

Protože chemik se nepřetržitě – aniž by si to explicitně uvědomoval – v celé své činnosti pohybuje v některé části chemického univerza, známého jako koncepce chemického prostoru, je občas účelné připomenout si definici této kategorie. Jenže těch se najde v literatuře i v diskuzích mnoho. Metaforická je definice Lipinského a Hopkinse analogizující chemický prostor s kosmologickým univerzem: je bez hranic. Úzce vymezena je definice popisující chemický prostor jako množinu všech možných molekulárních struktur a jejich možných interakcí, limitovanou možinou deskriptorů a parametrů v definovaném podprostoru.

CHEMICKÝ PROSTOR

CHEMICKÝ PROSTOR

Co všechno je prostor

- Prostor matematiků
- Prostor psychologie
- Filozofické implikace
- Grupy a podobnost v prostoru molekul a jejich reakcí
- Klasifikace množin a podmnožin podle určených klíčů
- Mapování chemického prostoru

Analýza chemického prostoru

- Databáze chemického prostoru
- Systémy navigace v chemickém (pod)prostoru
- Metoda chemigrafie
- Klastrování

Sítě v chemickém prostoru

- Model umělé chemie
- Programy reprezentující biomolekulární aktivity
- Síť chemických reakcí v chemickém prostoru
- Sítě v chemickém a biologickém prostoru ve farmakologii

Navigace v chemickém prostoru

- ChemGPS-NP Web a metoda hlavní komponenty
- Analýza chemického prostoru
- Procházení chemického prostoru

Mapování chemického prostoru

- Matematický popis reakcí v termech teorie grafů
- QSAR v chemoinformaticce
- Grafy reakčních mechanismů, maticový a kódový přepis
- Alfanumerický kód ANCOD při shlukování podobných modelů reakčních mechanismů v chemickém prostoru
- Nacházení podobností mechanismů matematickým modelem
- Shlukování modelů reakčních mechanismů
- Databáze *in silico* virtuálních molekul

Molekulární prostor

- Metrické mapy prostoru
- Princip teorie nekonečného hierarchického navršování hmotného světa
- Řešení molekulárního prostoru teorií AIM

Princip dokonalé kauzality

Prostor je jakákoli množina. Obecně množina, jejíž studium je předmětem některého matematického oboru a která má některé výrazné rysy reálného fyzikálního prostoru. *Euklidovský prostor* je základním matematickým modelem fyzikálního prostoru. n -Rozměrným euklidovským prostorem nazýváme afinní prostor A_n , jehož zaměřením je n -rozměrný prostor vektorový V_n (lineární prostor) se skalárním součinem. n -Rozměrný euklidovský prostor značíme E_n , číslo n je jeho dimenzí. Neprázdnou množinu $E_k \in E_n$, která je euklidovským prostorem se zaměřením V_k ($V_k \subset V_n$) nazýváme k -rozměrným lineárním euklidovským podprostorem E_k v euklidovském prostoru E_n . Pro $k = 1$ dostáváme směr, pro $k = 2$ dvojsměr v E_n . Tím, že je pro zaměření V_n euklidovského prostoru E_n definován skalární součin, jsou do E_n zavedeny *metrické pojmy*, např. vzdálenost dvou bodů, *úhel dvou směrů* apod. E_n je metrickým prostorem, jehož metrika je definována vzdáleností dvou bodů. *Bod* v nejširším smyslu je prvek libovolné množiny, zejména je-li pro označení této množiny užito termínu prostor. *Afinní prostor* A_n je neprázdna množina abstraktních matematických objektů (bodů). *Vektor* v geometrii afinního A_n a euklidovského E_n prostoru je uspořádaná dvojice bodů afinního, resp. euklidovského prostoru.

CHEMICKÝ PROSTOR

Pro matematické myšlení je charakteristická abstrakce: její doménou je abstrakce potenciálního a aktuálního nekonečna. Potenciálním nekonečnem rozumí matematik předpoklad, že je možno každou veličinu zvětšit nebo zmenšit, zatímco aktuálním nekonečnem rozumí okolnost, že je možné vyšetřovat množiny zahrnující nekonečný počet objektů, o kterých se předpokládá, že existují současně aspoň v matematikově myslí.

V abstraktní teorii množin jsou studovány ty vlastnosti množin, které jsou důsledkem základního vztahu Cantorem uvedené teorie množin „být prvkem množiny“. V jejím rámci jsou studovány problémy související s určením počtu prvků nekonečné množiny a jde o ekvivalenci množin, kardinální číslo, operace s kardinálními čísly, spočetnost a mohutnost (každá nekonečná množina se dá prostě zobrazit na svou vlastní část).

Druhý směr, teorie bodových množin, zavádí mezi prvky další vztahy, analogické vztahům platným v množině reálných čísel (a představují-li reálné objekty, tedy i v množině reálných objektů. To je pro naše další úvahy a výpovědi důležitý moment.

Při uspořádávání kardinálních čísel dospěl Cantor k poznání, že je nezbytné vyšetřovat úplné uspořádání v množinách. Také to je pro studium chemie v prostoru významným úkolem.

Intenzivní studium topologických struktur, které zachycuje v podstatě abstraktní matematickou formulaci pojmů okolí, limity a spojitosti, vedlo k postupnému zpřesňování představ o prostoru v širším slova smyslu. Pojem prostor tak nabývá kvalitativně nový obsah.

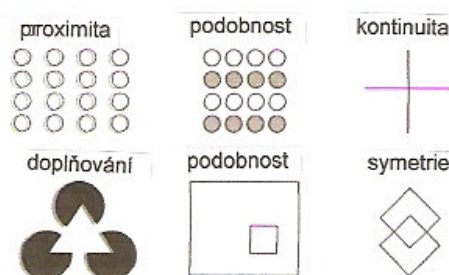
Tím novým prostorem je pro nás **chemický prostor**. Také **biologický prostor**. A dnes jsou oba tyto prostory neoddelitelné a prolínají se s prostorem supramolekulární chemie.

V *prostoru psychologie* je podobnost definována jako sblížení, blízkost (proximita) dvou myšlenkových představ. Kognitivní psychologie má pro koncepci podobnosti několik přístupů, ovšem každý je vztažen k určité partikulární soustavě předpokladů reprezentace znalostí.

Sheppard (1962) soudil, že myšlenkové představy názornosti se dají koncipovat jako jistý druh **myšlenkového prostoru**: koncepce jsou reprezentovány body („prvky“) v tomto prostoru a podobnost mezi nimi je funkcí vzdálenosti mezi jednotlivými body – koncepcemi. Jak velká je tato vzdálenost, umí řešit matematika. Tversky (1977) navrhl přístupy na základě vymezení mentálních vzdáleností daných znaků v určitém prostoru. Strukturní přístup (navržený Gentnerem a Markomanem, 1997) vychází z výzkumů analogie. Na typu myšlenkové reprezentace nezávislé jsou transformační přístupy (Hahn, Charter 2003), definující nejkratší množiny kroků nezbytných k určení vzdálenosti mezi srovnávanými systémy. A o koncepci podobnosti se zajímají i *sociálně psychologické přístupy* (shoda názorů, mezilidské vztahy, obratnost v komunikaci, demografické analýzy apod.).

„Gestalt“ principy formulovali psychologové ve dvacátých letech minulého století. V podstatě jde o vizuální zapracování prvků skupin (množin) do větších celků. Těch principů je uváděno šest a psychologové k nim dospěli na základě poznání, že lidské myslí je vpečetěna dispozice vkládat pozorované objekty do větších celků a ty do širších souvislostí.

Co si psychologové myslí o podstatě principů proximity, blízkosti, podobnosti, kontinuity, zcelování (closure), prostoru a symetrie, vypoví stručně skicy uvedených zákonů.



Zákon proximity: jsou-li percepce je seskupujeme jako objektu, a ty, které k nim objektu.

Zákon podobnosti: Objekty se stejnými atributy se snáze seskupují. Na skice si všimnete, že kroužky v horizontální poloze jsou si podobnější než ty v poloze vertikální. Podobnost má ovšem mnoho podob a stránek, proto jí bude věnována delší stať.

Zákon kontinuity: Naše oko vnímá konturu jako souvislý objekt; na schematu si spíše všimáme nejprve křížících se linií než toho, že úsečky se protínají nebo že dva pravé úhly mohou představovat vrchol grafu.

Zákon zcelování: Pozorovatel vyobrazení vnímá úplné, uzavřené obrazce, i když nejsou dokresleny. Nevadí, že bílý trojúhelník zakrývá jejich úplnost.

objekty stejné, při jejich by byly částí téhož nepatří, náležejí do jiného

Zákon podobnosti prostoru: Pokud se dva prvky pronikají jako na vyobrazení, pozorovatel to interpretují spíš jako malý čtverec umístěný ve velkém než že ve velkém čtverci je umístěn malý.

Zákon symetrie: Lidské oko preferuje interpretaci větší souměrnosti, tedy že na vyobrazení jsou dva překrývající se čtverce a až poté konstatuje, že vidí tři separátní čtyřúhelníky. Také o symetrii si uvedeme víc.

K termínu *Gestalt princip*: vypovídá, že (při zachování ostatních objektů) u částí posuzované oblasti, které se sobě podobají, sledujeme tendenci soudržnosti ve sjednocené formaci. WordNet® 3.0, © 2006 by Princeton University. *Gestalt* (termín převzatý z němčiny do češtiny i do angličtiny) znamená organizovaný celek, vnímaný jako více než suma jeho částí. Je překládán různě jako kompletnost, konfigurace, uspořádanost, esence, forma; Termín je užíván i v jiných disciplínách (Gestalt psychologie, v té se hovoří také o „sjednoceném celku“).

MNOHOROZMĚRNÉ PROSTORY VLASTNOSTÍ

Při seznamování s koncepcí chemického prostoru se setkáváme s mnohorozměrnými prostory vlastností, ve kterých jsou dimenze přiřazeny určitým, vybraným numerickým molekulárním vlastnostem deskriptorů. (Jako příklad můžete hledat Lipinského parametry „pravidla 5“.) Do těchto podprostorů vlastností jsou vřazeny molekuly (v této souvislosti označujeme „molekulou“ všechny entity – prvky množiny spadající do prostoru) mající souřadnice, které odpovídají hodnotám jejich vlastností; tak se formuje prostorové uspořádání molekul (tato myšlenka je připisována Pearlmanovi a Smithovi). Analýzou hlavní složky, PCA (Principal Component Analysis) se dospěje k projekci mnohorozměrného prostoru vlastností do prostorů s menší dimenzí (2D a 3D), které jsou nám vizuálně dostupné.

Poznámka: Lipinského pravidlo je s úspěchem aplikováno v oblasti syntéz léčiv a zájemce je může hledat ve farmakoinformaticce.

O chemické podobnosti, molekulární podobnosti, podobnosti chemických reakcí a s nimi spjatými fenomény izomerií, symetrií, periodicitou bylo napsáno doslova na sta monografií, na sta zasvěcených publikací a přílehlých sdělení, takže esej nazvaná Poznámky k tématu Podobnost v chemii by už nemusela přimět čtenáře k její četbě.

Pokud ano, začneme – a jak jinak – pátráním ve starých časech. Ve filozofii starých Řeků se Příroda řídí principem homogenity. Tato homogenita (zatím řecký základ slova nepřekládáme) je posuzována a studována ze dvou hledisek: je jednotou materie, látky, a jednotou struktury látky, objektů, věcí. Pythagoras se, jak se o něm obecně ví, shlédl v číslu, ve struktuře. Dnes je samozřejmé, že při popisu struktur, forem a jejich vztahů operujeme s číselnými parametry a můžeme přitakat rovněž na Pythagorovu myšlenku „Číslo je kvantita a kvantita je forma a forma je kvalita“ (Gow 1923, Russell 1939 ad.). Pilířem tohoto usuzování je Mendělejevův periodický zákon prvků a na něj navazují matematické teorie s aplikacemi v teorii informace, teorie množin, parciálně uspořádaných množin, fuzzy množin, teorie deterministického chaosu, teorie grafů, topologie a to nejen v chemii a ve fyzice, ale také v řadě dalších disciplín. (Matematická hlediska periodického zákona řešili podrobně např. Restrepo a Pachón a co jste právě čtli, jsou v podstatě jejich vývody.)

G.K. Vemulapalli dokládá (v *Philosophy of Chemistry*, Baier D., Scerri E., Lee McIntyre. Vyd. Springer Verlag 2006) názor, že ve druhé polovině minulého století došlo k jedinečnému rozvoji vědeckého poznání. Jedinci pak nezbyvá než se seznamovat s malými částmi vědy a to ještě s „ptačí perspektivou“ a při tom dbát, aby neztrácel alespoň povšechné kontakty s ostatními oblastmi poznání. Jeho srovnání s impresionistickým obrazem krajiny není až tak dalece přehnané. Vemulapalliho láká hledání filozofických implikací a vychází z toho, že vlastnosti symetrie objektu nejsou vzájemně nezávislé – jsou vzájemně spjaté. Soubor symetrií objektu představuje grupu. V matematice se takovými soubory zabývá teorie grup. Filozofové diskutují symetrii (a teorii grup) především

v systémech řešených kvantovou mechanikou. Rosen (1995) definuje symetrii jako „immunity to a possible change“.

(A protože poznatky narůstají takřka geometrickou řadou, hledám v tom neúprosném soupeření lidského poznání s nevyčerpatelnými tajemstvími Přírody pro sebe ospravedlnění publikování tohoto textu – že není jen nošením dříví do lesa – nebo je aspoň jeho přeskládáním?)

Čtenáře určitě zaujme sdělení D. Svozila z Laboratoře informatiky a chemie VŠCHT v Praze nazvané *Welcome to Chemical Space* (2011). Volně reprodukováná hesla charakterizují sdělnou informaci

- Všechny možné organické sloučeniny představují „chemický prostor“ ; k tomu bych dodal, že všechny chemické entity (atomy, ionty, radikály, fragmenty, molekuly, makromolekuly, supramolekuly...) a všechny jejich reakce spadají do chemického prostoru.
- Není od věci analogie mezi kosmologickým Univerzem a chemickým prostorem v existenci bez hranic, ani srovnání chemických sloučenin s hvězdami ve srovnávaných prostorech.
- Srovnání velikostí: odhadovaná velikost chemického prostoru je 10^{100} - 10^{200} (SciFinder $\sim 6 \times 10^7$). V pozorovaném Vesmíru může být kolem sextilionů (10^{21}) hvězd. Můžeme porovnat: Může být až 10^9 možných derivátů *n*-hexanu. Pro obsáhnutí člověkem je chemický prostor nekonečný.
- Ne všechny teoreticky předpokládané sloučeniny spadají do mezí chemicky syntetizovatelných látek.

