy logiky

MQSM fragmentů

- míra kvantové podobnosti fragmentů
- předpovídání bioaktivity sloučenin

EXTENZE POJMU PODOBNOST V POJMECH **izo-** SUBSYSTÉMŮ

PODOBNOST

Mendělejev: při uspořádání prvků v pořadí rostoucích atomových hmotností se prvky s podobnými vlastnostmi periodicky opakují.

Podobnost systémů na různých úrovních organizace hmotných objektů je principem obsaženým v teorii nekonečného hierarchického seskupování materie; ta popisuje, které souvislosti a jaká propojení mezi různými soubory materie lze nacházet. Někteří autoři hovoří o zákonu podobnosti „nosníků“ odstupňování a měřítek různých úrovní podobných systémů. Podobnost různých věcí je konformní se symetrií $SP\phi$ a může být ilustrována vzdáleností nebeských těles, kvantováním parametrů kosmických systémů, a na druhém konci existencí systémů založených na vodíku. Vztahy podobnosti nám umožňují dospívat k parametrům pro objekty, které jsou nedostižitelné přímým pozorováním (tedy nejhlubších strukturních elementárních částic a rovněž objektů o větší velikosti než jsou metagalaxie) a to jejich hmoty, velikosti, spinu, elektrického náboje, magnetického momentu, energie, pohybu, teploty atp. včetně hodnot základních fyzikálních konstant objektů na jednotlivých úrovních.

Shlukování a prolínání různých úrovní systémů předpokládá, že „větší“ systémy se vytvářejí skládáním částic z nižších úrovní a že tedy existují také vzájemné souvislosti charakteristik objektů a stavů komponent a je dána i symetrie mezi vlastnostmi látkových částic a vlastnostmi objektů, projevující se vztahy podobnosti. Pokud naše klasifikace sahá až ke kosmickým objektům, nabízí se myšlenka dimenzí jednotlivých škál zahrnující pátý rozměr časoprostoru.

Shepard (1962) soudí, že myšlenkové představy názornosti je možno koncipovat jako jistý druh myšlenkového prostoru. Koncepty jsou reprezentovány body v tomto prostoru a podobnost mezi nimi je funkcí vzdálenosti mezi jednotlivými koncepty. Je pak úkolem matematiky „měřit“ tuto vzdálenost.

Podobnost je možno považovat za jedno z pravidel sdružování, která spadají do běžné lidské činnosti při uspořádání jevů, událostí a fakt ve světě obklopujícím člověka. Proto je podobnost považována (např. Blough, 2001 ad.) za jeden z ústředních problémů psychologie. Pro moderní kognitivní proces je základem rozpoznání a srovnávání objektů a jejich kategorizace.

Zamyšlení nad úzkou strukturní podobností mezi chemickými systémy jakými jsou izomery, analogy, homology apod., všem chemikům důvěrně známou pod širokým pojmem podobnost, je stále čím doplňovat.

Zopakujme si, že v tom nejširším smyslu slova se o chemické podobnosti pojednává při porovnávání dvou nebo několika chemických entit majících stejné zastoupení atomů prvků. Že se rozlišují různé míry a druhy izomerií, v první řadě na chemický a fyzikální izomerismus. **Chemickou izomerii** pak hledáme v polymerii (sloučenin o stejném procentovém zastoupení, ale lišících se molekulovou hmotností - třeba C_2H_4 (28) a buten C_4H_8 (56)); polymerismus je podle příslušnosti sloučenin do stejné nebo různé třídy příležitostný nebo generický; o metamerii se hovoří u sloučenin se stejným procentovým složením a molekulovou hmotností, které zahrnují homologické radikály ($C_3H_7NH_2$, $C_2H_5NHCH_3$ a $N(CH_3)_3$ apod.); řetězová izomerie se týká molekul o stejném složení ale různém uspořádání ($CH_3CH_2CH_2CH_3$ a $(CH_3)_3CH$, ..); polohová izomerie je další forma jevu (příklad $CH_3CH_2CH_2Cl$ a $CH_3CHClCH_3$ apod.); strukturně izomerní jsou sloučeniny které se při stejném procentuálním složení a molekulovou hmotností liší uspořádáním atomů v jejich molekulách.

Z hlediska charakterizace **fyzikální mezomerie** se posuzují proměnlivé agregáty chemicky podobných molekul. Jde proto především o látky řazené ke geometrickým, stereoizomerním nebo opticky izomerním systémům. U geometricky izomerních molekul se uvažuje o allostereometrii, pokud látky mají dokonale podobné spojení a prostorové uspořádání atomů a liší se pouze obsahem energie (pohybu atomů), je namístě označení dynamický izomerismus. Také si vzpomínáte na geometrickou stereoizomerii sloučenin majících stejnou konstituci a odlišují se prostorovým uspořádáním některých skupin nebo atomů. Opticky izomerní jsou stereoizomerní látky podle otáčivosti polarizovaného světla (případně jsou opticky inaktivní (racemáty nebo mezofomy).

Definice izomerie

V chemii: Izomerem nazveme jednu ze dvou nebo několika látek, sestávajících z týchž prvků ve stejných poměrech, ale lišících se vlastnostmi; tyto rozdíly jsou způsobeny uspořádáním atomů ve struktuře molekul.

Ve fyzice: Jedno ze dvou nebo několika jader se stejným hmotnostním číslem a s atomovým číslem, které mají rozdílné radioaktivní vlastnosti a mohou existovat v jednom z mnoha energetických stavů v měřitelné časové periodě. (*Encyklopaedia.*)

Chemickou podobnost definujeme vždy z hlediska zaměření analýzy daných systémů.

Chemická podobnost je zaměřena na chemické prvky, jejich atomy, ionty, substrukтуры, fragmenty, molekuly (chemické sloučeniny) a, jak označení napovídá, na nacházení a srovnávání jejich strukturních a funkčních vlastností – a přitom především jde o předpoklady vztahu systému k možnému partnerskému reagentu; biologické projevy se obvykle označují a jsou kvantifikovány jako biologická aktivita látky a ta má samozřejmě základ v chemické aktivitě.

Chemická podobnost patří k nejvýznamnějším koncepcím chemoinformatiky. A její význam podtrhuje skutečnost, že v soudobých teoretických přístupech získala postavení jednoho z postupů predikcí vlastností chemických sloučenin, vyhledávání systémů s předem definovanými vlastnostmi a k budování obsáhlých databází struktur chemických komponent pro syntézy farmak. (Zpravidla jsou citováni Johnson a Maggiora: *Podobné sloučeniny mají podobné vlastnosti.*)

Relace podobnosti je tak široká a může zahrnovat podobnost v tolika různých aspektech, že lze těžko podat jejich výčet. Ponechme pro naše účely stranou podobnosti v matematice a ty anorganického světa (byť třeba idea krychle v kamenné soli překvapí a renesanční autoři věnovali mnoho místa obrazům měst a ruin v mramorech a brekciích či květin v krystalech antimonu či ledu na okně).

Myšlenka přistupovat k řešení molekulární podobnosti prostředky teorie informací a teorie atomu molekule (AIM), sdělená Nalewajskim a Parrem (*Proc.Natl.Sci. USA* 2000, 97, 8879), zaujala řadu dalších teoretiků. Teoretickým nástrojem je princip minimálního úbytku entropie (Kullback, Liebler, *Ann.Math.Stat.* 1951, 22, 79) aplikovaný na problém podobnosti mezi molekulami, případně jejich fragmenty až atomy. Chemikovo porozumění molekulám jako kombinacím atomů začíná definováním atomu parcelací molekuly na části jako nositele podstatných znaků a vlastností („stockholders partitioning“: Hirschfeld, *Theor.Chim.Acta* 1977, 44, 129); ostatně teorie AIM (Bader R.F.W., *Atoms in Molecules – A Quantum Theory*. Univ. of Oxford Press, Oxford 1990) se stala jednou z nejpropracovanějších a nejužívanějších kvantověchemických metod. Hodnoty elektronových hustot izolovaných atomů představují indikátory chemických vlastností atomů a součet hustot izolovaných atomů nebo iontů v aktuálních polohách jader definují rozložení elektronových hustot v promolekule. Promolekula je klíčovou položkou v analýze rozdílů hustot v chemické vazbě (citovaný Hirschfeld). Od časů Paulinga a Mullikena jsou základní stavy atomu s případnými malými poruchami referenčními stavy detailního popisu AIM obsahujícího koncepcí

hybridizace, promoční energie, polarizace, iontového charakteru vazby a všech informací řešených teorií funkcionálu elektronové hustoty, DFT.

Koncepce molekulární podobnosti obsahuje charakterizování strukturní podobnosti různých molekul či jejich fragmentů pomocí míry překryvových integrálů elektronové hustoty, elektrostatického potenciálu a Fukuiho funkce včetně tvrdosti systémů (HSAB). Chemici vycházejí z premisy, že z podobnosti elektronových struktur vyplývá souvislost s obsahem informace rozložení elektronových hustot. To vyžaduje získat míru informační vzdálenosti, jak ji předpokládá funkcionál úbytku entropie (informační vzdálenosti)

$$\Delta S[P/P_0] = \int P(\vec{r}) \ln \left[\frac{P(\vec{r})}{P_0(\vec{r})} \right] d\vec{r}$$

Pokud se počítá reaktivita specifického reakčního centra, hustoty odpovídajících molekulárních fragmentů jsou vyhledávány jako údaje vstupní informace. Teorie informací umožňuje vyvozování charakteru vazby mezi atomy z funkcionálu úbytku entropie. Základní myšlenka je aplikována také na excitované elektronové stavy.

V pojednání o nových filozofických perspektivách v chemii uvedl ve *Science* (2000) Nalini Bhushan vztahy podobnosti jako prostor pro uspořádání mnoha ostatních vlastností počínaje atomovými čísly, atomovou vahou a řady dalších až po periodicitu.

ANALÝZA CHEMICKÉHO PROSTORU

Hledání podobnosti je praktickým úkonem identifikace kandidátů (molekul, fragmentů iontů, radikálů...) srovnáváním dvojic sloučenin. Prohledává se databáze standardními postupy. Struktury se srovnávají na základě podobnosti hodnot vypočtených z jejich molekulárních deskriptorů. Nejširší uplatnění má dnes nepochybně vyhledávání podobnosti struktur v biologické chemii a ve farmaceutickém výzkumu.

