MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV BIOCHEMIE

Bakalářská práce

BRNO 2019 RADKA SEDLÁKOVÁ

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV BIOCHEMIE

Atomové typy v metodách pro výpočet parciálních atomových nábojů

Bakalářská práce

Radka Sedláková

Vedoucí práce: RNDr. Tomáš Raček **Brno 2019**

Bibliografický záznam

Autor:	Radka Sedláková Přírodovědecká fakulta, Masarykova univerzita Ústav biochemie
Název práce:	Atomové typy v metodách pro výpočet parciálních atomových nábojů
Studijní program:	Biochemie
Studijní obor:	Chemoinformatika a bioinformatika
Vedoucí práce:	RNDr. Tomáš Raček
Akademický rok:	2018/2019
Počet stran:	počet stran od druhé stránky dokumentu po poslední stránku obsahu + počet stránek práce od první strany Úvodu
Klíčová slova:	atomy, magie

Bibliographic Entry

Author:	Radka Sedláková Faculty of Science, Masaryk University Department of Biochemistry
Title of Thesis:	Atom types in methods for calculation of partial atomic charges
Degree Programme:	Biochemistry
Field of Study:	Chemoinformatics and Bioinformatics
Supervisor:	RNDr. Tomáš Raček
Academic Year:	2018/2019
Number of Pages:	počet stran od druhé stránky dokumentu po poslední stránku obsahu + počet stránek práce od první strany Úvodu
Keywords:	atoms, magic

Abstrakt

Parciální atomové náboje

Abstract

In this thesis we study

Místo tohoto listu vložte kopii oficiálního (podepsaného) zadání práce.

Poděkování

Na tomto místě bych chtěla poděkovat stránce LOTR University Memes za zlepšování nálady ve chvílích největší bezradnosti.

Prohlášení

Prohlašuji, že jsem svoji bakalářskou práci vypracovala samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 15. května 2019

....
Radka Sedláková

Obsah

Přehled použitého značení

Pro jednodušší orientaci čtenářů v textu bakalářské práce uvádím seznam zkratek, které jsou v textu použity.

$\mathbb{C}$	množina všech komplexních čísel
$\mathbb{R}$	množina všech reálných čísel
$\mathbb{Z}$	množina všech celých čísel
$\mathbb{N}$	množina všech přirozených čísel

Kapitola 1

Úvod

Distribuce elektronů v m

Kapitola 2

Teorie

2.1 Parciální atomové náboje

Parciální atomové náboje [1] jsou reálná čísla, která popisují asymetrické rozložení elektronové hustoty na chemické vazbě. Vznikají v důsledku rozdílných elektronegativit vazebných partnerů. Pokud v chemické vazbě figuruje vysoce elektronegativní atom, pak tento k sobě přitahuje vazebný elektronový pár, čímž se zvyšuje elektronová hustota v jeho okolí a dochází ke vzniku parciálního záporného náboje (δ-). V okolí elektropozitivnějšího vazebného partnera se elektronová hustota naopak snižuje a na atomu dochází ke vzniku parciálního kladného náboje (δ+).

Koncept parciálních atomových nábojů je pouze teoretický, hodnoty nábojů proto nelze získat pomocí experimentu [2]. Jelikož jsou ale významným faktorem pro predikci fyzikálních, chemických a biologických vlastností molekul, bylo pro jejich stanovení vyvinuto množství výpočetních metod. Tyto se dělí na metody kvantově-mechanické a metody empirické. Kvantově-mechanické metody poskytují přesnější výsledky, ovšem za cenu vysoké časové náročnosti. Empirické metody dosahují v porovnání s QM metodami velmi dobrých výsledků, a to ve výrazně kratším čase. Žádná z vyvinutých empirických metod však není uznána za všeobecně platnou a použitelnost konkrétních metod se hodnotí na základě reprodukovatelnosti výsledků [3].

Aplikaci parciálních atomových nábojů lze nalézt ve výpočetní chemii a chemoinformatice, kde slouží k predikci elektrostatických a termodynamických vlastností popisujících reaktivitu molekul. Uplatňují se v molekulových simulacích [4], ve virtuálním screeningu [5], při hledání vazebných míst proteinů nebo při návrhu farmakoforů. Prokázaly se jako platné deskriptory v QSAR a QSPR modelech [6, 7]. V anorganické chemii se uplatňují při popisu toku elektronů v bateriích a katalyzátorech [8].