Je-li chemický prostor ve své nekonečnosti těžko představitelný podobně jako je tomu s nekonečným Univerzem, jehož jsme „prvkem“, bodem ve vesmírném prostoru – a on takový je (Douglas Adams odhadl počet možných organických molekul v chemickém prostoru na více jak 10^{60} ...), je jen jedna možnost, jak se v něm „vyznat“. S vynucenou skrovností pátrat v jeho subjektivně volených podprostorech, o které máme zájem. Koncepce „chemický prostor“ předpokládá reprezentaci podobnou geografickým mapám, abychom mohli ilustrovat rozložení molekul a jejich vlastností. Proto nejprve vytváříme prostory společných vlastností, které nám umožní formovat serie, grupy, shluky, podprostory, třídy (z uvedených určení si vyberte) na základě molekulárních deskriptorů.

Vyhledávání adeptů pro vytváření množin – grup v chemickém prostoru a v prostoru chemických reakcí nám pomáhá nesporně hledání a nacházení chemické podobnosti a reakční podobnosti a izomerie. To první, co chemik zaregistruje v synoptickém pohledu, je příslušnost posuzovaného systému, formovaného kolem výrazného společného znaku, v části podprostorů.

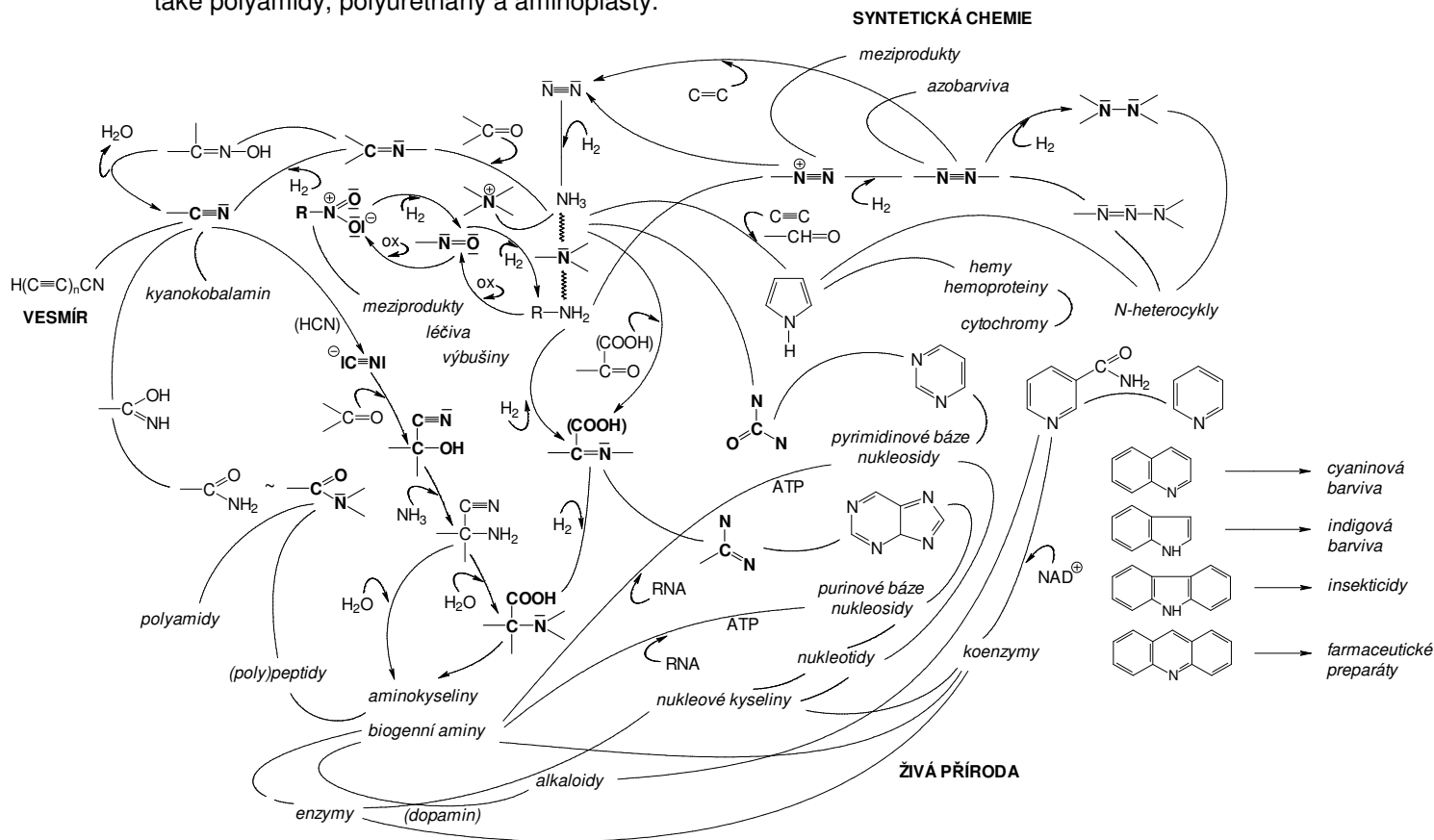
Příroda to zařídila tak, že ze stovky chemických prvků upřednostňuje pro životní procesy pozemských tvorů jen několik desítek prvků a těch, kterým říkáme organogenní a biogenní prvky, je sympaticky malá podmnožina. Mezi nimi zaujímají bezesporu základní postavení atomy dusíku, vbudované do rozlehlého chemického podprostoru.

Dusík je vestavěn do molekul svých sloučenin některými ze způsobů uspořádání valenčních elektronů, uvedených v grafu VS a jejich konverzí, které probíhají přes pestrou množinu dalších nestabilních vazebných situací. Dusík vytváří vazby σ i π se širokým okruhem dalších prvků, také sám se sebou, a pro život na Zemi má mimořádný význam zapojování atomů dusíku do vodíkových vazeb.

Substruktury molekul představované atomy C a N mohou být (a jsou) doplňovány různým okolím a jsou společné řadě funkčních skupin (např. $-\text{NH}_2$, $-\text{NHNO}$, $-\text{NCO}$, $=\text{NNH}_2$ atd.) i vyšších strukturních celků (pyrrolidinů, pyridinů, aldiminů apod.). Toto sjednocování nám usnadní poznání organických dusíkatých sloučenin v jednotlivostech i v celku a umožní nám aplikace modelů řešení jejich reakcí na analogické, přesto něčím si podobné i na první pohled různé systémy.

Atomy dusíku ve strukturně rozmanitých sloučeninách nacházíme v mezihvězdném prostoru, v živých organismech a samozřejmě v reaktantech preparativní chemie. V biomedicině a v biomateriálech se intenzívně rozvíjí proteinová chemie: předpokladem pochopení vztahů mezi sekvencí proteinů, jejich

strukturou a funkcí je poznání na úrovni reakčních mechanismů. Pro poznání 3-D struktury proteinů a hledání nových center pro vázání kovů je třeba znát redoxní vlastnosti systémů a hodnotit reaktivitu metodami osvědčenými při studiu malých molekul a jednodušších reakčních soustav. Poznatky jsou použitelné při aplikacích na makromolekuly biogenních aminů, proteinů, metaloproteinů, enzymů a koenzymů a nukleových kyselin. Z aminokyselin vytvářejí rostliny alkaloidy a syntetická chemie to umí také. Atomy dusíku jsou v molekulách některých antibiotik a řady dalších léčiv: jejich syntézy patří k mistrovským výkonům chemiků. Makromolekulami obsahujícími charakteristické *N*-substruktury jsou také polyamidy, polyurethany a aminoplasty.

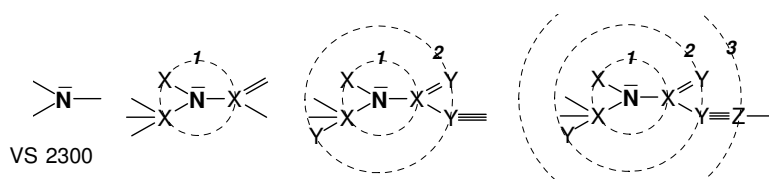


Ukázka výseče chemického prostoru osazeného dusíkatými sloučeninami. Každé heslo představuje svůj vlastní svět, svůj bohatě rozprostraněný prostor, do kterého spadá široká množina jednotlivých, něčím sobě podobných metaforických „molekul“.

Mapování chemického prostoru

Při **mapování chemického prostoru** se soustřeďuje chemik na soubory molekul, které jsou si v něčem podobné. Ono „něco“ je vždy nejmenší společný znak – a můžeme tímto postupem dospět až k atomu prvku, který je oním majákem, který obhlédává požadovanou množinu nebo dokonce třídu sloučenin v nedozírném oceánu chemického prostoru.

K modelování postupu zvolme oblast dusíkatých sloučenin, dost širokou a významnou, abychom v ní mohli posléze dospět k cílové struktuře se signifikantním atomem dusíku. V tom chemikovi pomáhá určující znak – valenční stav atomu. Můžeme si vybrat VS_N 2300 a naznačit odvozování reprezentantů nabalovatelných kolem centra v prvním, druhém,... *n*-tém okolí,

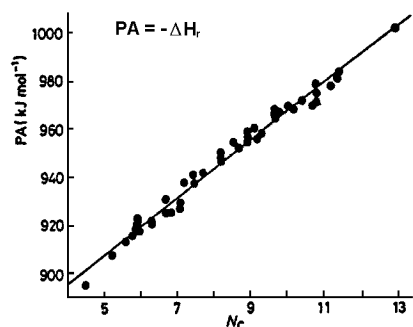


Úlohou mapování chemického prostoru je porozumění vztahům mezi strukturami malých molekul a jejich vlastnostmi a to až po jejich fenotypové efekty v živých systémech. Prostředkem je mapování chemického prostoru využívající molekulární deskriptory.

Těmi běžně aplikovanými deskriptory jsou především známé parametry molekul a atomů v nich (aim)

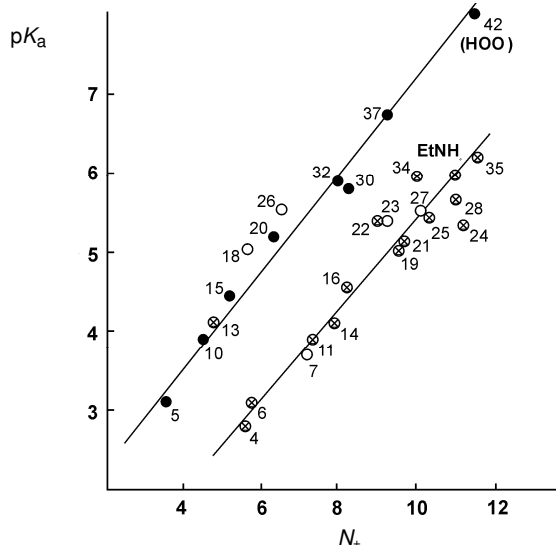
jakými jsou m.j. ionizační potenciály IP a elektronové afinity EA , polarizovatelnost α , protonové afinity PA , nukleofilicita N_+ , samozřejmě elektronegativity χ , dipólové momenty μ , pochopitelně tvrdost η a vůbec všechny korelovatelné veličiny, protože poskytují náhled na zastoupení (co největšího) počtu jednotlivých molekul dané třídy (souboru) v prohledávaném podprostoru.

Na ukázkou můžeme vyhledat třeba korelace protonových afinit alifatických aminů s hodnotami efektivní polarizovatelnosti (upravené číslem konektivity N_c).



K vytváření map chemických prostorů využívají chemici řadu technik a prostředků. Nejběžnější jsou například Vennovy diagramy.

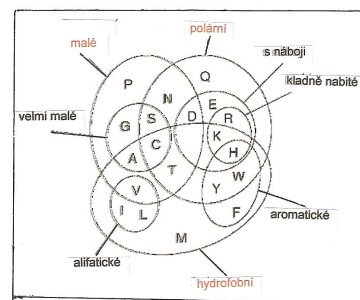
Nebo existuje závislost parametru N_+ na pK_a aminů; ta dospívá dále, separuje aminy primární a sekundární $\odot$ od terciárních $\circ$ a nukleofilů $\bullet$ s α efektem



Tak klasifikaci aminokyselin v chemickém prostoru řešil už v roce 1986 Taylor (*J.Theor.Biol.* 1986,119, 205). Algoritmus jím použitého programu PROMIS v jazyku PROLOG je založen na stochastické technice vyhledávání. Pravidla vyvozovaná z výchozí množiny obecných pravidel byla získávána zobecňováním, specializací a extenzí. Taylorova klasifikace aminokyselin je využitelná k popisu podobnosti mezi fragmenty uplatněnými v klasifikaci sekvencí. Dospívá se k sestavení zobecňujících zobrazení tvořenými sekvencí identifikátorů tříd (ALL, SMALL, CHARGED, HYDROPHOBIC, AROMATIC, ALIPHATIC, POSITIVE, POLAR; další podklasifikace je samozřejmě možná).

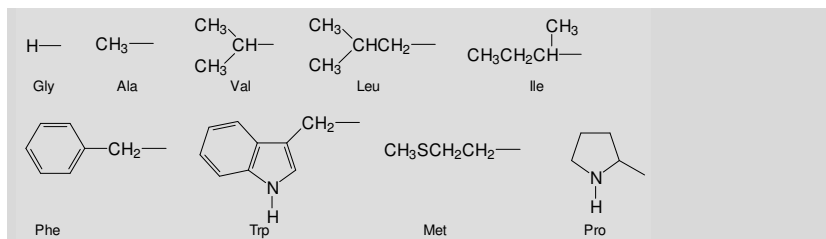
Vennův diagram zobrazuje vztahy tříd dvaceti geneticky kódovaných aminokyselin podle Taylora. Tato seskupení podmnožin umožňují úspěšnou aplikaci nalézání zobecněných obrazců v sekvencích aminokyselin.

Univerzum **ALL** [A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y]
 množina **SMALL** [A,C,D,G,N,P,S,T,V]
 podmnožina VERY SMALL [A,C,G,S]
 množina **HYDROPHOBIC** [A,C,F,G,H,I,K,L,M,T,V,W,Y]
 podmnožina AROMATIC [F,H,W,Y]
 podmnožina ALIPHATIC [I,L,V]
 množina **POLAR** [C,D,E,H,N,R,S,T,Q,W,Y]
 podmnožina POSITIVE [H,K,R]
 podmnožina CHARGED [D,E,H,K,R]

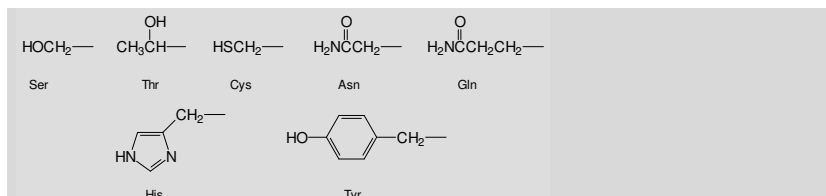


Ještě si připomeneme, že elektronové a sterické faktory v aminokyselinách spoluurčují substituenty na $-COOH$. Které to jsou, uvedeme s jejich klasifikací: určujícím znakem je zařazení aminokyselin k neutrálním, kyselým a bazickým systémům. Pak

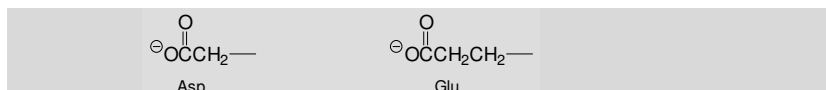
nepolární skupiny jsou



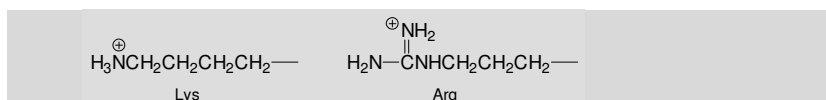
neutrální při pH 7



se záporným nábojem při pH 7



s kladným nábojem při pH 7



Každou množinu a podmnožinu lze klasifikovat podle určených klíčů dále na konkrétní prvky. Jedním takovým již zmíněným užitečným parametrem je nepochybně tvrdost η . Jsou ovšem další:

Základní molekulární entitou atomu v molekule je valenční stav atomu, VSA. Valenční stav atomu popsáný (Ruedbergem) jako atom v molekule, AIM, zbavený všech vlivů interferencí, k nimž dochází při vzniku vazeb, zmrazený na hodnotách, které měl v molekulárním celku, může být svou podstatou dalším klasifikačním parametrem.