Praktickou mírou úspěchu v hledání je faktor obohacení, *ef* (enrichment factor), udávající poměr podílu aktivních molekul ve vybrané podmnožině k podílu aktivních entit v celé databázi (pool),
 $ef = \text{frakce aktivní podmnožiny} / \text{frakce aktivního poolu}.$

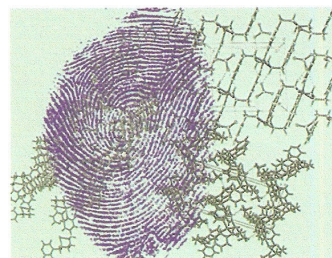
V literatuře můžete nacházet stále další a sofistikovanější molekulární deskriptory, ty ovšem vždy vycházejí z vhodné molekulární reprezentace definující chemický prostor a z modifikace uspořádání molekul v tomto prostoru. Molekulární deskriptor je konečným výsledkem logické a matematické procedury transformující chemickou informaci zakódovanou v symbolické reprezentaci molekuly do zvládnutelného počtu výsledků standardizovaného experimentu (Todeschini a Consonni).

Významnou pasáží deskriptorů jsou deskriptory substruktur, fragmentů a dvojic vzájemně vázaných atomů, představujících (potenciální) reakční centra (v textu jsou označovány jako atomové vektory, AV). **Hledání maximální společné substruktury** je praktikováno již dlouho a je využíváno k nacházení substruktur, které jsou podobné substruktuře templátu. Pro řešení reaktivity eduktů je to zásadní úkol, od počátku tohoto přístupu je, jak o tom napovídá bohatá literatura, studován a byla publikována řada řešení. V některých případech se vystačí při definování aktuální podobnosti s fuzzy metodikou. Je třeba si uvědomit, že hledání substruktur vyžaduje dokonalé zobrazení molekulárních konektivit a druhů atomů, časově náročné vzhledem k NP-úplné povaze problému. Postupy jsou uplatňovány a zdokonalovány zejména při studiu biologické aktivity v souvislosti s hledáním nových léčiv. Ve vstupu do problému si můžeme všimnout znázorňování molekul (Rarey, Dixon) jako struktur podobných stromům v teorii grafů (označovaných „Feature Trees“), které mají vztah k redukovaným grafům; v nich soubory atomů spadají do jednotlivých vrcholů. Jiným způsobem reprezentace molekulárních deskriptorů jsou „otisky prstů“ (fingerprints) molekulárních stromů (MOLPRINT 2D), významné při posuzování dvou vazeb vycházejících

z centrálního atomu, a samozřejmě QSAR studie. Rozsah oblasti podobnosti je postupně rozšiřován, dosud na pět až sedm seřetězení (třeba Daylight fingerprints).

Ve farmaceutickém průmyslu je široce využíván klastrovací algoritmus Jarvise-Patricka (J-P) s Daylight programy: pracuje s Daylight otisky prstů a Tanimotovým indexem podobnosti. Ve srovnání s jinými procedurami zdůrazňuje autor (Butina D., 1998) významnou časovou nenáročnost klastrování báze dat a možnost využití k řešení velkých heterogenních i malých souprav.

Parkin s kolektivem spolupracovníků publikovali (*CrystEngComm* 2007, 9, 648, ref. v *Highlights Chemical Science* 2007) metodu, kterou nazvali **strukturní genetické otisky prstů** a její význam spatřují hlavně pro hledání podobnosti krystalických látek (ve farmaceutickém výzkumu, v materiálové chemii a vůbec všude, kde se chemici setkávají s látkami v tuhém stavu).



Ke zobrazení (které má spíš ilustrativní účel) autor dodává, že například naftalen je více podobný anthracenu než benzenu.

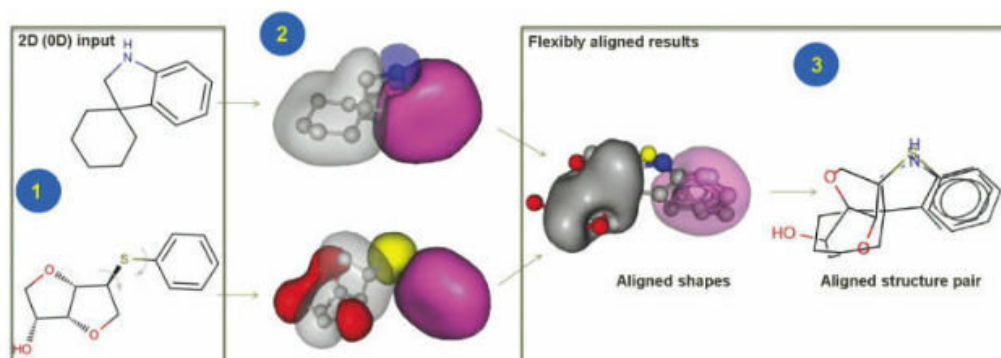
Virtuální prosívání založené na podobnosti předpokládá, že všechny sloučeniny v databázi, podobné hledané molekule, mají stejné chemické a biologické vlastnosti. Přesto, že tento předpoklad nemusí být vždycky splněn, je tato prohledávaná množina mnohdy obohacena o aktivní systémy. Při prohledávání milionů sloučenin jsou molekulární struktury obvykle reprezentovány strukturními klíči (*molecular screens*) nebo molekulárními „otisky prstů“ (*molecular fingerprints*); obojí obsahují 2D a 3D informace, závisí to na velikosti molekulárních systémů. Pro zpracování bohatých databází se nabízejí procesy založené na metodice otisků prstů, které mají větší informační hustotu (Daylight, BCI, UNITY 2D, Tripos jsou nejznámější příklady). Populární je Tanimotův (nebo Jaccardův) koeficient T . Je-li hodnota $T > 0,85$, jsou dvě srovnávané struktury považovány za podobné.

Pro chemii je významným znakem, že změny molekul při jejich interakcích nejsou omezeny na kvantitativní fyzikální vlastnosti jako jsou volná energie, hustota náboje nebo koncentrace. Interakcemi molekul jsou generovány nové sloučeniny. V tom je hlavní rozdíl pro teoretické zpracování chemických systémů. Modely dynamických systémů vycházejí z fixní množiny molekulárních species a sledují jejich koncentrace v závislosti na čase, podobně jako dynamika populací popisuje frekvence genů v organismech. Přijmeme-li tento obraz, molekuly, geny a biologické struktury lze redukovat na několik indexů proměnných koncentrací, zatímco jejich vlastnosti jsou implicitně popsány rychlostními konstantami, které je nutné stanovit mimo tyto teorie. Běžně je popsáno v chemických databázích zhruba 30 milionů sloučenin, ve skutečnosti je jich nekonečně mnoho.

Nejzajímavější otázkou v chemii je postavení jmenovitých molekul: Fontana a Buss zavedli notaci konstruktivního dynamického systému, v němž interakce = chemické reakce jsou posuzovány jako algebraické operace na množině typů molekul. Pro molekuly má Fontana λ kalkul a pro jejich reakce zavedl „aplikaci“ termu λ na reakčního partnera. Výsledkem je nový λ term. Uvedené termy jsou ovšem hrubými abstrakcemi reálných molekul.

K virtuálnímu screeningu došli chemici při řešení problému, jak redukovat enormní virtuální prostor chemických sloučenin, aby mohli pracovat s přijatelným počtem složek syntéz, zejména při hledání potenciálních farmak. K virtuálnímu screeningu se dochází dvěma způsoby: metodou založenou na ligandech (LBVS, ligand-based virtual screening) vycházející z 2-D a 3-D podobnosti mezi velkými databázemi sloučenin a z jejich známých aktivit, a metodou založenou na struktuře (SBVS, structure-based virtual screening), která aplikuje různé techniky modelování tvorby vazeb. Metoda SBVS se opírá o strukturní informace cílové sloučeniny (rtg krystalografie nebo NMR atp.). Postupné kroky nezbytné v metodě SBVS jsou dány v aktuálních ukládacích algoritmech; přitom všechny kroky mají jistý stupeň předpokladů, takže je třeba počítat s artefakty. Významným momentem metody je práce

s dynamickými modely. Ukázka principu SBVS je převzata ze sdělení Solt, Tominová, Niesz (ChemAxon Ltd. 2013):



Ukázka návazných kroků při 3-D flexibilním spojování formací: 1. struktury eduktů; 2. v tomto stadiu je vytvářen konformer 3-D a útvary jsou vybarveny podle typů atomů; 3. dochází k průniku objemů, který se maximalizuje v průběhu vzniku vazebného propojení. (*J. Chem. Sci.*, 2005, 117, 425).

Kvantifikace podobnosti využíváním míry podobnosti a indexů podobnosti spolu se srovnávacím výzkumem jsou využívány při hledání substruktur a prostředkem je zmíněná metoda otisků prstů. Je to velice abstraktní reprezentace určitých strukturních znaků molekuly a její užitečnost podtrhuje to, že tato technika byla v počátcích informačního systému známého jako Daylight Chemical Information Systém.

Zájemci o tento prostředek především chemoinformatiky mohou vyhledat například sělení *Structural genetic fingerprinting* autorů Parkin A. a d. (*CrystEngComm* 2007, 9, 648) s dalšími 40 odkazy, případně publikaci Eckert H., Bajorath J., *Drug Discovery Today* 2007, 12, 225 ad.

Předpoklady pro hledání molekulární podobnosti jsou uváděny v **principu molekulární podobnosti MSP (Molecular Similarity Principle)**:

- Vždy jde o srovnání v daném souboru a v závislosti na hledaném reakčním centru;
- Běžně užívané deskriptory podobnosti popisují statické systémy, je třeba brát v úvahu prostorovou měnivost interakcí reakčních center substrátu a reagentu (to platí zejména o receptorech proteinů);
- Je třeba znát odpověď na otázku, která z vlastností systému je dominantní pro srovnávání;
- Podobnost je zpravidla nelineární problém; jsou využívány učící algoritmy, ty nemusí být vždycky aplikovatelné;
- Je-li možných aktivních míst několik, nemusí být platný vztah „podobné molekuly = podobný efekt“;
- Zjednodušený může být názor „v jedné molekule se uplatňuje H-vazba, ve srovnávané molekule také“, přehlíží jemné rozdíly, které se mohou uplatňovat vlivem solvatace;
- Častým procesem je kooperace ve vazebných vztazích, která může situaci komplikovat; Podobnost substruktur a fragmentů je možno kvantifikovat pomocí analýzy databáze.