2.2 Kvantově-mechanické metody

Kvantově-mechanické metody pro výpočet parciálních atomových nábojů jsou založeny na poznatcích kvantové mechaniky. Dělí se na tři hlavní skupiny, a to metody semi-empirické, metody odvozené od teorie funkcionálu hustoty a metody *ab initio*. *Ab initio* metody (lat. *ab initio* - od počátku) staví výpočty na teoretickém aparátu a k řešení Schrödingerovy rovnice přistupují numericky, z čehož vyplývá jejich velká výpočetní náročnost. Metody semi-empirické jsou stejně jako metody *ab initio* založeny na řešení SE, pro zjednodušení výpočtů ale využívají kromě značné míry aproximací také data z experimentu.

Limitujícím faktorem pro použití QM metod je jejich složitost, konkrétně pro *ab initio* metody až $O(B^4)$, kde B je číslo rovno počtu elektronů v molekule nebo větší.

2.2.1 Základy kvantové mechaniky

K rozvoji kvantové mechaniky došlo ve 20. letech 20. století v reakci na newtonovskou mechaniku, jejíž aparát nepostačoval pro popis mikrosvěta. Základním principem QM je vlnově-korpuskulární dualismus. Vlna je v kvantové mechanice reprezentována matematickou funkcí Ψ, tzv. vlnovou funkcí, která popisuje dynamický stav částice a nese veškeré informace, které lze o částici získat [9].

Základním úkolem kvantové mechaniky je výpočet vlnové funkce systému. Vlnová funkce je řešením Schrödingerovy rovnice

$$H\Psi = E\Psi$$

kde H je operátor Hamiltonián a E je energie systému. Hamiltonián bere na vstup vlnovou funkci Ψ a transformuje ji na funkci jinou. Řešením Schrödingerovy rovnice je soubor funkcí, které lze po aplikaci Hamiltoniánu zapsat jako součin původní funkce a skaláru E. Takovéto funkce označujeme jako *vlastní funkce* a odpovídající skaláry jako *vlastní hodnoty* operátoru [10].

Schrödingerova rovnice je exaktně řešitelná pouze pro vybrané problémy. Jedním z nich je atom vodíku. Pro víceelektronové systémy je nutno do výpočtu zavádět velké množství aproximací, z nichž nejznámější je Born-Oppenheimerova aproximace. Jejím základním konceptem je oddělení pohybu jader atomů od pohybu elektronů, vycházející z předpokladu, že jádra, mnohonásobně těžší než elektrony, se pohybují výrazně pomaleji než elektrony samotné. Řešení Schrödingerovy rovnice

$$H\Psi(r,R) = E\Psi(r,R)$$

se tak rozkládá na řešení popisující elektrony v souboru fixních jader, po němž následuje řešení rovnice zahrnující kinetickou a potenciální energii jader obklopených polem elektronů. $\Psi(r,R)$ je vlnová funkce systému, závislá jak na souřadnicích elektronů (r), tak na souřadnicích jader (R).

2.2.2 Přehled kvantově-mechanických metod

Důležitým krokem *ab initio* a semi-empirických metod je výběr bázové sady. Bázová sada je soubor vlnových funkcí reprezentující atomové orbitaly, jejichž vhodnou lineární kombinací (LCAO) lze následně vyjádřit vlnovou funkci molekuly. Pro popis funkcí reprezentujících atomové orbitaly se používají orbitaly Gaussova typu (GTO). Kombinace několika Gaussových orbitalů přibližuje tzv. Slaterův orbital (STO), který je pro výpočet vlnové funkce molekuly méně vhodný z důvodu složitosti výpočtů. Bázových sad existuje nepřeberné množství, např. bázové sady STO-3G, STO-4G či obecně STO-nG, kde n je počet orbitalů Gaussova typu reprezentujících jeden atomový orbital.