Dva atomy v molekule se separují až do valenčních stavů se zmrazenými AO; z enegetického hlediska je to přesně výsledek, k němuž se dospěje jednodeterminantovou SCF Harthreeho-Fockovou metodou. V modelu zpracovaném aparátem teorie grafů je atom každého prvku A_i a pro atomový vektor A_i a A_j popsán takto:

- symbolem prvku pro vrchol V
- vlastním číselným indexem atomu příslušejícího do množiny A , získaným kanonickým indexováním strukturního vzorce (např. algoritmem CANON ad.)
- protonovým číslem Z atomu A_i příp. atomů A_i a A_j
- počtem valenčních elektronů, které jsou popsány
- lokalizovanými valenčními elektrony
- volnými elektrony n
- počtem delokalizovaných elektronů v systémech obsahujících atomy A_i a A_j
- počtem atomů ve všech delokalizovaných systémech s A_i
- formálním elektrickým nábojem, který je
- lokalizovaný q_i
- delokalizovaný ve všech delokalizovaných systémech
- elektronovou hustotou $\rho(r)$ na atomech A_i a A_j
- počtem atomů posuzovaných jako virtuální vrcholy molekulového grafu, které nejsou zahrnuty v EM a nemají proto vlastní deskriptory
- množinou sousedních atomů vázaných na atomu A_i kovalentními a nekovalentními vazbami (včetně vodíkových vazeb), danou koordinačním číslem atomu
- valenčním stavem VSA popsáným numerickými atomovými vektory.

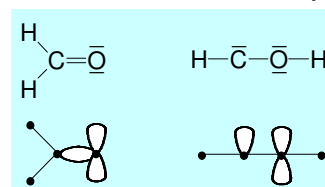
O atomu v molekule v modelu AIM je třeba znát:

• **valenční stav atomu, VS**

- hybridní stav
- energii DE
- chemický potenciál μ
- elektronovou hustotu $\rho(r)$ a její Laplacian $\nabla^2\rho(r)$
- atomový (kovalentní), iontový a orbitalový poloměr r_i, R_i
- náboje na atomech q_i, Q_i
- ionizační potenciál IP a elektronovou afinitu EA
- elektronegativitu χ
- polarizovatelnost ϵ
- tvrdost η / měkkost s
- Fukuiho funkci $f(r)$

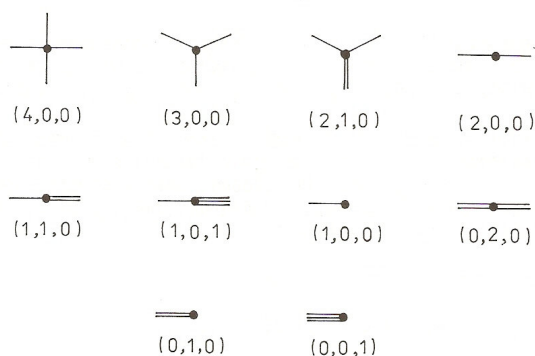
Vzhledem k tomu, že v každé "molekule", kterou hledáme v chemickém (pod)prostoru, se postupným osekáváním struktury dospívá k finální entitě, k atomovému vektoru, je pochopitelné, že ta se může brát jako klíč k mapování **chemického prostoru**. S chemickým strukturním vzorcem pracujeme jako s grafem, jímž skutečně pro matematika je, a jednou z operací nad grafy je operace rozkladu grafu. Postupně jsou odstraňovány marginální vrcholy, až se dospěje k dvojici vázaných atomů, která představuje reakční centrum, a v ní k dominantnímu atomu s jednoznačným valenčním stavem, VSA.

Formální grafová struktura „molekulový graf“ je v jednoduchém vztahu se strukturním vzorcem molekuly, hrany jsou přiřazené vazbám, přičemž násobnost hran vyjadřuje násobnost vazeb, smyčky jsou přiřazené volným elektronovým pářům. Tato jednoduchá korespondence mezi molekulovým grafem a strukturním vzorcem platí jen pro tzv. ohraničenou chemii, která se zabývá molekulami se sudým počtem elektronů v singletovém základním stavu.



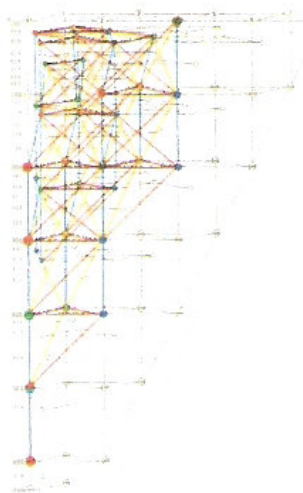
Přiřazení mezi strukturním vzorcem a molekulovým grafem.

Nad takto definovaným pojmem molekulového grafu můžeme řešit standardní úlohy teorie grafů, které jsou významné pro matematickou chemii a chemickou informatiku. Kvasnička s Pospíchálem formulovali teorém pro molekulové grafy: Vrcholy molekulových grafů ohodnotíme valenčními stavy (VS). Pro běžné prvky je třeba jen několika málo VS; máme-li konstruovat celou množinu grafů přiřazených dané sekvenci VS, je třeba znát podmínku, za které je splněn grafický sled procesů. Jmenovaní autoři řešili problém konstruktivní enumerace MG. Nejjednodušší se jeví specifikace molekulárních grafů počtem vrcholů a hran, i když není úplná. Vhodnější je použít nejen počet vrcholů, ale také některé subgrafy, molekulární fragmenty. Autoři zavedli označení technika superatomů (supervrcholů) a formulovali teorém grafické sekvence valenčních stavů (1989). Řešením dospěli autoři k deseti valenčním stavům, které označili jako základní valenční stavy (Target VS):

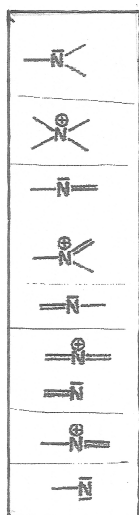


Soubor všech vybraných valenčních stavů pro molekulární grafy. U každého grafického znázornění jsou v závorkách uvedeny uspořádané trojice nezáporných celých čísel. K tomu tato ukázka.

Stabilní valenční stavy atomů dusíku je možno získat z grafu konverzí valenčních stavů atomu VSA_N



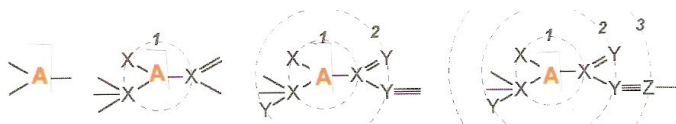
(Ilustrativní graf elementárních konverzí atomů dusíku; stabilním stavům odpovídají zeleně vybarvené vrcholy. Jsou to tyto)



V sousední minibance je naznačeno vytváření shluků struktur případně sloučenin majících společné superatomy, které lokalizujeme do mapy prohledávaného chemického prostoru.

Příklady (sub)struktur případně celistvých molekul jsou uspořádány na principu proximity a virtuální vrcholy (neobsazené valenční čárky) naznačují možnosti rozvíjení daných systémů do šíře. Každá množina v analyzovaném prostoru je obsaditelná prvky, určenými jednou podmínkou: společným valenčním stavem atomu **A** určeného za potenciální reakční centrum, v teorii grafů za kořen stromu, za dominantní složku substruktur a v klasifikačních schématech za určující znak třídy sloučenin (poslední sloupec v tabulce).

$\text{—}\ddot{\text{N}}\text{—}$	$\text{C—}\ddot{\text{N}}\text{—}$		alifatické aminy 1°, 2°, 3°
			enaminy
			aromatické aminy 1°, 2°, 3°
			hydroxylaminy
			O-alkyl, N-alkyl
			hydraziny
			hydrazolátka
			aziridiny
			azetidin(y)
			pyrrolidin(y)
			piperidin(y)
			diaziridin(y)
			pyrrol(y) (N-alkyl-, N-aryl- ...)
			N-oxidy azaaromátů
			N-nitrosamin(y)
			N-nitramin(y)
			amoniové soli protických kyselin
			amoniové soli Lewisových kyselin
			kvarterní amoniové soli (dtto v aromatické řadě)
			isokyanidy
			nitrososloučeniny
			nitroslooučeniny
			diazohydroxid
			azosloučeniny
			azidosloučeniny, azidy
			triazeny
			isokyanáty
			karbodiimidy
			iminy
			isonitrososloučeniny, oximy
			hydrazony
			amidiny
			aziny
			keteniminy
			pyridin(y)
			pyridazin(y)
			isokyanáty
			isothiokyanáty
			iminiové soli
			nitrony
			diazoalkany
			nitrily, kyanidy
			kyanáty
			kyanamidy
			nitrilový ion
			diazoniový ion
			alifatické diazoslooučeniny
			azidy
			nitreny



Koncepce chemického prostoru v souvislosti s možností jeho mapování při studiu obsáhlých bazí dat existuje nejenom v představách chemiků, biochemiků, bioorganických chemiků a fyziků a chemiků zabývajících se chemií materiálů. Tu jsou ukázky:

Reymond ad. (2012) si položili v souvislosti s konstruováním databáze chemického univerza GDB (Chemical Universe Database) otázku, kolik molekul odpovídajících jednoduchým chemickým pravidlům a syntetické dostupnosti je v chemickém prostoru možných. Odpověď má velkou vypovídací hodnotu při hledání nových léčiv a projekt GDB může zahrnovat také extenzi na větší molekuly a na syntézy sloučenin namodelovaných GDB. (Reymond, Ruddigkeit, Blum, van Deursen, *WIREs Comput. Mol. Sci.* 2012, 2, 717).

Nástroj k exploraci chemického prostoru mezi startovými a cílovými molekulami navrhli Hoksza a Svozil a nazvali jej Molpher (v BIBE, IEEE, 2011, 201 jej najdete s názvem *Exploration of chemical space by molecular morphing*). Program je určen k vyhledávání cest (hran) mezi dvěma molekulami (vrcholy) v chemickém prostoru (když morphing si přeložíme jako změnu ve tvaru oblasti dosahovanou přímo posouváním vrcholů jednotlivých elementů).

Procházení chemického prostoru je technika umožňující identifikovat v prostoru všech chemických sloučenin cestu, která obsahuje kandidáty na látky mající podobnou aktivitu jako start (S) a cíl (C) této cesty. Motivace problému je taková, že molekuly S a C mají podobnou funkci, tj. vážou se do podobného místa na povrchu proteinu a mají tudíž podobnou strukturu. Molpher pak prochází chemický prostor mezi S a C a hledá posloupnost strukturních modifikací definující cestu z S do C. Mnoho molekul na této cestě ovšem nemusí mít požadovanou funkci, tj. nemusí být schopny vázat se na dané místo v proteinu. Start (S) a cíl (C) tvoří vstup algoritmu Molpher vyvinutý v rámci softwarového projektu na MFF UK, který je schopen najít cestu v chemickém prostoru. Tato cesta ovšem obsahuje i molekuly, které nemusí být syntetizovatelné. Cílem práce je tedy implementace algoritmu, který by byl schopen odhadnout syntetizovatelnost molekuly. Předpokládá se algoritmus založený na identifikaci často se vyskytujících se fragmentů ve stávajících databázích chemických sloučenin a odvození syntetizovatelnosti sloučeniny z této informace (Hoksza 2011)

Algoritmus je schopen najít cestu v chemickém prostoru, ovšem řídí se pouze strukturní informací. Cílem citované práce je tedy nastudovat stávající metody dockingu, vyhodnotit je z hlediska časové složitosti a některou (některé) z nich naimplementovat do algoritmu Molpher a sledovat vliv aplikace dockingu na nalezenou cestu. (Hoksza 2009)

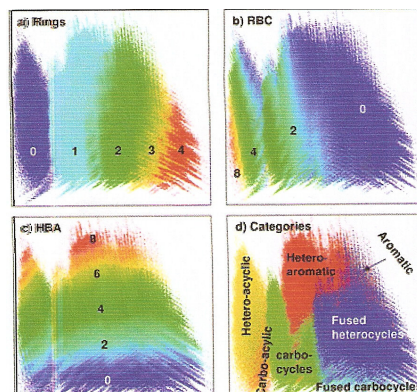
Jen poznámka k počítačové reprezentaci molekulárních struktur: 1. Založené na grafech jsou uzpůsobeny uchovávat informace dostupné v 2D i 3D strukturních vzorcích. (Při prohledávání souborů dat pro studium podobnosti jsou počítačově náročné.) 2. Reprezentace otisku prstů, reprezentující strukturní vzorce. Dané polohy v řetězci bitů odpovídají specifickým znakům (molekulárním fragmentům nebo fyzikálně chemickým vlastnostem). Příklad je uveden v tématu Podobnost. Protože tento způsob znázorňování prohledávaných systémů je výhodný a efektivní, vyvinuli chemici už několik typů programů (MACCS (MDL keys), ECFP ad.

Analýza chemického prostoru

Je v možnostech chemika obsáhnout chemický prostor? Je, pokud si z nepředstavitelného prostoru vymezí jistý podprostor. Autoři Reymond, van Deursen, Blum a Ruddigkeit (*MedChemComm.* 2010) publikovali svoje zkušenosti s realizací možností vytváření virtuálních knihoven molekul, mapování využitím vizualizace na základě analýzy hlavní komponenty (PCA) nebo samoorganizujících map a strukturní analýzy MQN

systému. V databázi chemického prostoru (GDB) se systém prodírá pomocí algoritmů k procházení definovaných regionů a současně generuje zvolené virtuální knihovny. Studie je zaměřena na molekulární systémy a záměrně také na hledání molekul nových léků.