Metody hledání podobnosti mají pochopitelně zásadní význam ve farmaceutickém výzkumu, a rovněž lékařská chemie využívá koncepci biosterie, nabízející záměnu podobných substruktur, u kterých se předpokládá výrazná míra aktivity. A je namístě zmínit rostoucí popularitu metod založených na podobnosti v chemických technologiích, zejména při pokročilých aplikacích výkonných počítačů. Snad jen pro zajímavost se zmíním o počátcích těchto aplikací: Koncem šedesátých let minulého století Corey a Wipke (a po nich řada dalších autorů) uvedli ve známost počítačový program (LHASA) využívaný poté k návrhům syntéz organických molekulárních sloučenin. Tento průkopnický projekt odstartoval sestavování dalších programů a o významu navrhování syntéz s pomocí počítačů svědčí i to, že počítačové programy první a druhé generace adoptovala bezprostředně k praktickému využívání jako první firma Merck. Uvedený námět sestavovat stromy syntéz zejména léčiv s pomocí počítačů se postupně rozvinul do dnešních podob predikcí vhodných a zamítaných kandidátů chemických struktur s požadovanými vlastnostmi.

Klasifikaci chemických kategorií založenou na podobnosti zavedli Salinerová, Patlewicz a Worth (2005).

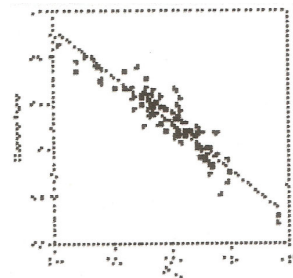
Obecně se uvádí 12 charakteristik a projevů chemických systémů, které jsou dány jejich strukturou, a kromě toho byla publikována řada kvantitativních měr (indexů MS) pro exaktní vyjádření molekulární podobnosti. Definice podobnosti molekul vychází z **mapování chemického prostoru**: molekulu reprezentují relevantní deskriptory v jednorozměrném prostoru reálnými čísly. Takže definice MS odpovídá reprezentaci molekul v prostoru deskriptorů. Molekulární struktury mohou být popsány jako uspořádaná množina složek vystihujících informace o vztazích mezi těmito složkami. Informaci je možno uvádět jako seznam indexovaných atomů a vazeb, ze kterých molekula sestává (molekulárními kódy), nebo formou „účetnictví“ složek různých druhů, kterými jsou vyjádřeny matematické vlastnosti struktury. Tyto matematické soubory musejí být strukturně invariantní v tom smyslu, že zůstávají nezměněny i když se do struktury přidá další komponenta. Různé úrovně strukturního popisu závisí na stupni informace a Saliner, Patlewitz a Worth (2005) je shrnují v těchto bodech:

- 1) Soupis druhů atomů obsažených v molekule.
- 2) Empirický vzorec vyjadřující poměry zastoupení jednotlivých atomů.
- 3) Molekulární vzorec, který udává počet atomů jednotlivých druhů (protonových čísel); z něj se počítá „molekulová váha“.
- 4) Dvourozměrný strukturní vzorec znázorňuje vzájemné uspořádání atomů - topologii molekuly a konektivitu atomů. Alternativní reprezentace jsou známy jako SMILES („Simplified Molecular Input Line Entry Specification“) a InChI kódy.
- 5) Trojrozměrné struktury popisují strukturu molekuly jako prostorový útvar, ve kterém jsou atomy uloženy v daných polohách v souřadnicovém systému x,y,z a tak poskytují geometrické a prostorové informace o molekule.
6. Řešení Schrödingerovy rovnice a popis rozložení elektronových hustot dávají přesný popis (závislý samozřejmě na úrovni teorie použité k výpočtům) a jsou počítačově zpracovatelné.

Reprezentace chemického systému je řešena v termtech konstituce, konfigurace a konformace. Konstituce poskytuje informaci o sledu vazebné konfigurace a popisují ji topologické deskriptory, přítomnost nebo absence fragmentů a deskriptory reprodukcující dvourozměrnou stavbu molekuly. Konfigurace je pak definována trojrozměrným nebo prostorovým uspořádáním atomů popisujícím délky vazeb a úhly mezi nimi a deskriptory tvaru molekuly. Prostorové uspořádání atomů, které umožňuje rozlišit stereoizomery přecházející jeden v druhý otáčením kolem (formální) σ -vazby, je konformace.

Uvedené informace umožňuje získat řada kvantitativních popisů molekulárních struktur pomocí vhodných deskriptorů: jen zmínkou jsou to široce uplatňovaný QSAR, pak DRAGON, ADAPT, OASIS a CODESSA a další.

MQS-SM (Molecular Quantum Self-Similarity Measure) představuje dílčí případ obecné formulace MQSM (Molecular Quantum Similarity Measures). Podle Gironese, Carbó-Dorcy a Ponce (*J.Chem.Inf.Comput.Sci.*, 2003, 43, 2033) je vhodný pro kvantitativní charakterizaci podobnosti mezi dvěma molekulárními strukturami A a B na základě srovnání jejich funkcí hustot. Obecná definice míry



kvantové podobnosti molekul A a B, $Z_{AB}(\Omega) = \iint \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1,\mathbf{r}_2)\rho_B(\mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2$, když $\rho_A(\mathbf{r}_1)$ a $\rho_B(\mathbf{r}_2)$ jsou funkce hustoty (DF) molekul A a B a $\Omega(\mathbf{r}_1,\mathbf{r}_2)$ je arbitrární (kladný) operátor. V jednoduchém případě, když $\Omega(\mathbf{r}_1,\mathbf{r}_2) = \delta(\mathbf{r}_1,\mathbf{r}_2)$, se získá míra podobnosti překryvu (Overlap similarity measure) $Z_{AA} = \int \rho_A(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r}$. Elektronová hustota fragmentu $\rho_A^X(\mathbf{r})$ je pak mírou kvantové podobnosti fragmentu.

Grafická forma korelační rovnice MQS-SM vs. Hammettovy σ parametry pro 110 substituentů benzoových kyselin.

Ponec R.: Molecular similarity and LFER. Molecular similarity and reactivity from quantum chemical to phenomenological approaches (Ed. R. Carbo), Kluwer . Ponec R.: Similarity models in the theory of pericyclic reactivity. Springer Verlag, Heidelberg

K vyjádření podobnosti slouží především deskriptory charakterizující chemické sloučeniny, jak jich rutinně využívá každý chemik:

- Empirické parametry užívané v klasických modelech QSAR (např. v Hanschově metodě, a připomínáme si její fyzikálně chemické parametry klasifikované jako elektronové, hydrofobní a sterické k vystižení rozložení elektronové hustoty, afinity k vodě a k interpretaci tvaru, geometrie a sterických nároků molekul). Samozřejmě jsou užitečné i hodnoty rozpustnosti, teplot tání a varu a spektroskopické údaje.
- Důležité jsou teoreticky počítané vlastnosti: topologické deskriptory a všechny údaje nabídnuté výpočetní chemií.
- Parametry získávané 3D-QSAR technikami, umožňující procedury superpozicí molekul. Patří k nim indexy molekulární podobnosti (MI) a topologické indexy molekulární podobnosti (QMI). Jsou aplikovatelné na molekuly jako celky i na jejich substrukтуры.

Tak se dozvídáme informace o tom, jak souvisí reakční rychlosti s rozdíly v molekulárních strukturách a získáváme obraz uplatnění funkčních skupin prostřednictvím jejich konstant substituentů a sterických konstant. Celková reprezentace soustavy je dána uvedenými parametry nebo nově zaváděnými deskriptory, získávanými především kvantově mechanickými výpočty zahrnujícími trojrozměrné konformační informace.

Ještě o jednom programu použitelném k hledání chemické podobnosti se zmíníme: Je jím **Princip hledání podobnosti**. Chemik aplikuje kritérium podobnosti s cílovou molekulou a získá transformovanou cílovou strukturu. Tu srovnává s každou z transformovaných sloučenin v katalogu; pokud jsou transformovaná cílová molekula a transformovaná katalogová molekula shodné, je výchozí molekula z katalogu proponována programem WODCA (Workbench for the Organization of Data for Chemical Applications) za vhodný startový edukt. Program WODCA obsahuje pro definování molekulární podobnosti čtyřicet různých kritérií. Podobnost je v něm definována na základě substruktur nebo zobecněných chemických reakcí. Tato kritéria jsou kombinována.

Chemoinformatics: A Textbook. Edited by Johann Gasteiger and Thomas Engel Copyright 2003 Wiley-VCH Verlag GmbH & Co. KGaA

Bender (2004, 2005) vyjádřil a dokládal jisté pochybnosti o oprávněnosti představ o šíři využitelnosti principu molekulární podobnosti (MS) v různých souvislostech teoretické chemie (soustředil se zejména na biologickou aktivitu).

Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S.. *J. Chem. Inf. Comput. Sci.* 2004, 44, 170-178.

A jakkoliv byla teorie podobnosti označena za jeden z nosných a osvědčených opěrných pilířů chemie, je dobré zamyslet se nad sdělením Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping (publikovaném kolektivem Muchmore S.W., Debe D.A., Metz J.T., Martin Y.C., Hajduk P.J., *J.Chem.Inf. Model.*, 2008, 48 (5), 941).

Teorie důvěryhodnosti (Belief Theory) je rozpracována v epistemologii a psychologii a tak jen stručně: Primárním problémem epistemologie je přesné pochopení obsahu naší znalosti o daném objektu, problému; důvěryhodnost je vědění (znalost), je-li to, čemu věříme, pravdivé a máme-li o jeho pravdivosti oprávněnou jistotu. Psychologové hovoří tradičně o důvěryhodnosti jako o nejjednodušší formě myšlenkové reprezentace, konstruktech, z nichž se skládají naše úvahy. Z Benderovy úvahy převzeme tyto názory: Pro naše „selské myšlení“ (common-sense) je důvěryhodnost korektním způsobem porozumění věci; nemusí být korektní a přitom nám umožňuje vyslovovat užitečné prognózy; naše běžné chápání důvěryhodnosti je chybné a mělo by být nahrazeno jinou teorií; naše „selské myšlení“ není bezchybné, ale pro člověka a zejména pro počítače nabízí použitelnou strategii.

V pojednání o podobnosti bychom si mohli pohrávat se slovní hříčkou podobnost – pravděpodobnost. V pravděpodobnostním usuzování (Approximative Reasoning, AR) slouží k reprezentaci a ke zpracování neurčitého vědění **fuzzy logika**.

Operace se symboly v umělé inteligenci (AI) jsou obohaceny o operace s lingvistickými proměnnými. Na rozdíl od klasické logiky, založené na booleovské dvouhodnotové logice (+, - ; ano-ne ; P-N) je fuzzy logika (FL) vícehodnotová (zcela P, částečně P, spíš N, určitě N). Jistý druh propozice je zpracováván jako elastický projev proměnné. Proměnná je ve fuzzy logice buď lingvistická, nebo fuzzy, nebo v klasické formě (predikátové logiky). (Ještě poznámka: predikát fuzzy je překládán do češtiny neostrá, mlhavá množina, a také se nepřekládá.)