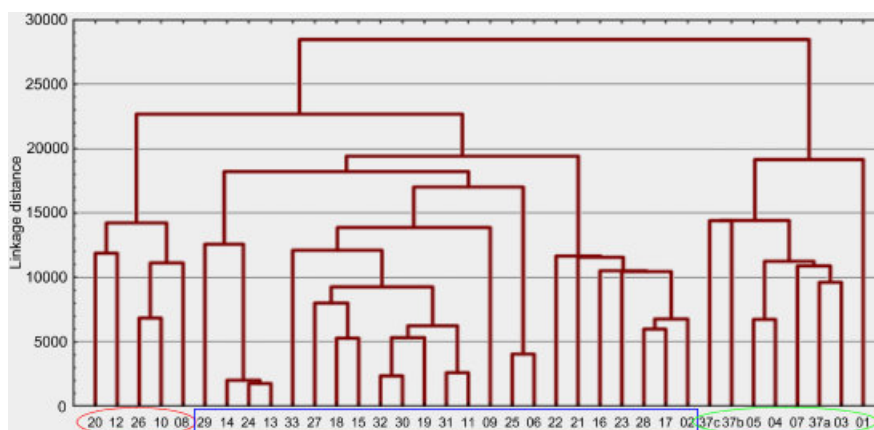
V mapě MQN prostoru GDB-11 je označen (a) počet cyklů v molekule, (b) počet vazeb s volnou otáčivostí, (c) počet atomů – akceptorů H-vazeb, (d) kategorie molekul. V části (d) jsou podle pravidel priorit uvedeny: 1. Heteroaromatické (červené) > 2. aromatické (fialové) > 3. kondenzované aromatické (modré) > 4. kondenzované karbocyklické (modrofialová) > 5. heterocyklické (zelená) > 6. karbocyklické (oranžová) > 7. heterocyklické (žlutá) > 8. karboacyklické (oranžová). Každý bod v mapě je zbarven podle majoritní kategorie sloučenin seskupených kolem tohoto bodu se šedým ovroubením (saturace v HSL stupnici), ukazující na čistotu kategorie).



Jak se může chemik orientovat v nedozírném chemickém prostoru. Vždyť i v univerzu vymezených okruhů sloučenin je nesnadné dávat k sobě, zvláště při hledání nových sloučenin ve farmaceutickém výzkumu nebo v supramolekulární a nanochemii, podobné systémy. Oprea a Gottfries (*J. Comb. Chem.* 2001, 3, 157) rozpracovali systém navigace v chemickém prostoru (*The Art of Navigating in Chemical Space*) na principu kombinatorické chemie v podprostoru, který nazvali *drugspace*. Na tomto základě lze sestavovat mapy prostorů léků v připodobnění autorů ke geografickým souřadnicím. Vybraná pravidla obsahují velikost molekul, lipofilicitu, polarizovatelnost, elektronovou hustotu, flexibilitu, rigiditu a schopnost vytvářet H-vazby (připomínáte si parametry QSAR?). Byly syntetizovány kombinatorické knihovny řádově s 10^2 - 10^9 strukturami. Globální mapy prostoru léků, GPS (Global Positioning System), byly extrahovány metodou hlavní komponenty (PCA) ze 72 deskriptorů. Struktury ležící mimo, satelity, nejsou zamítány, protože mohou obsahovat substrukтуры vhodné pro zkoumané syntézy. Testovací soupravou bylo m.j. 20 aminokyselin. ChemGPS je využitelná také v agrochemii, polymerní chemii apod., všude, kde se uplatní standardní chemometrické metriky. Autoři nazvali jimi uvedenou metodu **chemografií** a v citovaném sdělení je ji možno vyhledat.

Jinou užívanou technikou analýzy chemického prostoru je **shlukování**, klastrování aplikované na výzkum sloučenin, analýzu hledání substruktur a predikci molekulárních vlastností.

Například k analýze podobnosti, k návrhům syntéz a biologickému testování antibakteriálních cyklických peptidomimetik jí použili Katritzky, Berhanu ad. (*Beilstein J. Org. Chem.* 2012, 8, 1146). Analýza příslušnosti sloučenin podle jejich funkcí vychází z jedné dobře definované typové sloučeniny a výsledný obraz ukazuje, že většina testovaných systémů jsou klastrované klastry.



Klastrová analýza založená na molekulárních deskriptorech. Vzdálenost zařazování je v euklidovských mírách. Dvourozměrná reprezentace chemického prostoru je členěna do klastřů podobných sloučenin na principu hierarchické kastrovací metody. Struktury sloučenin označených v excerpované publikaci 37a-37c reprodukuje schéma. Osa X je určena pro sloučeniny a osa Y

sleduje vzdálenost zařazování (linkage distance). V červeném oválu je uveden klaster 1, modře orámované systémy patří do klasteru 2 a v zeleném oválu je klaster 3. (Převzato z publikace uvedeného autorského kolektivu.)

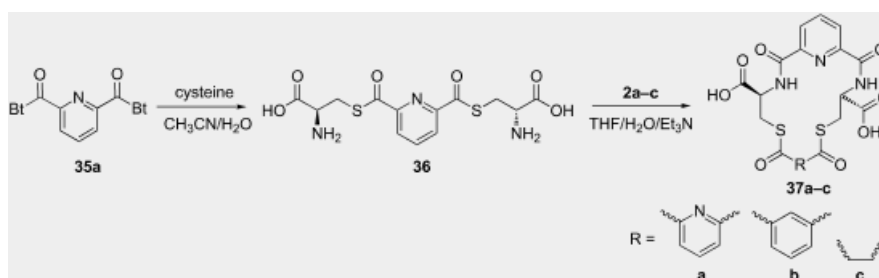
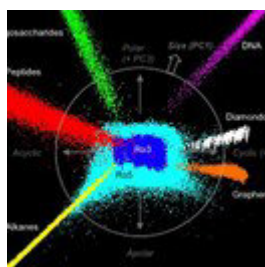


Schéma syntézy makrocyclů obsahujících cystein a vycházející z pyridinových struktur.

Malé organické molekuly jsou jak samy o sobě, tak také jako komponenty větších molekulárních systémů v bioorganické chemii, farmacii, makromolekulární až supramolekulární chemii posuzovány rovněž jako prvky v chemickém prostoru. Už zmínění Raymond, Blum, van Deursen (*Chimia*, 2011, 65, 863) se pokusili o enumeraci všech organických molekul pro sestavení databáze chemického prostoru, GDB, která obsahuje 977 milionů molekul obsahujících až 13 atomů C, N, O, Cl a S. Formou MQN systému vytvořili jednoduchou klasifikační metodu kterou považují za jeden z ekvivalentů periodického systému prvků. Přes jednoduchost 42-rozměrného MQN systému je jejich soustava schopná vystihovat bioaktivitu v seskupeních v MQN systému. Uživatel může v MQN prostoru interaktivně vyhledávat známé organické molekuly v PubChem, a dosud neznámé molekuly v databázi chemického prostoru GDB-13 volně přístupném na www.qdb.unibe.ch.



Mapa MQN pro GDB-11 reprodukuje osm kategorií vyznačených vybarvením podle pravidla priorit: 1. Heteroaromatické (červené) > 2. aromatické (fialové) > 3. kondenzované heterocyklické (fialové) > 4. kondenzované karbocyklické (červenofialové) > 5. heterocyklické (zelené) > 6. karbocyklické (světle zelené) > 7. heterocyklické (žluté) > 8. karbocyklické (oranžové).

Sítě v chemickém prostoru

Suzuki navrhl (2004) **model umělé chemie**, ve kterém je prostor reprezentován **molekulární sítí**. V tomto systému NAC (Network Artificial Chemistry) jsou vrcholy grafu abstraktními reprezentacemi molekul nebo atomových klastrů a hrany představují kolize a vazebné vztahy mezi nimi, k nimž srážky vedou. Podle řešeného systému jsou vytvářeny různé druhy hran. Slabé hrany jsou opakovaně zapojovány podle daných pravidel, a silné hrany mohou být upraveny katalýzou aktivními vrcholy. Například v biologických systémech mohou být v NAC aktualizovány hydrofobní klustry nebo participace na součinnosti v klastrech případně umožňují replikaci templátové informace v kódech molekulárních řetězců.

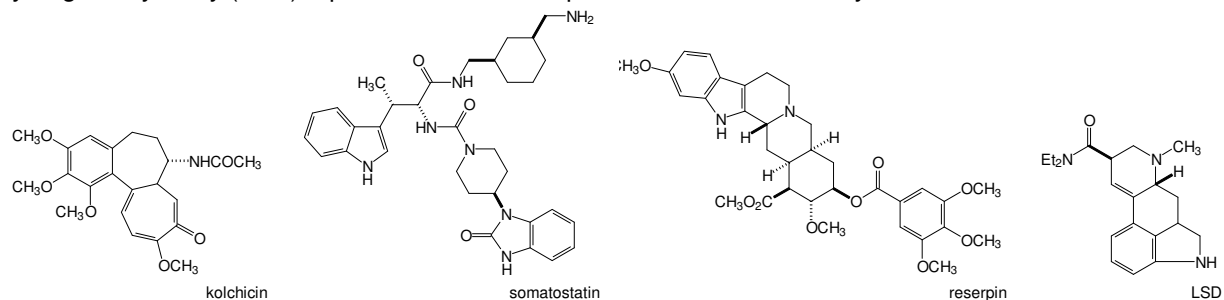
NAC je výzkumný program reprezentující především biomolekulární aktivity formou matematických grafů. Pracuje s prostorovými informacemi kterými jsou koordináty, tvary molekul, úhly a na těchto podkladech umí navrhovat hodnotná schemata, odpovídající logice interakcí mezi dvojicemi vrcholů.

K tomu matematik: Základní strukturou pro reprezentaci sítí je orientovaný graf. Vrcholy grafu modelují jednotlivé vrcholy sítě a hrany jejich spojnice. V orientovaných grafech je každá hrana tvořena uspořádanou dvojicí (u, v) vrcholů grafu, a tudíž taková hrana má směr z vrcholu u do v . Definice: Sítě je čtveřice $S = (G, z, s, w)$, kde

- G je orientovaný graf,
- vrcholy $z \in V(G)$, $s \in V(G)$ jsou zdroj a stok,
- $w : E(G) \rightarrow \mathbb{R}^+$ je kladné ohodnocení hran, zvané kapacita hran.

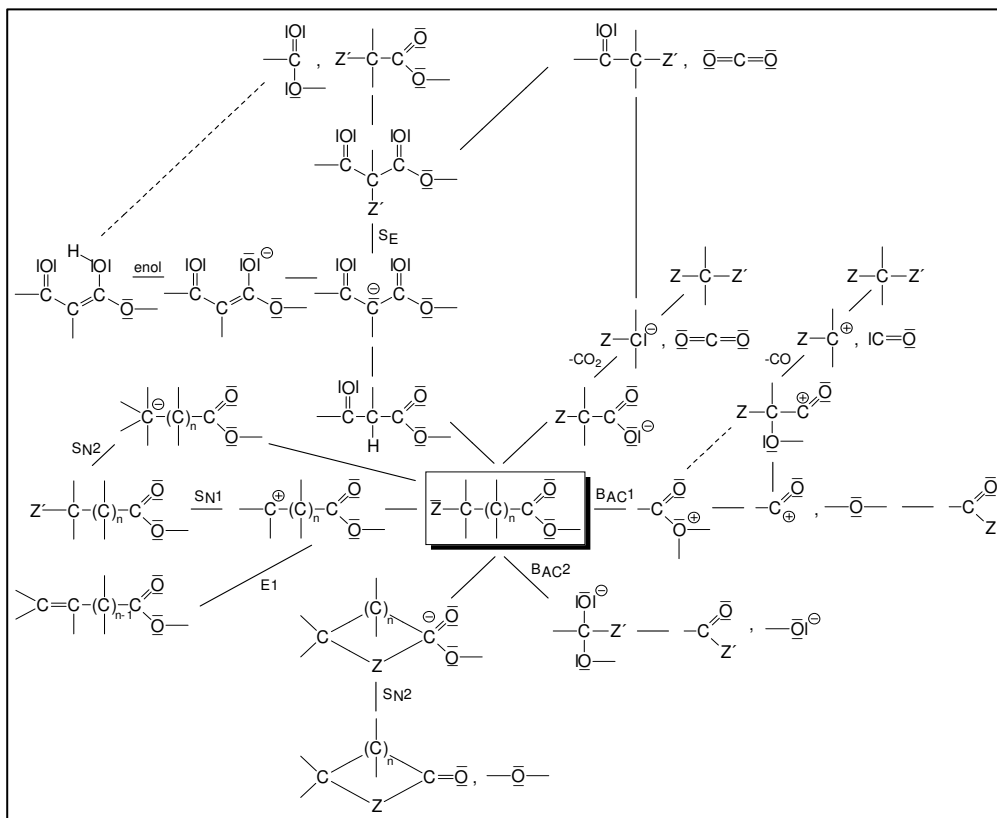
Chemický prostor si tedy můžeme představovat jako kombinatorický a konfigurační prostor osazený všemi chemickými entitami: supramolekulami, makromolekulami, velkými a malými stabilními molekulami, jejich fragmenty, radikály, ionty, přechodovými stavy – všemi chemickými systémy, se kterými se chemici setkávají jako s energeticky stabilními i nestabilními (např. přechodovými stavy, některými radikály, ionty,...) prostorovými uspořádáními (napsal Sukumar N., *Science* 2011).

Přitom „malá molekula“ může být třeba kolchicin, somatostatin, reserpin, *N,N*-diethylamid lysergové kyseliny (LSD) a podobné strukturně poměrně složité molekuly,



V uvažovaném chemickém prostoru se ale setkáváme i s nestabilními intermediárními systémy objevujícími se na „spojích“ (vektorech) mezi reaktanty a produkty (prvky prostoru) chemických reakcí. Myslíme tím všechny kombinace atomů všech prvků, možné (byť přechodně, s nadsázkou efemérně, za vymezených podmínek, např. v argonové matici apod.) při splnění podmínek pravidel kombinací valenčních stavů atomů. Sukumar odhaduje počet všech možných malých organických molekul v chemickém prostoru na více jak 10^{60} a ty obsazují jeden z podprostorů univerza chemického prostoru.

V uvedené ukázce podprostoru karboxylových sloučenin si všimněte, že na „vektorech“ ~ hranách subgrafu spojujících stabilní molekulární útvary jsou zastoupeny i fragmenty s náboji, ionty a radikály, které vznikají v průběhu následných konverzí valenčních stavů atomů:



Je znázorněna síť části podprostoru obsazení okolí startové struktury substitučních derivátů karboxylových kyselin s odpovídajícími interkonverzemi procházejícími nestabilními valenčními stavy reakčních center..

Mluvíme-li o chemickém prostoru, nemáme na mysli jen statické různě sofistikované banky molekul a případně i jejich reakcí. Chemické entity jsou v daném prostoru v zákonitých vztazích uplatňujících se při jejich vzájemných transformacích, které jsou podmíněny a limitovány zákonitými, a tedy přípustnými konverzemi valenčních stavů atomů (radikálů, iontů, transičních stavů). Jsou spjaty vztahy formulovanými principy podobnosti: je molekulární podobnost, chemická podobnost, strukturní podobnost, konstituční podobnost, podobnost molekulárních parametrů (chemických vlastností), podobnost biologických vlastností (např. schopností vázat se na podobné proteiny), podobnost reakcí, podobnost valenčních stavů atomů, vektorová analogie – tak pojmenováváme při popisu chemického prostoru obecně a speciální podoby podobnosti.