Fuzzy logika se zabývá reprezentací a operacemi s neurčitým vědění. Využívání symbolů umělé inteligence vede k operacím s lingvistickými proměnnými. Na rozdíl od klasické logiky založené na dvouhodnotové Booleově logice je fuzzy logika vícehodnotová. Na základě fuzzy logiky se teorie „approximate reasoning“ rozvinula mj. v expertní systémy (ES).

M. Otto, Anal. Chim. Acta 1990, 235, 169

Vyvozovací pravidla FL připomínají pravidla deduktivního vyvozování závěrů v klasické logice. Na bázi FL je rozvíjena teorie pravděpodobnostního usuzování, je úspěšně aplikována v expertních systémech, speciálně pak v ES založených na pravidlech (rule-based expert systems).

Relace fuzzy nerozlišitelnosti (a fuzzy podobnosti) mohou splňovat fuzzy podmínky reflexivity, symetrie a tranzitivity, aniž by narážely na Poincarého paradox

- Fuzzy reflexivita: $R_{xx} = 1$
- Fuzzy symetrie: $R_{xy} = R_{yx}$
- Fuzzy tranzitivita: $R_{xy} \& R_{yz} \Rightarrow R_{xz}$ (pro všechna x, y, z)
- Fuzzy relace ekvivalence (splňující tyto 3 podmínky) se nazývají fuzzy relace podobnosti (či nerozlišitelnosti)
- Od relací fuzzy rovnosti se navíc vyžaduje podmínka, že $R_{xy} = 1$ jen pro $x = y$
- Podobně se zkoumají relace fuzzy uspořádání apod., obecně jde o teorii fuzzy relací (část teorie fuzzy množin).

Mezey považuje podobnost za fuzzy ekvivalenci.

Máme-li fungující matematický model, který nás přivede přes zákonitosti konverzí valenčních stavů atomů k postižení pravděpodobných sekvencí kroků reorganizace valenčních elektronů na daných reakčních centrech reaktantů a k ocenění míry filicity reakčních center substrátu a reagentu výpočty jejich tvrdostí, pak nám týž matematický model umožní vstoupit do pole More O'Ferrallových diagramů s „fuzzy přesností“. Ta však postačuje k odhadu intervalu hodnot β_{LG} (nebo jiného parametru na příslušné souřadnici). Aparát chemické podobnosti (Chemical Similarity) nám nabídne v referenčních seriích fuzzy hodnotu rychlostní konstanty k modelované reakční soustavy (nebo jiný užitečný parametr reaktivity) a přes Brønstedův vztah $\log k = \beta \log K + c$ se nabízí možnost odhadu fuzzy hodnoty rovnovážné konstanty, blízké reálné hodnotě. Přes ni ($-\Delta G^0 = 2,303RT \log K$) pak můžeme vyslovit výrok o pravděpodobné uskutečnitelnosti suponané reakce za standardních podmínek.

O fuzzy podobnosti se zmíníme ještě v další významné souvislosti. V hledání strukturní podobnosti je podstatnou fází **shlukování**, klastrování, definované jako příslušnost množiny objektů do podmnožin na základě jejich podobnosti. Fuzzy podmnožina je definována jako funkce

$$c : X \rightarrow [0,1] .$$

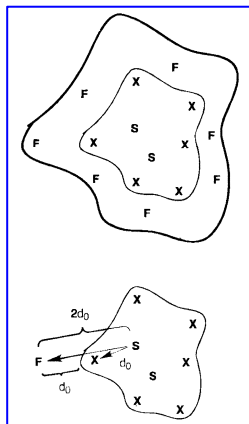
Její hodnota je považována za „stupeň příslušnosti“ nebo za „určitost příslušnosti“ (že do množiny určitě spadá) k dané množině. na základě tohoto paradigmatu lze definovat obsah termínu „neostré shlukování“ $X: \{c_1, \dots, c_n\}$ s následujícím řešením:

1. Pro každé i , c_i je neostrá podmnožina X
2. Pro každé i existuje alespoň jedno $x \in X$ s $c_i(x) = 1$
3. Pro každé $x \in X$ je alespoň jedno i s $c_i(x) = 1$

4. Je-li $i = j$, pak pro každé $x \in X$ platí $c_i(x) = 1$ nebo $c_j(x) = 1$.

Úplný popis algoritmu 5, A Technique for Fuzzy Similarity Clustering of Chemical Inventories. Doman ad., *J. Chem. Inf. Comput. Sci.*, 1996, 36 (6), 1195, který překonává některé nesnáze široce využívaného algoritmu Jarvis a Patricka, najde čtenář v citované publikaci.

A dnes už klasická publikace Chemical Similarity Searching: Willett, Barnard, Downs, *J. Chem. Inf. Comput. Sci.* 1998, 38, 983 stojí stále za přečtení.



Je znázorněn idealizovaný algoritmus 5 ve dvou dimenzích. Symbol „S“ označuje výchozí sloučeninu, „X“ je neostře klastrovaná sloučenina mimo ní a „F“ představuje neostře shlukované sloučeniny. V horní části vyobrazení představuje vnitřní cykl neostrý shluk, zatímco vnější kruh obsahuje souhrnný fuzzy klastr. V dolní části náčrtu je vzdálenost mezi „S“ sloučeninou a „X“ sloučeninou 0,15 (v Tanimotově prostoru), vzdálenost od „X“ k jednomu z „F“ je také 0,15, takže vzdálenost od „S“ k „F“ je 0,30 nebo $2d_0$. Z toho vyvodíme, že fuzzy sloučeniny jsou přibližně d_0 od „hrany“ neostrých klastrů.

Model vytvořený na matematickém fundamentu je v podstatě založen na pravděpodobnostním usuzování. Modelem prolíná a posouvá jej do teorie chemické reaktivity parametrizace výstupů a kvantifikace jejich řešení.

Podobnost dvojice funkcí může být vyjádřena určením kartézského součinu množiny objekt-funkce s tím, že tento součin množin je **fuzzy množina S**. Neostrost nad S se zavede do dichotomické situace tím, že o některém prvku x v S platí $0 < M(x) < 1$. Pak lze sestavit kartézský součin neostrých množin s dvojicemi funkcí členů $M(x)$, uplatňující se na uspořádaných dvojicích prvků množin. Míru podobnosti lze tedy vystihnout i v režimu fuzzy množin a fuzzy logiky.

(Zalewski, Krygowsky, Shorter, *Similarity Models in Organic chemistry, Biochemistry and Related Fields*. Elsevier 1991).

Míra $C(F_a, F_b) \approx 0$ přísluší tzv. marginálnímu členu. Protože každá funkce může být spjata korespondencí 1:1 s přesně definovanou kolekcí molekulových orbitalů, molekul, stavů, reakčních sérií ap., platí to i o míře C a o vytvoření uspořádání nad množinou funkcí. Důsledek – uspořádání množiny funkcí F indukuje analogické uspořádání množiny objektů S :

$F = \{f_1, f_2, \dots\}$ a $S = \{s_1, s_2, \dots\}$;

$K = [F \leftrightarrow S \parallel \{f_1 \leftrightarrow A_1, f_2 \leftrightarrow A_2, \dots\}]$;

je-li vyvozeno pravidlo uspořádání R v F , pak $R[F] = \{F_1 < F_2 < \dots\}$, a to indukuje $R[S] = \{s_1 < s_2 < \dots\}$.

Racionální jádro této koncepce se opírá o **fuzzy podobnost**.

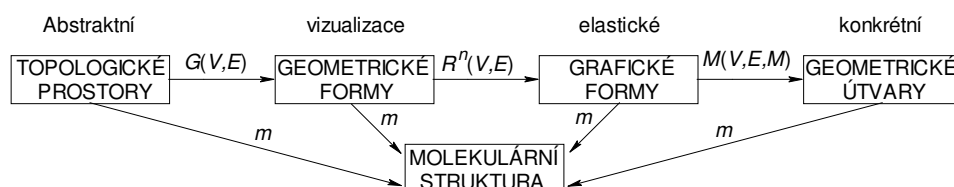
Logickou stránkou této koncepce se zabývá (m.j.) práce Godo, Rodríguez *Logical approaches to fuzzy similarity-based reasoning: an overview*, také Kvasnička ve studii Mentálne modely logiky v kognitívnej vede. *Kognice a umelý život* (2007). vystižení hodnot fuzzy podobnosti můžete nalézt např. v publikaci Sridevi, Nadarajan, Fuzzy Similarity Measure for Generalized Fuzzy Numbers, *Int. J. Open Problems Comput. Math.*, 2009, 2; v *Annals of Fuzzy Mathematics and Informatics* 2013, 6, 251 vyšla práce Beg, Ashraf o Kleeneově fuzzy podobnosti a mírách podobnosti s řadou příkladů aplikací. Většina zavedených hodnot míry podobnosti je řešena s fuzzy čísly, která jsou definována diskontinuuitními funkcemi. Jejich analýzou se zabývali Mitrovič a Zoran (*FME Transactions* 2006, 34, 115): rozbor vedli k míře podobnosti mezi různými množinami objektů a zavedli míru ZA, B s uspokojivými řešeními.

V našem rozboru zůstává stálým pozadím vědecká metoda. O ní víme, že je specifickou formou myšlenkových procesů a v souvislosti s ní je zmíněno **geometrické myšlení**: supluje formální a konkrétní operační myšlení topologickým a geometrickým myšlením. To je založeno na geometrické představivosti („zvnitřněná názornost“). Uskutečňuje se na různých stupních abstrakce a pochopitelně se u různých posuzovatelů může lišit. O geometrické představivosti platí, že je schopností, dovedností poznávat geometrické útvary a jejich vlastnosti, abstrahovat od reálné skutečnosti a v konkrétních objektech vidět geometrické útvary v jejich čisté podobě, i na základě rovinných obrazců si představovat prostorové formy v jejich různých vztazích (i takových, na jaké nemohou být převedeny pomocí hmotných modelů), pracovat s tesaurem představ geometrických útvarů a umět si je vybavovat a umět si představit geometrické útvary na základě pouhého popisu. Je potřeba formovat v mysli představy dynamiky změn (velikosti, tvarů a poloh) útvarů mít schopnost mentální manipulace s útvary.

Prostorová představivost je restrikcí geometrické představy na 3D, prostorové myšlení je schopnost restrikce geometrické myšlenky na 3D.

Elementární chemická gramatika - strukturní teorie - je svým formalismem klasických strukturních vzorců (a nepochybně díky analogii mezi strukturními vzorci a grafy) blízká topologickému myšlení. Jednoznačnost přiřazení mezi strukturními chemickými vzorci a molekulovým grafem je zřejmá nejen na úrovni klasického popisu, ale i v rámci kvantové topologie (Turro).

MYŠLENÍ



Schematické znázornění vztahů mezi topologickými prostory, grafickými formami, geometrickými formami a geometrickými útvary. Graf je matematická struktura sloužící k modelování skutečnosti, že mezi prvky nějaké množiny X existují určité vztahy. Prvkům množiny X odpovídají vrcholy grafu V , vazbám hrany grafu E . R^n je euklidovský n -rozměrný prostor ($n = 1, 2, 3$). Do vztahů vstupuje metrika M . Mapovací proces je označen m . Elastická geometrie je topologie.

Podobnost a chemotopologické studie.

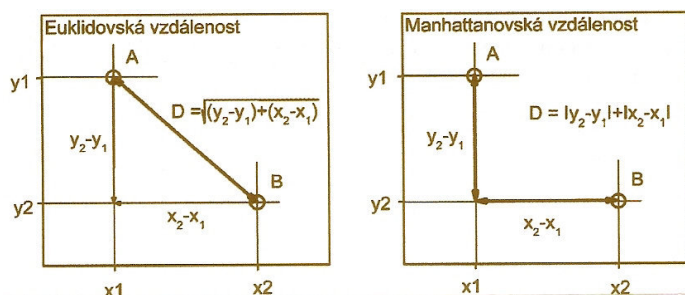
Informační teorie a pravidla uspořádanosti zaujala Boncheva (2006) v souvislosti se zobecňováním názoru, že distribuce elektronů v atomech souvisí s jistým principem maximálního informačního obsahu. Jiný přístup nabízí teorie množin: je množina Q chemických prvků q , lze tedy definovat Q jako $Q = \{q \mid q \text{ je chemický prvek}\}$. Mendělejevova explikace říkala, že q je určen svými vlastnostmi a v jeho době tou vlastností byla atomová hmotnost a periodicitu některých dalších vlastností označil jako valenci. K hlubšímu řešení přispěli Robert a Carbó-Dorca, (1998, 2000) aplikací teorie množin. Zavedli koncepci **kvantové atomové podobnosti** v návaznosti na kvantovou molekulární podobnost. U dvaceti vybraných atomů srovnáním míry kvantové podobnosti počítali (Ponec a Carbó-Dorca, 1998, 2000) míru „self-similarity“ a dospěli k míře prostorového obsazení molekuly atomy.

Další metodou studia podobnosti chemických prvků je **shluková analýza**, sloužící jako matematický prostředek hledání vztahů (Zhou ad., 2000; Sneath, 2000; Resptrepo ad., 2006). Podobnost je počítána ze vzdálenosti nebo z metrické funkce. (a a b jsou si podobné, jsou-li v těsné blízkosti v matematickém prostoru, ve kterém jsou definovány.) Shlukovou analýzou se získá dendrogram (Restrepo, 2006), graf, který může být interpretován jako mapa podobností. Ukázka je v kontextu Izovalentní hyperkonjugace. Nejvýznamnějším ziskem uvedeného přístupu (Zhou ad., 2000) jsou výsledky plynoucí z možnosti využívat matematické prostředky ke studiu chemických prvků a využívání vlastností k definici každého chemického prvku; a přínosem shlukové analýzy je definice podobnosti v matematických termínech.

Shluková analýza (také clusterová analýza, v angličtině cluster analysis) je vícerozměrná statistická metoda, která se používá ke klasifikaci objektů. Slouží k třídění jednotek do skupin (shluků) tak, aby si jednotky náležící do stejné skupiny byly podobnější než objekty ze skupin různých. Shlukovou analýzu je možné provádět jak na množině objektů, z nichž každý musí být popsán prostřednictvím stejného souboru znaků, které má smysl v dané množině sledovat, tak na množině znaků, které jsou charakterizovány prostřednictvím určitého souboru objektů, nositelů těchto znaků.

Elegantní vstup do problematiky shlukové analýzy může čtenář číst v informativním přehledu prof. Melouna a prof. Militkého (milan.meloun@upce.cz) *Přednosti analýzy shluků ve vícerozměrné statistické analýze*.

V chemii využíváme shlukovou analýzu v taxonomii (empirické klasifikaci objektů) a zejména k identifikaci vztahů mezi objekty. Kritériem vytváření shluků objektů je podobnost: po stanovení znaků, podle nichž hledáme podobné systémy, se provádí jejich kombinace do podobnostních měr. Měření se provádí různými způsoby. Jsou korelační míry (korelační koeficienty mezi dvěma sloupci matice představují korelaci mezi dvojicí objektů); jsou míry vzdálenosti, nejčastěji využívané v mnoha oborech aplikovaného principu podobnosti – geometrická metrika, také Euklidovská vzdálenost, a Manhattanovská vzdálenost (Hammingova metrika), případně Minkovského metrika. S termínem vzdálenosti se při četbě o podobnosti setkáváte, proto pro přiblížení názornosti uvedu ukázky převzaté z citovaného Melounova pojednání:



Euklidovská vzdálenost D a Manhattanovská vzdálenost D představují míry založené na prezentaci objektů v prostoru, jehož souřadnice tvoří jednotlivé znaky.

Další míry jsou míry asociace založené na otázkách a odpovědích ano – ne, využitelné například v expertních systémech.

Vyjádřit uvedené vztahy čísly je možné indexy podobnosti.

Uvedeme si ukázky klasických SI. S elektronovými strukturami operuje index, který zavedl Carbó se spolupracovníky a který je široce využíván ve studiích QSAR,

$$R_{AB} = \frac{\int \rho_A(\vec{r}) \rho_B(\vec{r}) d\vec{r}}{\{[\int \rho_A^2(\vec{r}) d\vec{r}][\int \rho_B^2(\vec{r}) d\vec{r}]\}^{1/2}}$$

a Hodgkinův-Richardsův index

$$H_{AB} = \frac{2 \int \rho_A(\vec{r}) \rho_B(\vec{r}) d\vec{r}}{[\int \rho_A^2(\vec{r}) d\vec{r} + \int \rho_B^2(\vec{r}) d\vec{r}]}$$

V nich $\rho_A(r)$ a $\rho_B(r)$ jsou elektronové hustoty molekul A a B , a integruje se přes celý prostor. Čítec ve vztahu vystihuje míru překryvu elektronové hustoty dvou superponovaných molekul, jmenovatel je normalizační faktor.

Fyzikální obsah těchto indexů diskutoval Ponec, který navrhl též topologický index podobnosti

$$R_{AB} = \frac{\int \rho_A \rho_B dv}{(\int \rho_A^2 dv)^{1/2} (\int \rho_B^2 dv)^{1/2}}$$

Index R_{AB} představuje koeficient křížené korelace a nabývá hodnot $0 \leq R_{AB} \leq 1$; pro dokonalou podobnost 1, 0 pro úplnou dissimilaritu funkcí elektronové hustoty. Carbóův index je vhodný zejména pro porovnávání optických izomerů i koeficientů při rentgenografické krystalografii proteinů.

Úspěšnost metody závisí na přenositelnosti a aditivitě vlastností atomů, což jsou podmínky operací s fragmenty molekul i v intencích teorie atomu v molekule, AIM.

Index R_{AB} modifikovali Meyer a Richards do tvaru, který reflektuje podobnost tvaru jak van der Waalových, tak také povrchů stejné hustoty (izodensity) $S_{AB} = B_{AB} / (T_A \cdot T_B)^{1/2}$.

B_{AB} je počet bodů shodných u obou molekul, T_A počet bodů připadajících na povrch molekuly A (srv. informaci o skládání MESP).

KVANTOVĚ MOLEKULÁRNÍ PODOBNOST

Přednostní zájem chemie je soustředěn na porozumění vztahům mezi strukturou látek a jejich reaktivitou, QSRR (Quantitative Structure-Reactivity Relationships), a na jejich využití především ve farmaceutickém výzkumu. Spolu s bioorganickou chemií, biologií, toxikologií, farmaceutickou chemií a environmentální chemií se organická chemie angažuje také na studiu vztahů mezi strukturou a aktivitou látek, QSAR (Quantitative Structure-Activity Relationships): organická chemie se zabývá interakcemi mezi reaktanty, které za daných podmínek probíhat mohou, biochemii zajímají interakce mezi chemickými sloučeninami a všemi formami složek živých systémů, DNA, enzymy, organelami, buňkami, membránami ad., které v organismech probíhají.

Porozumění vztahům „chemie – živé systémy“ zprostředkovávají vztahy QSAR (Hansch) a základem jejich řešení je Hammettova rovnice rozšířená o sterický parametr E_s a hydrofobní konstanty F .

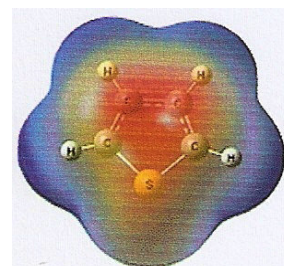
Od poloviny minulého století se chemici začali zabývat teoretickým základem a aplikacemi vztahů mezi strukturou a vlastnostmi (SAR). Koncepte kvantitativních vztahů struktura – aktivita, QSAR, má pendant v QSPR, kvantitativním vztahu mezi strukturou a vlastnostmi (Quantitative Structure-Property Relationships). Na tradiční modely QSAR, pracující s matematickými deskriptory, navazují aplikace technik molekulového modelování, dovolující vysoké rozlišení s kvalitou *ab initio* výpočtů jedné molekuly (může jí být jedna molekula peptidu s více jak 1000 atomů). Ve výbavě techniky MEDLA (Molecular Electron Density Lego Assembler) jsou geometrie atomových jader získané rentgenostrukturní analýzou, optimalizované konformace, elektronové hustoty a elektrostatické potenciály, informace ekvivalentní kvantově chemické vlnové funkci; výpočty jsou uspokojivě rychlé. ukázka povrchu elektronové hustoty taxolu, generovaného metodou MEDLA,. Aplikace techniky molekulového lega na modelování elektronových hustot aminokyselin je v některé z dalších statí.

Technikou MEDLA vytvořený povrch elektronové hustoty taxolu; srovnáme se strukturním vzorcem tohoto kancerostatika.

Technika QShAR (Quantitative Shape-Activity Relationships) je založena na kvantově chemických deskriptorech elektronové hustoty na povrchu molekul. Data jsou rozložena do podmnožin a uspořádána v matici s mnoha sty prvky – modely. Speciální variabilní výběr je spojen s kvalifikací predikce a rutina odpovídá principu geometrické podobnosti molekulárních útvarů, chápané jako topologická ekvivalence.