Naznačené chemické transformace eduktů v produkty interpretuje síť chemických reakcí v chemickém prostoru.

Matematika definuje síť jako ohodnocený, případně speciálně souvislý ohodnocený graf.

Vztahy podobnosti vystižené v síti mají určitou **metriku**, která definuje vzdálenost mezi dvěma molekulami; metriky lze samozřejmě využít při posuzování vážené a nevážené sítě. Třebaže tyto míry podobnosti jsou vzájemně souvztažné, identické nejsou a to se odráží v různosti sítí. Topologické charakteristiky sítí chemického prostoru se uplatňují například při aplikacích na hledání léčiv. Vzhledem k enormní velikosti chemického prostoru je separátně definován biologický prostor extrahující z něj molekuly s biologickou aktivitou. I v biologických sítích jsou pak studovány okruhy, podmnožiny s cílenou aplikovatelností, se strukturní odlišností či blízkostí, a také s nečekanými vlastnostmi nebo s nežádoucími vedlejšími účinky.

Tým vědců z institutů farmaceutického výzkumu vytvořil virtuální sběrný panel, který jako počítačový model využívá přístup ensemblů podobnosti (SEA, similarity ensemble approach). Detekuje podobnosti mezi chemickými strukturami léčiv a vyděluje soubory molekul, které vykazují vedlejší účinky (Pasolini ad. 2012):



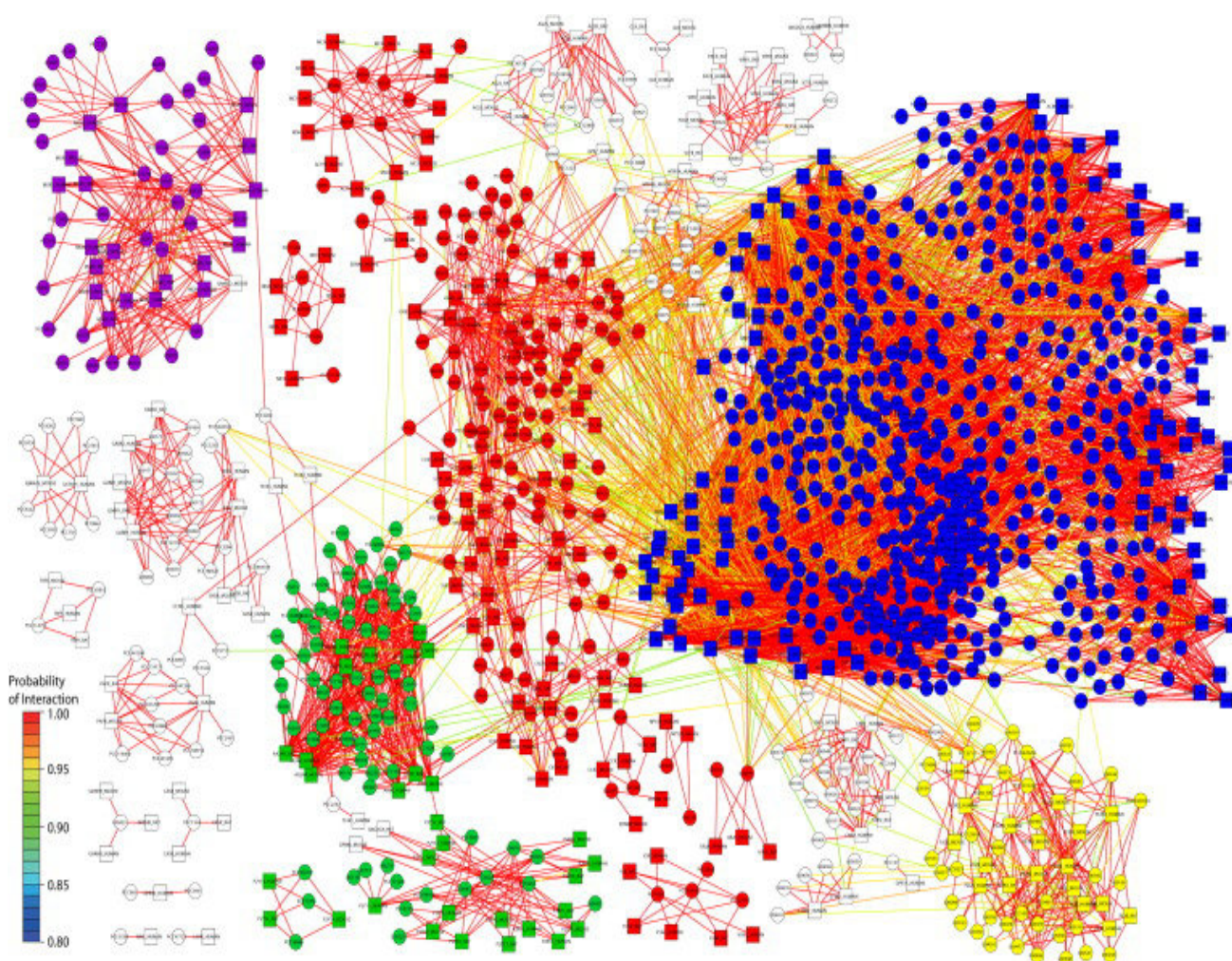
Barevně jsou odlišeny ensembly podobných struktur a jejich prvky (tečky). Šipkami jsou naznačeny směry, ve kterých by mělo být bádání zastaveno pro zjištěné nebo předpokládané vedlejší účinky syntetizovaných a testovaných léčiv.

Síť v chemickém a biologickém prostoru a jejich vzájemného prolínání využívá bohatě především farmaceutický výzkum například při vyhledávání systémů s optimální specifitou k danému proteinu. Jeden příklad za mnohé: Brown a Okuno (*Chemistry & Biology* 2012, 19, 23) úspěšně vytvořili počítačový systém umožňující řešení polyfarmaceutických interakcí. Strategii „jedna k jedné“, jeden ligand-jeden protein přeformulovali na systém oceňující interakce mezi více ligandy a více proteiny, GPCR (G-protein-coupled receptor) přenáší signály (zrakové, sluchové, aktivity imunitního systému a dalších fyziologických funkcí) v organismu. Rozpoznávání povahy a podstaty těchto interakcí mohli sledovat v okruzích (množinách) vytipovaných chemických systémů v podmnožinách chemického prostoru a řešili možnosti vztahů mezi známými receptory a molekulami léků; od známých systémů dosáhli až k prediktivním nálezům, k návrhům nových léků.

Z prací autorů je převzata v originálu ukázka takových sítí, jejich podklasifikací a interakcí mezi prvky v jednotlivých souborech.

Je naznačena komplexita interakcí v sítích mezi ligandy, současně se otevírají možnosti virtuálního vyhledávání dosud

nepodchycených či možných interakcí. Je-li interakce GPCR-ligand předpovězena na bioaktivitu 10 uM, v síti je zakreslena mezi nimi nová hrana. Výsledná síť pak představuje složitost interakcí mezi chemickým a biologickým prostorem a vybízí k posunu strategie hledání léčiv z postupu „jeden ligand-jeden protein“ k uplatnění „více ligandů-více proteinů“. Barevně jsou vyznačeny množiny sloučenin a GPCR podle jejich přiřazení (příbuznosti, podobnosti, uplatnění): modře jsou podloženy aminy, červeně peptidy, žlutě prostanoidy, zeleně nukleotidy. Zbarvení hran = linií od zelené ke žluté a k červené má vystihnout věrohodnost interakcí GPCR-ligand s doplněným číselným vystižením interakcí mezi množinami s vysokou predikční věrohodností.



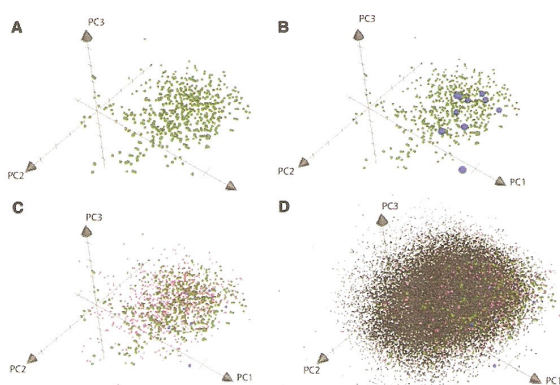
Rozvoj *in silico* hledání lepších a nových farmak je umožněn využíváním počítačových technik. Uvedení autoři aplikují v metodě CGBVS řadu dalších procesů (umělých neuronových sítí, genetických algoritmů, samoučících se strojů k vytváření výpočetních algoritmů ad.) k vyvíjení taktiky využitelné informace o obrovských souborech dat; tak se metoda CGBVS stává i predikátorem nových struktur a jejich reakcí.

„Neomezený“ průzkum dat rozšiřuje naše znalosti o chemicko-biologických interakcích. Integrace zdánlivě nesourodých oborů – teoretické a systematické chemie, matematické chemie, chemické biologie, systematické biologie, polyfarmakologie, klinických informací, makromolekulární, supramolekulární a nanochemie kromě dalších je předpokladem intenzivního pokroku jak v uvedených oblastech lidského poznání, tak v důsledcích – v aplikacích na lidské zdraví.

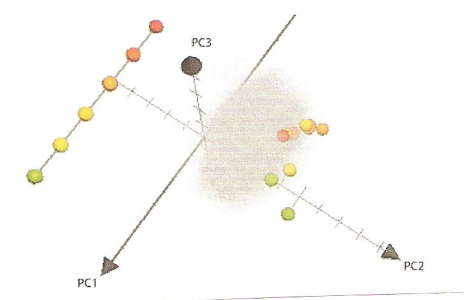
Navigace v chemickém prostoru

Pro chemika zabývajícího se přílehlými obory je internet nesporně hlavním zdrojem informací, prostředků a zprostředkování ve vědeckých aplikacích. O jedné možnosti **navigace v chemickém prostoru** publikovali Rosén ad. (*J Comput Aided Mol Des* 2009, 23, 253) studii zaměřenou na aplikace chemie malých molekul ve výzkumu biologicky aktivních molekul. Centrálním pramenem informací je World Wide Web modifikovaný v ChemGPS-NP Web s metodou analýzy hlavní komponenty, PCA; ta je zaměřena na průzkum **biologicky relevantního chemického prostoru**. V termech globálního mapování do konsistentní osmírozměrné (SD) mapy se dochází k fyzikálně chemickým charakteristikám referenční množiny sloučenin.

První čtyři dimenze ChemGPS-NP mapy čítající 77 % variací dat interpretují autoři takto: první rozměr (principal component one, PC1) zahrnuje velikost, tvar a polarizovatelnost; PC2 obsahuje systémy s vlastnostmi aromátů a konjugovaných struktur; v oblasti PC3 se nacházejí sloučeniny charakterizované lipofilicitou, polaritou a schopnostmi vytvářet H-vazby; PC4 pak reflektuje flexibilitu a rigiditu molekul.



Ukázka výsledků, ke kterým autoři dospěli spojením SMILES z PubChem a skóre predikcí z ChemPGS-NP. Výsledky získané systémem Apple programu Grapher 2.0. **a** jsou v souladu pro aktivní molekuly (587, zelené), **b** aktivní sloučeniny s 2-fenylbenzooxazoly, prosvětlené (11, modré), **c** sloučeniny s nepřesvědčivými vlastnostmi (793, růžové) a **d** všechny testované sloučeniny včetně neaktivních (48, 174, šedé). Výrazným znakem v PC1 je velikost, PC2 aromaticita a v PC3 lipofilicita.



Mapování ChemGPS-NP privilegovaných skupin (zelené s nejnižší, červené s nejvyšší aktivitou).

Mapování chemického prostoru

Moffitt už v roce 1954 postřehl, jak významně závisí elektronegativita (v Mullikenově stupnici χ) na energiích valenčních stavů atomů. Koncepce Paulingova a Van Vleckova valenčního stavu (VS) se postupně zabydlovala ve způsobech uvažování teoretických chemiků řešících okolnosti vzniku chemické vazby párováním valenčních elektronů podle Slatera a Paulinga. Za předpokladu aditivity vazeb jsou vazebné energie přisuzovány valenčním a ne základním stavům atomů. Z tabulek valenčních stavů atomů jsou odvozovány

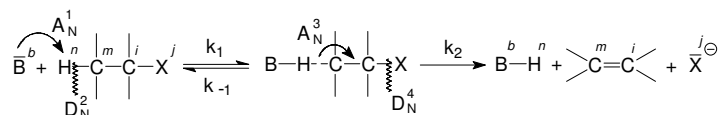
jejich promoční energie a vlivy orbitalových energií na vlastnosti atomů v molekulách. (Moffitt W., *Rep.Prog.Phys.* 1954, 17, 173.)

Kerber, Laue, Meringer, Rocker (2005) publikovali obecný matematický popis reakcí organických sloučenin v termech teorie grafů: Počítačový program MOLDEN-QSPR1. Východiskem byla pro ně teorie Dugundjiho a Ugiho, rozpracovávaná řadou dalších autorů, která reprezentuje organické reakce *in silico*. Hlavním zaměřením je generování chemických reakcí ze zadaných eduktů aplikací dané množiny chemických transformací s cílem získávat „všechny možné“ reakční produkty. Publikace se zabývá molekulárními grafy, substrukturami molekul, chemickými reakcemi, knihovnami získanými z centrální molekuly, konstrukcí reakčních sítí, generováním molekulárních fragmentů a v aplikaci na vytváření kombinatorických knihoven.

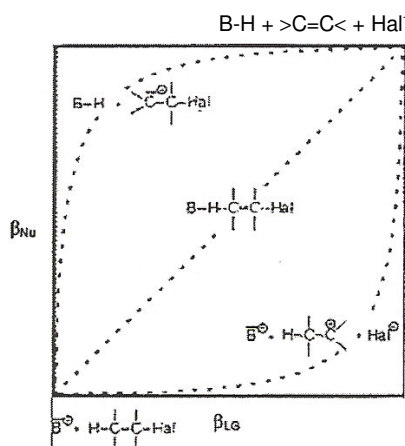
Dugundjiho-Ugiho modelu využili rovněž Leber, Egelhofer, Schomburg a Schomburg (*Journal Bioinformatics* 2009, 25, 23) k automatizovanému přisouzení reakčních operátorů na enzymatické reakce. Publikací s tímto námětem můžete nacházet stále v různých souvislostech.

V modelování kvantitativních vztahů mezi strukturou a reaktivitou, QSAR, spadajících do chemoinformatiky, najdeme řadu metod, založených na číselných informacích, alfanumerických notacích, strukturních otiscích prstů, na skládání subgrafů apod. s využitím počítačových metod. Ne všechny algoritmy jsou schopné poskytovat všechny informace, potřebné k danému účelu – třeba o biologické aktivitě. Vizualizace navazující na mapování chemického prostoru je výhodná při procházení množin dat k hledání závislostí mezi strukturami sloučenin, jejich forem (tvarů) a hodnotami hledaných výsledných situací.

Z tohoto hlediska jsou nejslabší matematické modely popisu sledů elementárních kroků v průběhu chemických transformací eduktů v produkty v intencích teorie AIM. Jejich předností je však automatické modelování možných variací reakčních mechanismů kombinacemi elementárních kroků, EP. Elementární kroky jsou disociace D a asociace A; příklad



Soubor 22 elementárních kroků a jejich notací je v tabulce ve stati Izomerní ensembly molekul. V uvedeném schematu je pořadí (sled) EP v alfanumerickém kódu ANCOD ADAD, celý ANCOD obsahuje kromě údaje posloupnosti kroků počet komponent, rozhodující kroky (určující rychlost děje) a stereochemii; na ně může navazovat umístění modelů reakčních mechanismů v More O'Ferrallových diagramech (MOF)



Grafy reakčních mechanismů mohou být přepsány jak do maticové podoby, tak také alfanumerickým kódem ANCOD: $G_{RM} \rightarrow S_1, S_2, S_3, S_4$. Položka S_1 specifikuje typ EP, $S_1 = x_2^{(1)} x_2^{(2)} \dots x_2^{(n)}$, S_2 násobnost vazeb (1,2,3), řetězec S_3 topologii G_{RM} v obecném označení vrcholů (I,J), (J,N),... a položka S_4 zahrnuje čtyři vstupy čtyřsložkového kódu tak, že specifikuje, zda daný EP nastává/nenastává na dvou vrcholech eduktového MG. Ukázka ANCOD mechanismu eliminace E1 popis specifikuje

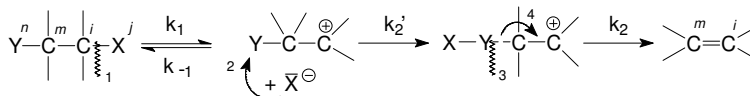
$\frac{\text{DADA}}{S_1} \frac{1112}{S_2} \frac{(\text{J,I})(\text{J,N})(\text{M,N})(\text{M,I})}{S_3} \frac{0100}{S_4}$

Čtyřsložkový ANCOD slouží také ke stanovení vektoru vzdálenosti dvou G_{RM}

$$d(G_{RM}, G_{RM}') = (d_1, d_2, d_3, d_4)$$

a ke shlukování modelů G_{RM} , založenému na jejich podobnosti.

Ukázka stenogramu mechanismu E1:



Počet zaniklých/vzniklých vazeb (2-2+)

reakční centra 4RC

ANCOD DADA 1112 (JI) (JN) (MN) (MI) 0100

matice MSR: R8.

Ukázalo se, že ANCOD může suplovat funkci diskriminátoru při shlukování podobných modelů reakčních mechanismů v chemickém prostoru. V klastrech se shromažďují soubory které mohou představovat reakční série.

ŘETĚZCE ANCOD ELIMINAČNÍCH REAKCÍ

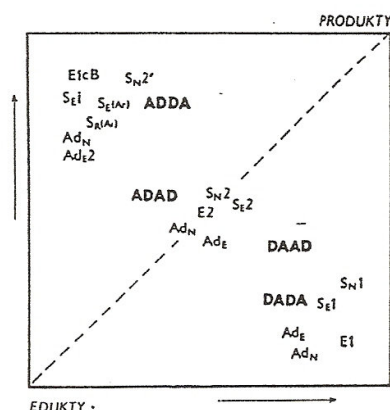
(e1)	E2	ADAD	1121	(BN)(MN)(MI)(JI)	1000
(e2)	E _{ox} (B)	ADAD	1121	(BN)(MN)(MI)(JI)	1000
(e3)	E _{ox}	ADAD AD	1121 12	(NK)(NM)(IM)(IJ)(LJ)(LK)	100010
(e4)	E _{2anion}	ADAD AD	1121 11	(BN)(MN)(MI)(JI)(JN)(LK)	100010
(e5)	E ₌	ADADAD	1132 11	(BJ)(IJ)(IM)(NM)(NJ)(BJ)	100010
(e6)	E1cB	ADDA	1112	(BN)(MN)(JI)(MI)	1000
(e7)	E1 _{therm}	ADDA	1112	(JN)(JI)(MN)(MI)	0000
(e8)	E _{ox} (A)	ADDA	1112	(NA)(NM)(IJ)(IM)	1000
(e9)	E _{ox} (A)	D ADDA A	11121	(LK)(NK)(NM)(IJ)(IM)(IJ)	010001
(e10)	α-E	ADD	111	(BO)(IO)(JI)	100
(e11)	α-E _{karben}	ADDADA	1111 12	(BO)(IO)(JI)(IN)(MN)(MI)	100000
(e12)	E2cB	ADDADA	1111 13	(BO)(IO)(JI)(BN)(MN)(MI)	100100
(e13)	E _{C=Z}	DA	12	(JI)(MI)	00
(e14)	E _{C=Z} B2	AD ADDA	111112	(BN)(MN)(JN)(BN)(JI)(MI)	101000
(e15)	α',β-E	AD ADDA	111112	(BP)(OP)(ON)(MN)(JI)(MI)	101000
(e16)	E1	DADA	1112	(JI)(JN)(MN)(MI)	0100
(e17)	E1(B)	DADA	1112	(JI)(JN)(MN)(MI)	0100
(e18)	Ei	DADA	1112	(JI)(JN)(MN)(MI)	0100
(e19)	E2 _{R+}	DADA	1112	(JI)(JN)(MN)(MI)	0100
(e20)	E2C	A DADA D	11112 1	(BI)(JI)(JN)(MN)(MI)(BI)	10 1000
(e21)	E1c	A DADA D	11112 1	(JL)(JI)(KN)(MN)(MI)(KL)	10 1000 0
(e22)	E1 _{kation}	AD DADA	111112	(JI)(BN)(MN)(MI)	10 0100
(e23)	tele-E	ADAD DADA	1111 1232	(JA)(BA)(JI)(BN) (MN)(ML)(KL)(KI)	1001 0000
(e24)	γ-E	ADAD DADA	1111 1112	(BO)(NO)(JO)(BO)(JI)(PI)(PM)(NM)	1010 0100
(e25)	E2 _{kation}	DAAD	1121	(JI)(BN)(MI)(MN)	0100
(e26)	E _{C=Z}	DAAD	1211	(JI)(BN)(MI)(MN)	0010
(e27)	E1 _{anion}	AD DAAD	11 1211	(BN)(MN) (JI)(MI)(JN)(BN)	10 0010

(e1)	E2	synchronní součinná bimolekulární eliminace
(e2)	E _{ox} (B)	oxidace alkoholů cestou esterů minerálních kyselin, katalyzovaná bázemi
(e3)	E _{ox}	synchronní oxidační eliminace s cyklickým přechodovým stavem
(e4)	E _{2anion}	eliminace za vzniku C=Z vazeb s bazickou katalýzou (E2 + S _N 2)
(e5)	E ₌	eliminace spojená s izomerací allenů na alkyly
(e6)	E1cB	asynchronní eliminace s karbaniontovým intermediátem
(e7)	E1 _{therm}	synchronní termická monomolekulární eliminace
(e8)	E _{ox} (A)	kysele katalyzovaná oxidace alkoholů s acyklickým přenosem H ⁺ (zkrácená verze)
(e9)	E _{ox} (A)	kysele katalyzovaná oxidace alkoholů v souhrnné prezentaci
(e10)	α-E	α-eliminace s tvorbou karbenů
(e11)	α-E _{karben}	α-eliminace cestou karbenů
(e12)	E2cB	eliminace vinylových sloučenin podmíněná silnou bází (obdoba e11)
(e13)	E _{C=Z}	dehydratace karbinolaminů
(e14)	E _{C=Z} B2	eliminace za vzniku C=N, obecně C=Z vazeb, katalyzovaná bázemi
(e15)	α',β-E	eliminace s ylidovým mechanismem (s cyklickým přechodovým stavem)
(e16)	E1	monomolekulární eliminace (s tvorbou karbokationtu)
(e17)	E1(B)	monomolekulární eliminace a s nezbytnou spoluúčastí báze
(e18)	Ei	pyrolytické eliminace s heterolytickým průběhem
(e19)	E2 _{R+}	eliminace na C-C vazbách s kationtovým TS (kationtový protějšek E1cB)
(e20)	E2C	eliminace substrátů stericky bráněných pro S _N 2 (Winsteinův-Parkerův mechanismus)
(e21)	E1c	dehydratace alkoholů s cyklickým přechodovým stavem

(e22)	E1 _{kation}	eliminace přes karbokation, vyžadující kyselou katalýzu
(e23)	tele-E	eliminace probíhající na α,ω-atomech uhlíku alkynů (tvorba C=C=C systémů)
(e24)	γ-eliminace	dehydratace alkoholů s anchimerní asistencí sousední skupiny
(e25)	E2 _{kation}	eliminace cestou karbokationtu s nezbytnou spolúčastí báze
(e26)	E _{C=Z}	eliminace vedoucí k C=Z vazbám
(e27)	E1 _{anion}	eliminace vedoucí k C=Z vazbám (také (E _{CO} 1)anion).

Grafové a maticové modely mechanismů eliminací (E) na C-C vazbách nasycených sloučenin, na C=C a C≡C vazbách substrátů a na C-Z a C=Z vazbách sloučenin se Z = O, N, S, P, ..., jsou uspořádány na základě jejich kinetických kritérií. I při zběžném synoptickém náhledu vám neujdou společné znaky plejády mechanismů eliminačních reakcí.

Sousední MOF diagram vypovídá o silné podobnosti modelů reakčních mechanismů tří základních typů – substitucí, eliminací a adicí.

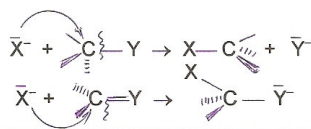


V podprostoru předznamenaném kódy ADDA, ADAD, DAAD a DADA se nacházejí reakční série (RS), do kterých mohou spadat studované reakce. Vzdálenost modelované reakce od reakční série (aktuální vzdálenost) se řeší na dvou úrovních:

Na topologické úrovni se posuzuje vzdálenost modelované reakce od RS z maximální odchylky v hodnotách srovnávaných parametrů (s normalizací a adjustací na míru významnosti jednotlivých parametrů). Předpokladem je dostatečně bohatý soubor parametrizovaných srovnávacích reakcí. Na semikvantitativní úrovni hledáme souvislost dvou stránek chemických reakcí, které může chemik experimentálně určit a kvantifikovat, a tou je souvislost mezi termodynamickou a kinetickou stránkou dějů; tu vystihl (snad poprvé) Brønsted svou známou rovnicí.

Za základní typy reakčních mechanismů jsou považovány substituce, adice a eliminace. Jejich modely jsou natolik tvárné, že další popisované mechanismy jako jsou molekulové přesmyky, redoxní reakce, cykloadiční reakce i speciálně pojmenované procesy, lze řešit využitím modelů S, Ad a E mechanismů.

Formální podobnosti nukleofilních substitucí a adicí si všiml už Shaik (1981) a vystihl ji touto formou

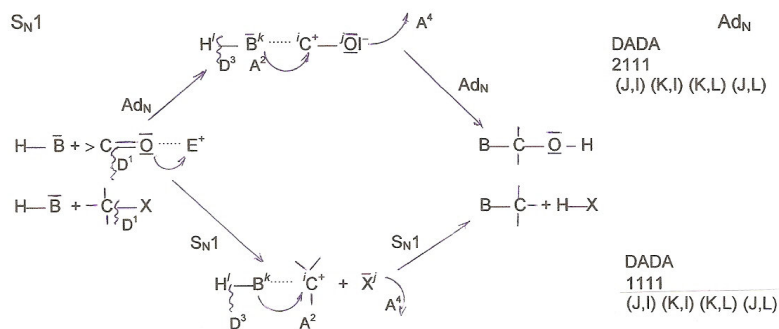


Shaik S.S., *J.Amer.Chem.Soc.* 1981, 103, 3692

V prvním přiblížení uvažoval hraniční orbitály:

vysokoležící obsazený (n) donor a dvojici vazebných a antivazebných orbitalů (X a X*) akceptoru. Je možno psát různé vlnové funkce (základní a excitované) systému jako lineární kombinace konfigurací reaktantů, vznikajících distribucí čtyř zúčastněných elektronů po třech orbitalech n, X a X*. To je korektní úvaha, pokud můžeme tyto orbitály definovat v delokalizovaných orbitalech D-A komplexu v každém bodu podél reakční koordináty.

V aparátu matematického modelu chemie můžeme k uvedené podobnosti mechanismů dospět díky ANCOD takto:



Využitím aparátu matematického modelu chemie je možno odkrývat také další podobnosti. Třeba modelů mechanismů S_N2 a Ad_N , $E2$, $E1$, S_N1 a $E1$, eliminaci můžeme považovat za interní substituci (čím: π -elektrony z prekursoru, ostatně o jejich kompetici víme (a mluvíme o reakci S_N vs. E), obrácenou sekvencí elementárních kroků při eliminaci je jejich „retro“ sekvence při adici (to píšeme $\overleftarrow{E}$) a A a D v modelu adice a substituce, pak při adici jsou interně „substituovány“ π -elektrony v substruktuře $K \rightleftharpoons I$. Z uspořádání notací v ANCOD1 lze vyvozovat molekularitu modelového mechanismu

ADAD odpovídá mechanismům bimolekulárních reakcí,

DADA odpovídá mechanismům monomolekulárních reakcí,

ADDA popisuje mechanismy s karbaniontovým, případně karbokationtovým TS,

DAAD popisuje mechanismy blízké monomolekulárním dějům,

DDAA najdeme v popisu α -eliminací s karbenovými stavy a v popisu přesmyků,

AADD najdeme v popisu jejich retroreakcí (reakcí karbenů, nitrenů apod.).

Chemik nepochybně ocení možnost, kterou mu nabízí **shlukování modelů reakčních mechanismů** v podobě ANCOD a grafů G_{RM} , k vytváření sjednocujícího pohledu na chemické reakce.

Koča J., Kratochvíl M., Kvasnička V., *Coll. Czech. Chem. Commun.* 1985, 50, 1433

Koča J., *Coll. Czech. Chem. Commun.* 1988, 53, 3119

Kvasnička V., *Coll. Czech. Chem. Commun.* 1983, 48, 2118

Při formulování modelu byl zaveden vektor vzdálenosti dvou grafů reakčních mechanismů

$G_{RM} = (V, I, \phi, L, \phi, B)$ a $G'_{RM} = (V', I', \phi', L', \phi', B')$ se stejným počtem elementárních kroků a se stringy $x_2^{(1)} x_2^{(2)} \dots x_2^{(n)} x_3^{(n)} (I_1, J_1)(I_2, J_2) \dots (I_n, J_n) x_4^{(1)} x_4^{(2)} \dots x_4^{(n)}$ a $x_2'^{(1)} x_2'^{(2)} \dots x_3'^{(1)} x_3'^{(2)} \dots x_3'^{(1)} (I_1', J_2')(I_2', J_2') \dots (I_n', J_n') \dots x_4'^{(1)} x_4'^{(2)} x_4'^{(3)} \dots x_4'^{(n)}$.