Operační jednotkou je „ekvivalent atomu“ TAE, (Transferable Atom Equivalent) ve smyslu Baderovy teorie atomů v molekulách. Pro TAE jsou počítány elektronové hustoty, ionizační potenciály, energetické parametry a další, indikující vztahy mezi reakčními centry eduktů. Na vyobrazení je reprodukován ionizační potenciál molekulového povrchu thiofenu jako jeden z mnoha parametrů, které jsou obsaženy v TAE.

A dodejme, že současná praxe zná také QSTR (Quantitative Structure-Toxicity Relationship) a využívá ho v projektování syntéz léčiv.



Od roku 1970 nabudila pozornost chemiků interesovaných v odvětví QSPR publikace

Carbó-Dorcy směřovaná na **molekulární kvantovou podobnost, MQS**. S rozvíjením jejích idejí a teoretického základu se etablovala grafově orientovaná QSPR; aplikací geometrie navazující na kvantovou mechaniku se rozšiřovaly její postuláty (Cioslowski, Fleischmann, Assessing molecular similarity from results of ab initio electronic structure calculations. *J. Am.Chem. Soc.* 1991, 113,64) a byly řešeny praktické problémy v řadě oblastí chemie. Po patnácti letech zakotvila MQS v QSPR a v QSAR v jejich teoriích i praktických aplikacích. To vedlo samozřejmě i k formulování nových teoretických vývodů a ke vzniku souvisících matematických a fyzikálních podkladů (Cooper, Allan, A novel approach to molecular similarity. *J. Comput-Aided Mol. Design* 1989, 3,253). Empirické QSPR modely přešly v korektně definované kvantově chemické formulace. Teorie DFT získala podobu kvantově mechanické funkce elektronové hustoty a hraje významnou roli v koncepci chemické podobnosti.

V publikaci Molecular quantum similarity using conceptual DFT descriptors (*J. Chem. Sci.*, 2005, 117, 425) popisují Bultinck a Carbó-Dorca užití základních deskriptorů elektronové hustoty $\rho(r)$, funkce tvaru $s(r)$, Fukuiho funkcí $f^+(r)$ a $f^-(r)$ a lokálních měkkostí $s^+(r)$, $s^-(r)$ k nacházení korelací mezi dvojicemi uvedených deskriptorů DFT a indexy podobnosti.

Lokální chemické potenciály $\mu_\alpha(r)$ vystihují příspěvky elektronových hustot jednotlivých partnerů subsystému (fragmentu, aim,...)

$$\mu_\alpha(r) = \mu_\alpha [\rho_\alpha, \rho_\beta; r] ;$$

Tvrdoost $\eta = (\delta_\mu / \delta_N)_{v(r)}$,

chemický potenciál $\mu = (\delta E / \delta N)_v$, a pokračujeme

$$\begin{aligned} \eta &= (\delta^2 E / \delta N^2)_v , \text{ také} \\ \mu &= -1/2(E^{\text{HOMO}} + E^{\text{LUMO}}), \\ \eta &= 1/2(E^{\text{HOMO}} - E^{\text{LUMO}}) \end{aligned}$$

a nově pro lokální chemické potenciály fragmentů

$$\mu_\alpha(r) = (\delta \epsilon_v [\rho_A, \rho_B] / \delta \rho_\alpha(r))_\rho .$$

Vodítkem je nám paradigma „Vlastnosti atomů determinují vlastnosti molekul“ (Hinze).

A vlastnosti atomů jsou trojího druhu: Primární, které můžeme a umíme stanovit přímo spektroskopicky, elektronové hustoty pomocí rentgenového záření, ty, které plynou z teoretických konceptů (atomové orbitály, valenční stavy atomů, skalární pole elektronové hustoty ρ , Laplacian $\nabla^2 \rho(r)$ a deskriptory založené na elektronové hustotě – chemický potenciál μ , elektronegativita χ , parametry tvrdosti η , Fukuiho funkce $f(r)$ apod.), a pak vlastnosti, které mohou být přisouzeny atomům z interakcí v molekulárních celcích (jako jsou experimentálně dostupné atomové poloměry, dipólové momenty μ ad.); z vyrovnání elektronegativit ve vazbách lze dospět k parciálním nábojům na atomech v molekule, od nich k molekulárním vlastnostem jako jsou ESCA data založená na měření vazebné energie emitovaných elektronů, protonové afinity a elektronové afinity ap.).

Zmíněný **Laplacian elektronové hustoty $\nabla^2 \rho(r)$** je dalším výborným deskriptorem reaktivity molekuly: vyznačuje v molekule oblasti lokálního zhuštění a zředění náboje. Je tedy vhodným kandidátem na vystižení míry podobnosti, která obráží reaktivitu molekuly a proto je označován jako index podobnosti v reaktivitě (RSI, Reactive Similarity Index. Index podobnosti, také

bezprostředně spjatý s reaktivitou, je založen na parametru lokální měkkosti $s(\vec{r}) = \left[\frac{\partial \rho(\vec{r})}{\partial \mu} \right]_{v(\vec{r})}$

$$R_{AB}^s = \frac{\int s_A(\vec{r}) s_B(\vec{r}) d\vec{r}}{\{[\int s_A^2(\vec{r}) d\vec{r}][\int s_B^2(\vec{r}) d\vec{r}]\}^{1/2}}$$

kde s_A a s_B jsou lokální měkkosti molekuly A nebo B.

Vztah pro lokální měkkost může být přepsán do tvaru

$$s(\vec{r}) = \left[\frac{\partial \rho(\vec{r})}{\partial N} \right]_{v(\vec{r})} \left[\frac{\partial N}{\partial \mu} \right]_{v(\vec{r})}$$

který je korespondentní se vztahem $s(\vec{r}) = f(\vec{r}) S$,

je-li S globální měkkost a f Fukuiho funkce. O přístupu, který přejímá dílčí formu funkcionálu a poskytuje „čísla“ o velikosti molekulární podobnosti, platí, že MS musí být vyvoditelná z vlnových funkcí srovnávaných molekul, musí mít zřetelnou fyzikální a matematickou interpretaci, a výpočet musí být rychlý.

Výpočty molekulární podobnosti mají neodmyslitelný význam při analýzách obsáhlých databází sloučenin v chemickém a farmaceutickém výzkumu. Molekuly jsou popsány binárními vektory s bity odpovídajícími přítomnosti nebo absenci podstatných strukturních znaků, takže jako míra podobnosti je vhodný Tanimotův koeficient případně chemická vzdálenost mezi dvěma molekulami. Pokud Tanimotův koeficient u některých malých molekul nedává uspokojivé odpovědi, může se uplatnit nový asociační koeficient uzpůsobený k vyrovnaní odchylek. Uvedený koeficient prezentovali Fligner, Verducci a Blower (2012) a testovali jej v databázi National Cancer Institute and Registry of Toxic Effects of Chemical Substances na obsáhlých kolekcích. Villet (2011) rozebírá asociační koeficient jako modifikaci Jaccardova-Tanimotova koeficientu.

Často zmiňovaný Tanimotův koeficient (index) vychází z poměru překryvu množin ke spojené množině a stal se mírou podobnosti. Je vyjádřen matematickým výrazem $T(a, b) = \frac{N_c}{N_a + N_b - N_c}$ v němž N představuje počet atributů v každém objektu (a, b)). C je symbol pro překryvovou množinu.

O kvantově molekulární podobnosti informuje spousta publikací. Zmíním například práci Molecular Quantum Similarity and the Fundamentals of QSAR známých autorů Besalú, Gironés, Amat a Carbó-Dorca (*Acc. Chem. Res.*, 2002, 35 (5), 289), nabízející obecný přehled o kvantové podobnosti a aplikacích v QSAR, o její koncepci vycházející z teoretických základů a o jejím prudkém rozvíjení zahrnujícím matematické formulace a míry podobnosti; samozřejmě s rozбором ukázkového řešení.

PODOBNOST V BCP PROSTORU

Rozložení elektronové hustoty $\rho(r)$, zkratkou také ED, představuje základní kvantitu, která určuje všechny chemické a fyzikální vlastnosti a jevy jako třeba intra- a intermolekulární síly, geometrické uspořádání molekul, elektronový potenciál, chemickou vazbu; další probíráme v kontextu teorie AIM. ED je kvantitou v DFT, a je dostupná postupy založenými na vlnových funkcích. Fyzikálně pozorovatelný fenomén lze měřit např. rtg difrakční metodou s vysokým rozlišením (zejména s novými zdroji synchrotronu třetí generace). Srovnání experimentálních poznatků s teoretickými výpočty ED umožňuje kvantová teorie atomu v molekule, QTAIM (Bader, Matta, *J.Phys.Chem. A* 2004, 108, 6661).

Topologických vlastností elektronové hustoty $\rho(r)$ využil Bader k formulování teorie molekulové struktury: Molekula je fragmentována až na atomy (skupiny) s povrchem o nulovém toku, splňující podmínku $\nabla \rho(r) \cdot n = 0$ pro každý bod – místo povrchu – subsystému $r \in S(r)$, kde n je jednotkový vektor ve směru normály k ploše. Body s povrchem o nulovém toku $\nabla \rho(r) = 0$ jsou kritické body. Kritické body na stopě vazby (bp), která popisuje trasu mezi vázanými atomy, a podél níž dosahuje elektronová hustota maxima vzhledem k bočnímu rozložení elektronů, jsou kritické body vazby (bcp). Elektronová hustota v bcp, $\rho(r)$, udává svým r polohu bcp. Atom v molekule, aim, je Baderem definován jako část prostoru vymezená povrchem $S(r)$ majícím vlastnost „nulového toku“; tím se myslí, že $S(r)$ se nekříží s žádným z gradientů vektorů $\rho(r)$.

Je-li k dispozici sofistikovaný popis molekuly v nějakém prostoru, dá se očekávat, že se najde několik způsobů měření podobnosti mezi molekulami. Baderem uvedený deskriptor může operovat s matematickými prostředky jako jsou ekvivalence, parciální uspořádání, proximity, teorie grafů a teorie grup. Popelier (*J.Phys.Chem. A* 1999, 103 (15), 2885) zavedl restrikcí na míru jednoduché Euklidovské vzdálenosti v prostoru BCP a vzdálenost d_{ij} mezi dvěma BCP i a j v 3-D prostoru definuje

$$d_{ij} = [(\rho_{b,i} - \rho_{b,j})^2 + (\nabla^2 \rho_{b,i} - \nabla^2 \rho_{b,j})^2 + (\epsilon_{b,i} - \epsilon_{b,j})^2]^{1/2} \quad (1)$$

Vzdálenost $d(A,B)$ mezi dvěma molekulami A a B je pak definována jako suma těchto BCP vzdáleností (rovnice 2)

$$d(A,B) = \sum_{i \in A} \sum_{j \in B} d_{ij} \quad (2)$$

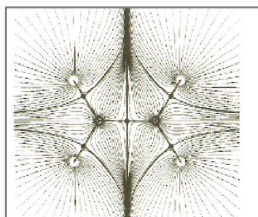
Čím je hodnota $d(A,B)$ nižší, tím jsou si obě srovnávané molekuly podobnější. Model řeší pouze vzdálenosti mezi dvěma korespondujícími BCP v molekule a a v molekule B. Vzdálenost definovaná výrazem (2) může být aktuálně modifikována, protože tři její složky vzdálenosti mají rozdílné dimenze. Obecnost a jednoduchost Popelierova modelu umožňuje predikce v systému QSAR.

K této pasáži může být užitečná glosa o Baderových kritických bodech vazby:

Elektronová hustota $\rho(r_c)$ je vypočtena pro vazebnou interakci v kritickém bodu vazby, (bcp), r , podél každé A-B stopy vazby (bp). Model kritického bodu vazby, bcp, je založen na topologických vlastnostech hustoty náboje $\rho(r)$. Vzorec molekuly je parcelován na fragmenty s povrchem s nulovým tokem splňující podmínku $\nabla \rho(r) \cdot n = 0$ pro každý bod plochy subsystému, když n je vektor ve směru normály k ploše. Body na plochách s nulovým tokem, v nichž $\nabla \rho(r) = 0$ jsou kritické body vazby, bcp. Každý bcp r_c může být označen podle svého zařazení a označení (l, s). Přiřazení l bcp odpovídá počtu nenulových vlastních hodnot Hessien matice $r(r_c)$, označení je dáno algebraickým součtem znaků vlastních hodnot (eigenvalues).

Kritický bod vazby (3, -1) má tři zakřivení, jedno kladné a dvě záporná a je to tedy sedlový bod v $\rho(r)$. Čili v bcp je elektronová hustota v minimu vzhledem k jejímu rozložení podél stopy vazby (bp, bond path) spojující v reálu dva vzájemně vázané atomy, a v maximu vzhledem k rozložení hustoty příčně ke stopě vazby (bp).

Kritický bod vazby (3, -1) najdeme mezi každou dvojicí vázaných atomů. O bcp byla publikována řada zjištění, která informují, třeba, o korelaci bcp s typem a řádem vazby, s elektronegativitami atomů a skupin apod.



Rozložení trajektorií končících v jádru. Množina trajektorií končí v daném jádru – atraktoru a pokrývá jámu atraktoru. Ve vyobrazení je zakreslena množina vazby cbp (jsou vyznačeny tečkami). Modelován je systém ethenu.

Popelierem odvozená nová míra podobnosti operující v abstraktním prostoru, ve kterém jsou vlastnosti soustředěny v kritických bodech vazby, umožňuje získávání relevantních informací z jejich *ab initio* vlnových funkcí. Praktickou aplikací metody jsou například výpočty LFER na principu definice vzdáleností mezi molekulami v BCP prostoru. Obecnost a využitelnost popsaného přístupu lze demonstrovat na vyvozování predikcí QSAR aplikovaných v lékařských oborech.

MOLEKULÁRNÍ KVANTOVÁ PODOBNOST A DESKRIPTORY DFT

Už v předchozím výběru deskriptorů podobnosti je napovězeno, že k jejich výpočtům je využívána řada základních parametrů jako jsou elektronová hustota $\rho(r)$, funkce tvaru $\sigma(r)$, Fukuiho funkce $f^+(r)$ a $f^-(r)$, lokální měkkost $s^+(r)$ a $s^-(r)$, i další funkce zahrnuté mezi deskriptory pole, například elektrostatický potenciál, sterická pole, hydrofobní pole ad., pole řešená konceptem DFT. Jsou sledovány korelace mezi indexy podobnosti pro každou dvojici užitých deskriptorů a jsou porovnávány informace poskytované jednotlivými modely.

Bultinck a Carbó-Dorca (*J.Chem.Sci.* 2005, 425,1217) vyhodnotili sílu aplikovaných indexů podobnosti, některé shledali redundantními a dospěli k významnému poznatku, že pro hodnocení porovnáváných molekul jsou nejvhodnějšími entitami popisu funkce elektronové hustoty a lokální měkkost, poskytující prakticky shodné informace jako Fukuiho funkce a funkce tvaru molekul (shape function).

Připomenutí: V abstraktní algebře představuje pole algebraickou strukturu složenou z množiny a na ní definovaných dvou operací; ty se chovají podobně jako sčítání a násobení na racionálních nebo reálných číslech. Daný stupeň abstrakce umožňuje studovat všechny podobné struktury současně. Definice pole v chemii: pole umožňuje vložení několika hodnot do jedné proměnné. Tyto hodnoty jsou přístupné pomocí indexu (zapisuje se do hranatých závorek); index může být celé číslo (v Pascalu) nebo řetězec (string).

Molekulární kvantová podobnost se soustřeďuje na kvantifikaci podobnosti dvou srovnávaných molekul A a B vyhodnocením měr **molekulárních kvantových podobností** (MQSM, Molecular quantum similarity measures)

$$Z_{AB} = \int F_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) F_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2,$$

je-li F_A deskriptor pole užitý pro molekulu A, který je v mnoha aplikacích korespondentní s elektronovou hustotou A. Parametr $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ se uplatňuje jako kladný přesně určený operátor podobně jako Coulombův operátor $(\mathbf{r}_1 - \mathbf{r}_2)^{-1}$, gravitační operátor $(\mathbf{r}_1 - \mathbf{r}_2)^{-2}$ a nejčastěji používaný operátor: $\delta(\mathbf{r}_1 - \mathbf{r}_2)$, Diracova delta funkce, která převádí MQSM předchozí rovnice v překryvovou míru mezi dvěma funkcemi,

$$Z_{AB} = \int F_A(\mathbf{r}_1) F_B(\mathbf{r}_1) d\mathbf{r}_1.$$

Tento MQSM používali Bultinck a Carbó-Dorca ve svých studiích, protože kvantová podobnost molekul A a B může být snadno kvantifikována pomocí Euklidovské vzdálenosti mezi vektory hustoty nebo základními vektory DFT,

$$d_{AB}^2 = Z_{AA} + Z_{BB} - 2Z_{AB}.$$

Z_{AA} a Z_{BB} jsou překryvové integrály nad deskriptorem prostorů a označují se jako míry sebepodobnosti. Jsou dobře korelovatelné s mnoha fyzikálněchemickými molekulárními vlastnostmi a jsou brány také za míru koncentrace hustoty elektronového náboje..

Jiný široce využívaný index podobnosti je Carbó-ův index C_{AB} , zobecněný kosinus s hodnotami mezi 0 a 1 (1 pro dokonalou podobnost).

$$C_{AB} = Z_{AB} / (Z_{AA} Z_{BB})^{1/2}.$$

Získané míry podobnosti se pak použijí k sestavení matice

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & \dots & Z_{1N} \\ \vdots & \ddots & \vdots \\ Z_{N1} & \dots & Z_{NN} \end{bmatrix}.$$

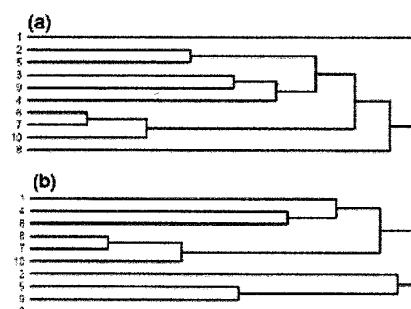
O této matici platí, že každý řádek a každý sloupec lze považovat za diskrétní reprezentaci odpovídajícího deskriptoru molekulárního prostoru v podprostoru, představujícím všechny prostory N molekul.

Tak každý i -tý sloupec Z , zapsaný z_i , uvádí diskrétní reprezentaci molekuly I v prostoru obsahujícím deskriptory prostoru N molekul. Sloupcové vektory vymezují reprezentace ∞ -rozměrného molekulárního prostoru na N rozměrnou reprezentaci vektoru, při čemž všechna čísla v daném vektoru jsou reálné, kladné definované deskriptory. Deskriptory vektorů jsou univerzální v tom smyslu, že je lze získat pro některou molekulu dané množiny.

Bultinck a Carbó-Dorca ve svých studiích doložili, že v koncepci DFT jsou elektronová hustota $\rho(r)$ a funkce tvaru molekuly entitami postačujícími k vystižení molekulární podobnosti a že poskytují informace shodné s funkcí obálky a Fukuiho funkce.

Pro sérii steroidů autoři připojili dendrogram, který v jiném kontextu bude uveden s komentářem.

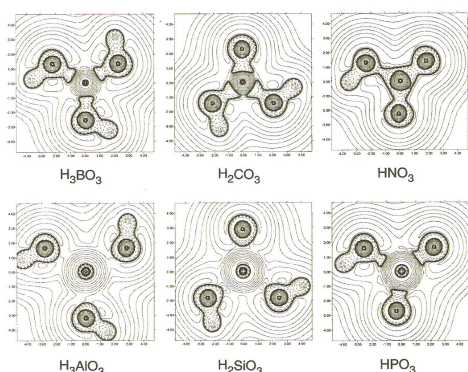
Dendrogram byl získán z Carbó-ova indexu založeného na funkci elektronové hustoty (a) a na lokální měklosti (b).



Ke zmínce o Laplacianu elektronové hustoty $L(r)$: vraťme se ve stručnosti k tomu, co o něm víme. Je jakousi stopou Hessovy matice. Jinak – je suma diagonálních prvků druhé derivace matice hustoty. My se zajímáme o Laplacian elektronové hustoty proto, že z podnětu Badera, který vypracoval koncepci atomu v molekule (AIM), poskytuje zajímavé informace o chemické vazbě. Zkusme je znovu rekapitulovat :

Molden řeší Laplacian elektronové hustoty souborem *ab initio* GamessUS/UK a Gaussian.

- V oblastech s negativními hodnotami je soustředěn záporný náboj a dominantní se stává energie potenciálu
- V oblastech s pozitivním Laplacianem je dominantní hustota kinetické energie a dochází k úbytku záporného náboje
- Laplacian pokrývá strukturu obálky atomu podle změny koncentrace náboje a úbytku náboje
- Chemická reakce odpovídá kombinaci koncentrace náboje ve valenční sféře donoru s úbytkem náboje ve valenční akceptoru elektronů.