Vektor vzdálenosti mezi těmito dvěma grafy je formálně dán uspořádanou čtveřicí

$$d(G_{RM}, G'_{RM}) = (d_1, d_2, d_3, d_4),$$

kde

$$d_1 = \sum \delta(x_2^{(i)}, x_2'^{(i)}),$$

$$d_2 = \sum \delta(x_3^{(i)}, x_3'^{(i)}) - \text{reflektuje podobnost topologie mezi grafy } G_{RM} \text{ a } G'_{RM},$$

$$d_3 = 0 \text{ je-li funkce } F \text{ identitou, a } d_3 = 1, \text{ jde-li o korespondenci jedna k jedné (ne však identitu),}$$

$$d_4 = \sum \delta(x_4^{(i)}, x_4'^{(i)}).$$

Ze stanovení vektoru vzdálenosti $d(G_{RM}, G'_{RM})$ mezi posuzovanou dvojicí grafů reakčních mechanismů o stejném počtu hran (elementárních kroků EP) plyne významná vlastnost symetrie

$$d(G_{RM}, G'_{RM}) = d(G'_{RM}, G_{RM}).$$

K tomu příklad: Určili jsme tři reakce pomocí ANCOD

(1) E1	DADA 1112	(L)(K,L)(K,L)	0000
(2) E1 ^(B)	DADA 1112	(J,I)(B,L)(K,L)(K,L)	0100
(3) E2	ADAD 1121	(B,L)(K,L)(K,L)(J,I)	1000

Vektory vzdálenosti mezi nimi jsou $d^{(1)(2)} = (0,0,2,1)$

$$d^{(1)(3)} = (4,2,3,1)$$

$$d^{(2)(3)} = (4,3,2,2).$$

Dalším rozpracováním vektoru vzdálenosti se dospěje ke shlukování reakčních mechanismů specifikovaných jejich kódem ANCOD. Příklad:

1)	E1	DADA 1112	(Jl)(JN)(MN)(MI)	0000
2)	E1cB	ADDA 1112	(BN)((MN)(Jl)(MI)	1000
3)	E2	ADAD 1121	(BN)(MN)(MI)(Jl)	1000
4)	E2*	DAAD 1121	(Jl)(BN)(MI)(MN)	0100
5)	Ad _N	DADA 2111	(MI)(Jl)(JN)(BN)	0101
6)	Ad _{E2}	ADAD 1211	(IK)(IJ)(LJ)(LK)	1010
7)	Ad _{E2}	ADDA 1211	(IK)(IJ)(LK)(LJ)	1001
8)	Ad _N ^{Gr}	ADDA 1211	(MN)(MI)(JM)(Jl)	1001
9)	S _N 1	DADA 1111	(Jl)(Kl)(KL)(JL)	0101
10)	S _N 2	ADAD 1111	(Kl)(Jl)(JL)(KL)	1010
11)	S _{E2}	ADAD 1111	(IK)(IJ)(LJ)(LK)	1010
12)	S _{N2} ^(C+)	DAAD 1111	(Jl)(Kl)(JL)(KL)	0110

Všechny různé vektory vzdálenosti jsou uspořádány podle lexikografického pravidla pro čtyři různé permutace π . Jsou čtyři různá uspořádání s klasifikačními klíči $\pi_1 = (1,4,2,3)$; $\pi_2 = (2,1,4,3)$; $\pi_3 = (4,1,2,3)$; $\pi_4 = (3,1,4,2)$.

Klastrování pomocí klasifikačního klíče π_1 preferuje sekvenci elementárních kroků (D,A). V prvním kroku jsou si zpodobněny mechanismy Ad_2 a Ad_N , v sedmém kroku jsou doplněny mechanismem $E1cB$; společným znakem těchto mechanismů je umístění jejich ANCOD v levém horním rohu MOF diagramu s aniontovým intermediátem. Všechny tyto reakce obsahují stejný čtvrtý elementární krok, asociaci, který signalizuje rychlost tvorby produktu. Ve druhém kroku shlukování podle klasifikačního klíče π_1 jsou pospojovány bimolekulární mechanismy s $S_1 = ADAD$; jejich přechodové stavy jsou umístěny ve středu hlavní diagonály MOF a jsou to S_N2 a S_E2 mechanismy. Ve čtvrtém kroku shlukování se k nim přidruží mechanismus $E2$, v pátém kroku pak Ad_E ; ve třetím kroku podle klasifikačního klíče jsou sdruženy mechanismy Ad_N a S_N1 se stejným stringem $S_1 = DADA$, v osmém kroku se připojuje mechanismus $E1$. Jejich reakční koordináta prochází pravým dolním rohem MOF; zahajovány jsou disociací a vznikem karbokationtu (*IM*). V šestém shlukovacím kroku podle klíče π_1 jsou sdruženy reakce $E2^+$ (s tvorbou karbokationtu v prvním EP a s následující spoluúčastí báze) a $S_N2^{(C+)}$ (spadají sem například reakce trifenylmethyl-halogenidů s kyslíkatými nukleofily s rychlou heterolýzou vazby C-Hal a následnou pomalou asociací reagentu ve druhém kroku) s $S_1 = DAAD$. V MOF diagramu jsou umístěny mezi středem a pravým dolním rohem. Klasifikační klíč π_1 seskupuje reakční mechanismy podle umístění TS v MOF diagramech.

Shlukování podle klasifikačního klíče π_2 poskytuje hrubou klasifikaci na základě změn hybridních stavů reakčního centra ($sp \rightarrow sp^2$ a $sp^2 \rightarrow sp^3$); seskupení substitucí je dosaženo po čtyřech krocích, eliminací také.

Klasifikace π_3 preferuje reakce se stejným způsobem participace reagentů. V prvním a druhém kroku se soustřeďují reakce, pro které je důležitá součinnost s další složkou eduktů (reagentem, katalyzátorem). Tak ve třetím kroku se sblíží mechanismy S_N1 a Ad_N , charakterizované krokem D na strategické vazbě substrátu (to bez spoluúčasti partnera) a významné jsou řetězce S_4 a S_1 ; jejich seskupení je v osmém kroku doplněno mechanismem $E1$. Klasifikace podle klíče π_4 vystihuje topologii studovaných grafů reakčních mechanismů. Označování vrcholů je jako dosud smluvní. Shlukováním reakčních mechanismů z tohoto hlediska se dospělo k hrubému rozkladu do různých klastřů.

Systematický průzkum chemického prostoru umožňují databáze *in silico* virtuálních molekul. Ty mohou být vizualizovány projekcí mnohorozměrného prostoru molekul a jejich vlastností do 2D nebo 3D formy.

Podrobné informace můžete nalézt v těchto pracech: L. Ruddigkeit, R. van Deursen, L. C. Blum and J.-L. Reymond (2012). "Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17". *J. Chem. Inf. Model* 2012, 52 (11): 2864. doi:10.1021/ci300415d. PMID 23088335.

M. Awale, R. van Deursen and J. L. Reymond. "MQN-Maplet: Visualization of Chemical Space with Interactive Maps of DrugBank, ChEMBL, PubChem, GDB-11, and GDB-13". *J. Chem. Inf. Model* 2013. doi:10.1021/ci300513m. PMID 23297797.

L. Ruddigkeit, L. C. Blum and J.-L. Reymond, "Visualization and Virtual Screening of the Chemical Universe Database GDB-17". *J. Chem. Inf. Model* 2013 53(1) 56. doi:10.1021/ci300535x. PMID 23259841.

Vyplatí se také projít v Google heslo Visualisation of Chemical Space with interactive Maps of DrugBank.

Mapování chemického prostoru je většinou založeno na **molekulárních deskriptorech**. Například autoři Haggarty, Clemons, Wong a Schreiber (*Comb Chem High Throughput Screen* 2004, 7, 669) s využitím analýzy hlavní komponenty a trojrozměrné vizualizace vytvářejí **metrické mapy prostoru** s morfologickými znaky náležejícími různým prvkům diverzity v knihovnách. Filtrováním těchto map pomocí fenotypových deskriptorů, získaných z měření parametrů malých molekul, se dospěje například k hodnotám biologické aktivity ve specifických podprostorech.

Reymondův tým vytvořil metodu jednoduché klasifikace v **systému MQN** (Molecular Quantum Numbers), který v sobě nese znaky periodického systému organogenních prvků. Přes svou jednoduchost je aplikován na řešení bioaktivity molekul. MQN prostor známých organických molekul v PubChem a nepopsaných molekul v Chemical Universe Database GDB-13 je volně přístupný na www.gdb.unibe.ch.

Počet molekul respektujících pravidla a podmínky chemických vazeb, stabilní geometrie a funkční skupiny je pro daný počet atomů vymezený. Enumerace vychází z dané kolekce matematických grafů a realizuje se buď v grafové reprezentaci nebo metodou otisku prstu (molecular fingerprints odpovídají strukturním vzorcům). Autoři (Reymond, Blum, van Deursen, *Chimia* 2011, 65, 863) uvádějí programy používané k vizualizaci chemického prostoru. Také autoři Škoda a Hoksza (Springer 2013) se zabývají v publikaci *Chemical Space Visualisation Using ViFrame* programováním ve sféře chemického prostoru.

Krátká poznámka: Smyth a McClave navrhli a popisují techniku založenou na principu „podobnost vs. diverzita“ (*Lecture Notes in Computer Science* 2001, 2080, 347). Systémy řešící porozumění konkrétním případům využívají konvenčního předpokladu podobnosti aby vyhledávaly soubory případů, které jsou co nejpodobnější standardu. V řadě situací a případů je to úspěšný přístup, někdy zdoluhavý, v tom ovšem nabízí řešení diverzita. Systémy nazvané Base-based reasoning (CBR) řeší nové problémy tak, že opakovaně procházejí problémy zpracované v předchozích úlohách a uložených v tzv. case-base. CBR kombinuje techniky vyhledávání informací, umělé inteligence a uživatelské zkušenosti (expertního systému). Smyslem kombinování podobnosti a diverzity je především urychlení hledání

řešení. Uplatnění mají v chemometrice, kombinatorických knihovnách při aplikacích v ekologii, genovém inženýrství a pod. Chemický prostor je přístupný vizualizaci a kvantifikaci chemické diverzity. To, co chemik označuje jako „prostor“ nemusí vždy poskytovat požadovanou informaci. Nepředpokládaný úsudek může ovlivnit interpretaci. Chemická diverzita je abstraktní koncepce obtížně přístupná důslednější normalizaci a standardizaci. Samozřejmě je navrženo a navrhováno více metod pro tyto typy průzkumu.

Zasvěcenou informaci může čtenář nalézt také pod *Space, Diversity, and Synthesis* Jeremy Henle, 4/23/2013. Chemical Space - [School of Chemical Sciences](#). Princip diverzity pak osvětlil Martin, Y.C., Diverse viewpoints on computational aspects of molecular diversity. *Journal of Combinatorial Chemistry* 2001, 3, 231. O užitečnosti mapování chemického prostoru pro biologickou chemii pojednává práce Lipinski, Lombardo, Domini, Feeney v *Adv. Drug Delivery Reviews* 1997, 23, 3

Problému explorační chemického prostoru osazeného známými i dosud neznámými organickými malými molekulami, které představují stavební kameny větších formací, se věnuje mnoho zájmových skupin. Jména jejich některých protagonistů se v rešerších stále opakují. Stejně jako jejich základní otázka jak analyzovat v záplavě molekul v obrovských chemických prostorech i podprostorech logickými postupy rozumný přehled o chemické podobnosti a diverzitě a jak konstruovat využitelné programy.

Reymond v citovaném sdělení reprodukuje jeden z nich, DrugBank (Knox, Jewison, Liu, Ly, Frolkis, Pon, Banco, Mak, Neveu, Djoumbou, Eisner, Guo, Wishart, *Nucleic Acids Res.* 2011, 39, D1035):

	database	description	size ^a	Web address	
Gütlein, Karwath Kramer	PubChem	known molecules from various public sources	32.5 M	http://pubchem.ncbi.nlm.nih.gov	a
	Chemspider	online resource from the Royal Society of Chemistry	26.0 M	http://www.chemspider.com/	
	ZINC	commercially available small molecules	21.0 M	http://zinc.docking.org	
	NCI Open	anticancer and AIDS compounds with screening data	0.25 M	http://cactus.nci.nih.gov/ncidb2.1	
	ChemDB	commercially available small molecules	4.1 M	http://cdb.ics.uci.edu	
	BindingDB	bioactive molecules with binding affinity data	0.36 M	http://www.bindingdb.org	
	ChemBank	small molecules annotated with screening data	1.2 M	http://chembank.broadinstitute.org/	
	ChEMBL	small molecules annotated with experimental data	1.1 M	https://www.ebi.ac.uk/chembl/db	
	CTD	comparative toxicogenomics database	0.17 M	http://ctdbase.org	
	HMDB	human metabolome database	0.0085 M	http://www.hmdb.ca	
	SMPDB	small molecule pathway database	0.001 M	http://www.smpdb.ca	
	DrugBank	experimental and approved small molecule drugs	0.0065 M	http://www.drugbank.ca	

(*Journal of Cheminformatics* 2012, 4, 7) vytvořili obecný interaktivní program aplikace mapování nazvaný CheS-Mapper umožňující prohlížení množin malých molekul, hledaných uživatelem ke kombinacím s reaktanty, a jejich obrazy. Program vizualizuje struktury sloučenin a jejich tvary ve virtuálním 3D prostoru. Detekuje substrukтуры (klastry) s příslušnými daty a je možno jej používat k analýze dat vhodných k získávání informací o vztazích struktura-reaktivita (SAR). Aplikace je rozdělena do dvou částí, nazvaných Chemical Space Mapping a Visualisation of Chemical Datasets in Chemical Space.