Mapy Laplacianu rozložení elektronové hustoty vypočtené pro $H_{6-n}^{+} M^{+}O_3$ molekuly s $M = B, C, N, Al, Si, P$. Kation M je v centru každé molekuly, H atomy jsou vázány na aniony oxidů pro udržení elektroneutality. Plné linie reprezentují kladné kontury Laplacianu, přerušované spojnice znázorňují negativní úroveň $\nabla^2\rho(r)$. Intervaly konturových linií jsou větší a menší v úrovních od $\pm 2 \times 10^{-4}$, $\pm 4 \times 10^{-4}$, $\pm 8 \times 10^{-4}$ s počátkem u $n = -3$.

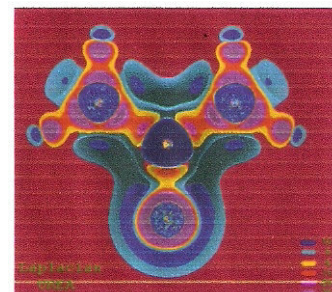
Mapování rozložení elektronové hustoty $\rho(r)$ vypočtené pro vazby M-O v řadě molekul oxidů bylo autorem (Gibbs 1994 ad.) nejprve provedeno pomocí jejich kritických bodů vazby, bcp. Výpočty byly v souladu také s hodnotami pro krystaly. Vzhledem k tomu, že v molekulách a krystalických formách se uplatňují typy vazeb od kovalentních, polárních až po iontové, soustředil Gibbsův autorský kolektiv pozornost na vyhodnocování rozložení elektronových hustot. Pro srovnání disponovali hodnotami Laplacianu distribucí valenčních elektronových hustot (Bader 1984; Kraka 1984) a studie zaměřili na porovnání těchto parametrů u oxidů a sulfidů. (Horní vyobrazení je převzato z publikace Gibbs ad., 1999).

Změny ve vlastnostech kritických bodů vazeb oxidů, sulfidů a nitridů vykázaly podobné a konsistentní trendy. Tyto poznatky mají význam pro posouzení vlivů funkčních skupin např. v zeolitech (používaných v organické syntéze) nebo pro predikce interakcí organických ligandů mikroorganismů na površích minerálů.

Fyzikální základ topologie Laplacianu elektronové hustoty hledejme v modelu VSEPR. Funkce je derivátem skaláru gradientu vektorového pole elektronové hustoty, veličina $\nabla^2\rho$ a udává, kde je elektronový náboj koncentrován, $\nabla^2\rho < 0$, a kde je snížen, $\nabla^2\rho > 0$; lokální koncentrace náboje umožňují mapování elektronových párů Lewisova a VSEPR modelů.

Na zobrazení Laplacianu elektronové hustoty močoviny snadno najdete karbonylovou skupinu a dvě NH_2 skupiny.

(Poznámka možná zbytečná: Teorie VSEPR (Valence Shell Electron Pair Repulsion) (Nyholm, Gillespie, 1957) umožňuje určit tvar neiontových molekul sloučenin nepřechodných prvků na základě odpuzování mezi elektronovými páry centrálního atomu. Tvar molekuly ovlivňují pouze valenční elektrony.)



FUZZY PODOBNOST

Model vytvořený na matematickém fundamentu je v podstatě založen na pravděpodobnostním usuzování. Modelem prolíná a posouvá jej do teorie chemické reaktivity

parametrizace výstupů a kvantifikace jejich řešení. Ještě jednou se vracíme k fuzzy podobnosti.

Podobnost dvojice funkcí může být vyjádřena určením kartézského součinu množiny objekt-funkce s tím, že tento součin množin je **fuzzy množina S**. Neostrost nad S se zavede do dichotomické situace tím, že o některém prvku x v S platí $0 < M(x) < 1$. Pak lze sestavit kartézský součin neostrých množin s dvojicemi funkcí členů $M(x)$, uplatňující se na uspořádaných dvojicích prvků množin. Míru podobnosti lze tedy vystihnout i v režimu fuzzy množin a fuzzy logiky.

(Zalewski, Krygowsky, Shorter, *Similarity Models in Organic chemistry, Biochemistry and Related Fields*. Elsevier 1991).

Míra $C(F_a, F_b) \approx 0$ přísluší tzv. marginálnímu členu. Protože každá funkce může být spjata korespondencí 1:1 s přesně definovanou kolekcí molekulových orbitalů, molekul, stavů, reakčních sérií ap., platí to i o míře C a o vytvoření uspořádání nad množinou funkcí. Důsledek – uspořádání množiny funkcí F indukuje analogické uspořádání množiny objektů S:

$F = \{f_1, f_2, \dots\}$ a $S = \{s_1, s_2, \dots\}$;

$K = [F \leftrightarrow S \parallel \{f_1 \leftrightarrow A_1, f_2 \leftrightarrow A_2, \dots\}]$;

je-li vyvozeno pravidlo uspořádání R v F, pak $R[F] = \{F_1 < F_2 < \dots\}$, a to indukuje $R[S] = \{s_1 < s_2 < \dots\}$.

Racionální jádro této koncepce se opírá o **fuzzy podobnost**. Mezeyovu myšlenku (1997) o paralele podobnosti s fuzzy ekvivalencí jste už v předchozích řádcích četli.

MQSM FRAGMENTŮ

Krátce o modelování elektronových efektů substituentů z LFER využitím **míry kvantové podobnosti fragmentů**:

MQS-SM (Molecular Quantum Self-Similarity Measure) představuje dílčí případ obecné formulace MQSM (Molecular Quantum Similarity Measures). Podle Gironese, Carbó-Dorcy a Ponce (*J.Chem.Inf.Comput.Sci.*, 2003, 43, 2033) je vhodný pro kvantitativní charakterizaci podobnosti mezi dvěma molekulárními strukturami A a B na základě srovnání jejich funkcí hustot. Obecná definice míry kvantové podobnosti molekul A a B,

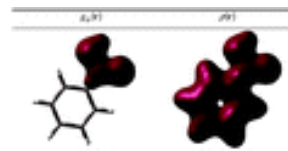
$Z_{AB}(\Omega) = \iint \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1, \mathbf{r}_2)\rho_B(\mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2$, když $\rho_A(\mathbf{r}_1)$ a $\rho_B(\mathbf{r}_2)$ jsou funkce hustoty (DF) molekul A a B a $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ je arbitrární (kladný) operátor. V jednoduchém případě, když $\Omega(\mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1, \mathbf{r}_2)$, se získá míra podobnosti překryvu (Overlap similarity measure)

$Z_{AA} = \int \rho_A(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r}$. Elektronová hustota fragmentu $\rho_A^X(\mathbf{r})$ je pak míra kvantové podobnosti fragmentu.

MQSM fragmentů odvodili autoři od partií Fermiho děr a nabídli tak nové výhodné teoretické deskriptory substitučních efektů z LFER

$$Z_{AA}^X = \int \rho_A^X(\mathbf{r})\rho_A^X(\mathbf{r})d\mathbf{r}$$

Srovnání kontur hustoty pro $g_\Omega(\mathbf{r})$, $\Omega = \text{COOH}$, a $p(\mathbf{r})$ na téže úrovni (0,03 e/au³).



Molekulární podobnost řeší Saggiomo, Kristova, Ludlow a Sijbren (*Journal of Systems Chemistry* 2013, 4 :2) aplikací termodynamicky kontrolovaných sítí. Metoda, založená na **dynamické kombinatorické chemii**, je určena především k využívání ve farmaceutickém a biochemickém výzkumu. K nacházení molekulární podobnosti v roztocích využívají dynamické molekulární sítě ve dvoustupňovém procesu. Shlukovou analýzou se stanoví diskriminační kapacita sítě. Klasifikační algoritmus je trénován ve schopnosti klasifikovat neznámé systémy. Dynamická molekulární síť byla testována na identifikaci aminů a amoniových iontů v množině 25 různých molekul. Upraveným způsobem mohl systém klasifikovat neznámé molekuly podle toho, jestli obsahují nebo neobsahují ethylaminovou skupinu. Studie se nabízí jako návod k použití na předpovídání bioaktivity sloučenin. Metoda patří k systémové chemii.

(Poznámka na okraj: Obecně je za systém považován integrovaný celek, sestávající z různých, interagujících, specializovaných struktur a subfunkcí.)

O kvantově molekulární podobnosti informuje spousta publikací. Zmíním například práci Molecular Quantum Similarity and the Fundamentals of QSAR známých autorů Besalú, Gironés, Amat a Carbó-Dorca (*Acc. Chem. Res.*, 2002, 35 (5), 289), nabízející obecný přehled o kvantové podobnosti a aplikacích v QSAR, o její koncepci vycházející z teoretických základů a o jejím prudkém rozvíjení zahrnujícím matematické formulace a míry podobnosti; samozřejmě s rozбором ukázkového řešení. Ze stovek publikací o chemické podobnosti v jejích různých podobách nemohu nezmínit jména Robert Ponec, David Hokzsa, Daniel Svozil se spolupracovníky a jejich Publications. V Chemických listech (1994) můžete nalézt informativní pojednání R. Ponce Podobnost v chemii.

Podobnost je obtížné definovat, ale lehce ji rozpoznáme, když ji vidíme. Pragmatická, praktická a vyhovující charakteristika podobnosti, kterou obecně a zejména v přírodních vědách uplatňujeme, staví na definici vzdálenosti.

Obvykle se pojem vymezuje buď definicí, která určí nutné specifické vlastnosti a znaky, nebo výčtem všech objektů, které pod pojem spadají. V logice podle Carnapa se tyto dva přístupy označují jako intenze pojmu (obsah nebo význam pojmu), a jako extenze (rozsah) pojmu, za kterou se považuje výčet objektů nebo podřízených pojmů. U Kanta (a vlastně už u Aristotela) se setkáváme s termíny *kategorie* a *nejobecnější pojmy*, které tvoří vrcholy různých hierarchií. Podřazený pojem je pak definován svým nadřazeným pojmem a druhovým rozdílem vzhledem k ostatním podřazeným pojmům.

Na předchozích stránkách, jak je uvádí název tématu, jsme za nejobecnější pojem označili podobnost. Navážeme na něj tématem zabývajícím se extenzí pojmu podobnost, tedy výčtem „podřízených“ pojmů, které v dané tématice podrobněji specifikují jejich intenzi. Jejich společným jmenovatelem je prefix **izo-**.