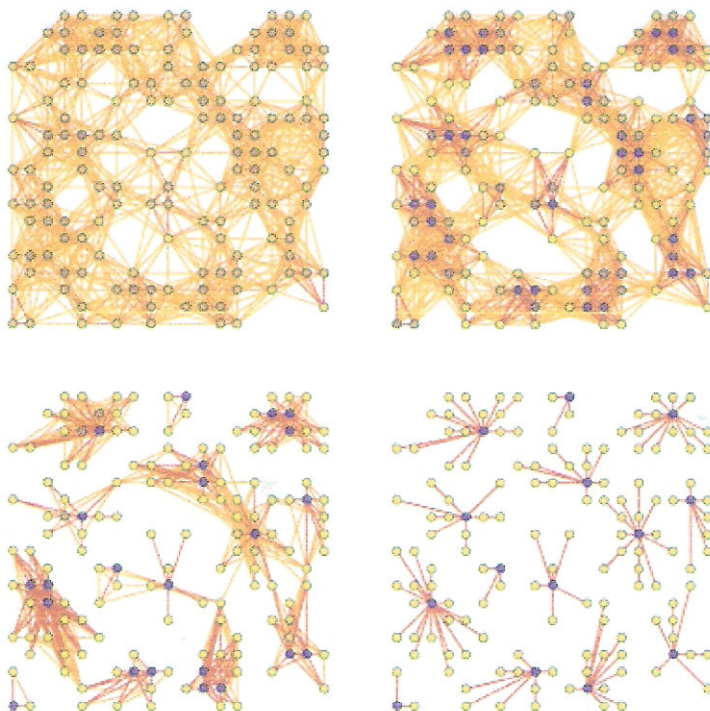
Molekulární prostor

V hmotných objektech kolem nás nacházíme podobnost a je vyjádřena v principu teorie nekonečného hierarchického „navršování“ hmotného světa (v angličtině „nesting“ je přílehlavější). Uvedený princip říká, že mezi jednotlivými a na pohled různými úrovněmi uspořádání hmotného světa existují vzájemná propojení. Podobnost objektů v jeho různých úrovních odpovídá $SP\phi$ symetrii a je ilustrována systémy od nejmenších pozorovatelných

subatomárních částic až po největší pozorovatelné superklastry galaxií s mezihvězdnými vzdálenostmi a kvantováním parametrů kosmických systémů, a objekty nedosažitelnými přímým pozorováním. Základní fyzikální konstanty představují hmotnost, velikost, spin, elektrický náboj, magnetický moment, energie, teplota ad. Na úrovni hvězd platí konstanty Planckova, Diracova, Boltzmannova a další. Všechny nám známé objekty jsou skládány z částic (částí) z nižší úrovně hmoty a už to znamená, že mezi nimi existují vztahy a především vztahy podobnosti. Z možnosti umístění kosmických objektů na různých úrovních hmoty na pomyslné škálované ose dimenzí vyvozují někteří autoři úvahy o pátém rozměru časoprostoru.

Vyjdeme-li z opačného konce naší úvahy, vstupují do ní Aristoteles, Descartes, Leibnitz a Newton, a již z 5. století p.n.l. Anaxagoras: u nich nacházíme myšlenku, že hmota je dělitelná do nekonečna; také že pohyb, atrakce a repulze se uplatňují jak mezi „nebeskými tělesy“, tak mezi nejmenšími částčkami, se kterými se na Zemi setkáváme. U Anaxagora můžeme s odvážnou představivostí a soudobým výkladem jeho terminologie nalézt i zárodky **koncepce fraktálů**. Stať končíme posouzením toho nejmenšího útvaru, kterým je pro chemika atom a prostor, který zaujímá v molekule.

Skládání větších systémů z menších útvarů si můžeme znázornit ilustračním vyobrazením, převzatým ze sdělení Alfred Wegener Institut o biologickém výzkumu v oblasti biodiverzity, biomineralizace a dalších biologických systémů s využíváním bioinformatiky:



Výchozí je obrázek vpravo dole, na kterém jsou klastrovacím algoritmem uvedeny shluky grafů, z nichž se postupně na principu podobnosti utvářejí složitější klastry. V obráceném sledu prohlížení můžeme hledat ukázkou uplatnění principu rozkladu grafové prezentace složitějších systémů až na prvopočáteční edukty.

Nekončící dělitelnost potkáváme v různých podobách a interpretacích ve filozofii, fyzice, chemii, matematice (v té zejména v teorii uspořádanosti), v teorii pravděpodobnosti,... i v ekonomii. Sledovanými objekty jsou

hmota, látka, prostor, čas a abstraktní matematické objekty, mj. kontinuum.

Zastavíme se u **molekulárního prostoru** řešeného teorií AIM.

Výpočty mnoha fyzikálních a chemických entit zavedené Baderovou teorií atomu v molekule (AIM, Atom-in-Molecule) vycházejí z parcelace molekulárního prostoru na atomové objemy obsahující přesně jedno atomové jádro. Při dalším rozvíjení a aplikacích teorie může aim představovat v přeneseném smyslu i víceatomové substrukтуры, fragmenty a promolekuly.

V dané souvislosti stačí zmínit, že přístup v intencích teorie AIM a kvantové teorie QAIM představuje kvantově chemický model, který charakterizuje chemickou vazbu systému na bázi topologie kvantové hustoty náboje. Za poslední tři desetiletí se rozvinula

teorie AIM v jeden z nejužívanějších teoretických nástrojů chemie. Teorie definuje atom jako vlastní otevřený systém, prostorovou formaci, která může sdílet energii a elektronovou hustotu lokalizované v 3D prostoru, tedy může být definována v terminologii lokálních zakřivení elektronové hustoty. Pro matematiky představují teorémy AIM topologii elektronové hustoty. Pro teorémy QTAIM je molekulární struktura dána stacionárními body elektronové hustoty. Teorie zvládá výpočty širokého okruhu fyzikálních vlastností na základě atomů díky parcelování prostoru na atomové objemy, které obsahují právě jedno atomové jádro (aim) uplatňující se jako atraktor elektronové hustoty.

Shrnutí:

Molekula může být rozparcelována na množinu atomových objemů, které jsou rozděleny do sérií podle gradientu vektorového pole. Vlastnosti atomů – náboj, dipólový moment a energie jsou počítány integrováním odpovídajících operátorů přes atomový objem.

Mezi dvěma atomy existuje vazba, pokud jejich atomové objemy sdílejí společný meziatomový povrch a na něm je kritický bod $(3,-1)$. Kritický bod definuje Bader jako bod v prostoru, jehož gradient je nulový. V kritickém bodu $(3,-1)$ jsou vlastní hodnoty („eigenvalues“) Hessianovy matice negativní, ostatní vlastní hodnoty jsou kladné. To můžeme vyjádřit také takto: kritický bod vazby, cbp, představuje sedlový bod prvního řádu ve skalárním poli elektronové hustoty. Stopa vazby je linie, podél které dosahuje elektronová hustota maxim vzhledem k okolí.

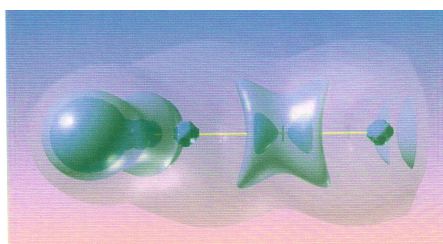
Meziatomové vazby jsou v teorii AIM klasifikovány jako spojení s uzavřenými slupkami nebo sdílené, je-li Laplacean elektronové hustoty v kritickém bodu kladný nebo záporný.

Geometrické pnutí vazby lze měřit zjištěním odchylky cbp od meziatomové vazby mezi dvěma atomy. Velká odchylka odpovídá většímu pnutí vazby.

O tom, co všechno QTAIM zahrnuje a zvládá, stačí heslovité informace:

Topologie elektronové hustoty určuje tvar, formu atomů v molekulách
 Umožňuje sestavování vazeb a viriálních spojů a molekulových i viriálních grafů
 Vyčlenění atomů z molekulárních vlastností
 Nodální povrch v Laplaceanu indukující reaktivní místa na povrchu molekuly.

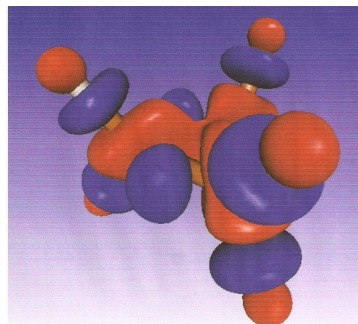
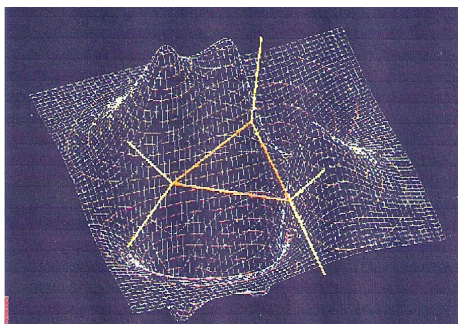
Takže Baderova teorie AIM je objektivní cestou k popisu molekulárních struktur a k předvídání mnoha molekulárních vlastností. Využívá topologie hustoty elektrického náboje, která supluje skalární pole v trojrozměrném prostoru, a kvantově chemických pozorovatelných k objektivní informaci o molekule. Levit (1996) byl Houtovou koncepcí aplikace elektronové hustoty a obrazy molekulových orbitalů inspirován k dalšímu rozvíjení teorie AIM při postupném zvyšování kapacity a možností počítačů k elaborátu, v němž se zabývá rostoucí přesností výpočtů (jmenovitě DFT), zdokonalovanou grafickou informací o topologickém obrazu a dynamické 3D produkci (jako VRML) spějící k možnosti modifikace výsledků pro využívání www.



Ze studie Levita (*A New Pictorial Approach to Molecular Structures and Reactivity*, 1996) je převzato vyobrazení kvadrantu molekuly ethanu. Je patrná eliptická povaha C-C vazby, v oblasti vazby je výrazné nahuštění náboje.

Simulace byly provedeny s Gaussian 94 využitím DFT na b3lyp/6-311++G(2d,2p)//b3lyp/6-311++G(2d,2p) úrovni teorie.

Molekulové orbitály jsou jakousi odpovědí kvantové chemie na otázku Kde jsou elektrony v molekule. Vyobrazení elektronové hustoty reprodukuje sumu všech atomových orbitalů AO a počet všech elektronů ve všech zaplněných MO; zobrazení 3-D jsou izopovrchem, protože povrch molekuly sestává z mnoha bodů stejné elektronové hustoty. Jsou obrazem zaplnění prostoru („space-filling“) molekulou. Elektrostatický potenciál EP je korelovatelný s dipólovým momentem, elektronegativitou, nábojem a dalšími parametry, které charakterizují molekuly. Tvar molekuly je dán elektronovou hustotou molekuly. Ta závisí na sestavě atomů a na jejich chemické konektivitě v molekule. Jednou z možností stanovení



Jsou modelovány povrchy tvarů molekul cyklopropanu VRML 2.0

tvaru molekuly je výpočet elektronové hustoty kvantově chemickými metodami, z nichž zmiňme populární program MOLDEN, kterým lze počítat povrchy elektronové hustoty a povrchy elektrostatického potenciálu na základě informace získané Gaussianem nebo Firefly (PC GAMESS). Kahn (2012) nabídl soubory vizualizovaných molekulárních prostorů. Ukázka dvou z běžných provedení vizualizace výpočtů povrchů molekulárních tvarů, které můžeme považovat za vizualizaci molekulárních prostorů, vydá za dlouhé komentáře:

Při navigaci chemickým prostorem, jak jsme si uváděli, slouží jako souřadnice molekulární deskriptory. Molekuly jsou transponovány do číselných kódů podle použité matematické rutiny. Hlavní třídy jsou

- 0D – deskriptory – počet atomů, počet vazeb, molekulová hmotnost
- 1D – deskriptory – sestavy fragmentů
- 2D – deskriptory – transpozice do grafové řeči
- 3D – deskriptory – zahrnují informace o geometrii a velikosti molekul, sterické informace a deskriptory povrchu
- 4D – deskriptory – 3D koordináty a konformace.

Princip dokonalé kauzality

Chemický prostor představuje kombinatorický a konfigurační prostor zahrnující všechny možné molekuly a jejich fragmenty existující díky vazbám splňujícím pravidla kombinací valenčních stavů mezi atomy, vedoucích k energeticky stabilním i intermediálním uspořádáním. Bylo už zmíněno, že celkový počet možných malých organických molekul obsažených v chemickém prostoru převyšuje 10^6 a je pochopitelně mimo prakticky dosažitelné možnosti chemiků i Přírody. Je i větší než celkový počet atomů ve známém Univerzu.

Molekuly v chemickém prostoru jsou ve vzájemných vztazích, představovaných mnoha a různými parametrizovanými projevy. Z nich nejdůležitější jsou **vztahy podobnosti**, na jejichž základě probíhají chemické přeměny na určitých reakčních centrech. Tak je vytvářena v chemickém prostoru (virtuální) síť chemických reakcí, v níž je mnoho transformací reálných.

Vztahy podobnosti nacházíme jako

- konstituční podobnost (podobnost atomů v molekule, aim),
- strukturní podobnost (podobnost substruktur a fragmentů molekul),
- podobnost trojrozměrných tvarů (forem v jejich grafickém zpracování),
- podobnost chemických vlastností (projevující se především v reaktivitě),
- podobnost biologických vlastností (... ve vazbách k podobným proteinům).

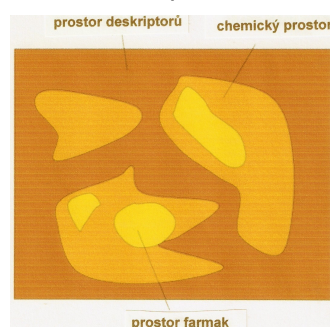
Podprostory (regiony) chemického prostoru, mapované chemikem zabývajícím se určitou oblastí výzkumu, jsou samozřejmě limitovány ústřední strukturou a jejím okolím. Nemusí jít

vždycky o syntetizování dosud nepopsaných sloučenin, uplatňuje se strukturní diverzita a omezení představuje dostupnost eduktů, jejich vlastnosti, vnější podmínky syntéz (katalýza, teploty, tlaky apod.), takže badatel se může pohybovat v různých regionech chemického prostoru, s různou metrikou podobnosti a tak se prolíná topologie sítí chemického prostoru s biologickými sítěmi nebo sítěmi společenských i kognitivních věd.

Široké aplikace a neustálé zdokonalování zaznamenaly programy řešící chemickou podobnost v oblasti studia biologických aktivit. Změny biologických aktivit spjatých se změnami molekulárních struktur umí chemici vystihnout prostředky teorie vztahů mezi strukturou a aktivitou (SAR). Princip podobnosti (SP) vyjadřuje předpoklad, že podobné molekuly vykazují podobné vlastnosti v biologických testech; výpadky platnosti uvedeného principu byly pozorovány v případech, kdy velice si podobné molekuly měly rozdílné biologické aktivity. (Maggiora je nazval „aktivity cliffs“, jejich mírou může být index terénu aktivity struktury SALI (Structure Activity Landscape Index) a ukazuje se, že tyto regiony jsou lákavým tématem výzkumů, protože malé změny struktury molekuly mohou vést ke sloučeninám s velkou biologickou účinností.)

Zasvěcené studie může čtenář hledat například pod heslem [similarity in chemical space publications](#) a jmény Robert Ponec, David Hoghza a Daniel Svozil, samozřejmě s publikacemi mnoha dalších autorů.

Kdybychom chtěli závěrem zhodnotit úlohu koncepce chemického prostoru, už těch několik předchozích úryvků je dostatečným potvrzením významu fenoménu podobnosti v chemii (a nejen organické), v biochemii, biologické chemii a chemické biologii, v supramolekulární chemii a především ve farmaceutickém výzkumu při hledání nových léčiv a sledování jejich interakcí s funkčními komponentami živých organismů.



Ilustrační obrázek k závěrečnému odstavci.