Krokem vedoucím k výpočtu hodnot parciálních nábojů je provedení populační analýzy, která popisuje rozložení elektronové hustoty v molekule. Příkladem je Mullikenova populační analýza (*Mulliken Population Analysis*, MPA), která elektronovou hustotu určuje dle obsazenosti atomových orbitalů elektrony. V rámci chemické vazby je elektronová hustota rovnoměrně rozdělena mezi vazebné partnery, není tedy brána v potaz možná rozdílnost elektronegativit. Výsledky MPA jsou také silně závislé na použitém kvantově-mechanickém přístupu a na velikosti bázové sady. Nevýhody MPA, zejména nepřesnost výsledků související s rozšiřováním bázové sady, řeší přirozená populační analýza (*Natural Population Analysis*, NPA), pracující s přirozenými atomovými orbitaly. Přirozené atomové orbitaly jsou nejprve vypočteny z bázové sady a jsou následně použity pro výpočet ortonormálních přirozených vazebných orbitalů (*Natural bonding orbitals*, NBO). Na základě NBO se poté provadí populační analýza.

Odlišný přístup finálního výpočtu parciálních atomových nábojů představuje metoda *Atoms-in-Molecules* (AIM), která přiřazuje náboje atomům na základě integrace elektronové hustoty přes prostor příslušící danému atomu. Dalším možným přístupem je výpočet parciálních atomových nábojů na základě elektrostatických potenciálů molekuly (metody založené na *Molecular Electrostatic Potential-derived charges*, MEP).

Hartree-Fockova metoda

Problém řešení víceelektronových systémů nastává při zahrnutí elektronových interakcí do výpočtu. Hartree-Fockova metoda rozkládá původní problém n-elektronové Schrödingerovy rovnice na řešení n jednoelektronových rovnic. Využívá přiblížení pomocí metody nezávislých částic (*Self-Consistent Field*, SCF), která pracuje s modelem elektronu pohybujícím se v průměrném poli ostatních elektronů. HF metoda tedy nezahrnuje korelaci pohybu elektronů. Jednoelektronové rovnice jsou určeny předpisem

$$\hat{F}\chi_i = \varepsilon_i \chi_i \tag{2.1}$$

kde Fockův operátor $\hat{F}$ je Hamiltonián aplikovaný na jednoelektronový (atomový nebo molekulový) orbital a ε_i je odpovídající Langrangeův multiplikátor. Metoda pracuje iterativně a konečné řešení rovnic určuje na základě ustálení výsledků jednotlivých iterací výpočtu.

DFT

Metody založené na teorii funkcionálu hustoty (*Density Functional Theory*, DFT) nevycházejí z řešení vlnové funkce, ale poznatky staví na rozložení elektronové hustoty v molekule, ze které následně odvozují energii systému a další vlastnosti molekuly. Do výpočtů DFT metod je narozdíl od Hartee-Fockovy metody zahrnuta i korelační energie elektronů. Ve výpočtech figurují pouze tři neznámé (souřadnice x, y, z), zatímco řešení Schrödingerovy rovnice obsahuje $4n$ neznámých, kde n představuje počet elektronů systému. Metody odvozené od DFT jsou tak výpočetně méně náročné a poskytují přesnější výsledky.

Jedním z cílů DFT metod je výpočet celkové energie elektronů na základě elektronové hustoty. *Funkcionál* je v rámci DFT metod chápán jako zobrazení, které zobrazuje funkci, představující elektronovu hustotu, do množiny reálných čísel popisujících energii elektronů. Říkáme, že energie elektronů je funkcionálem elektronové hustoty.

Semiempirické metody

QM metody byly v době svého vzniku limitovány nedostatečnými výpočetními zdroji. Problém vyřešil rozvoj semi-empirických metod, které část výpočtů parametrizují nebo aproximují na základě experimentálních dat, přičemž se snaží přiblížit QM výpočtům.

Raná semi-empirická metoda CNDO (*Complete Neglect of Differential Overlap*) využívá teorii SCF pro popis elektronových interakcí a je založena na ZDO aproximaci (*Zero Differential Overlap*). Ta zamítá interakce atomových orbitalů lokalizovaných na různých atomech molekuly a pracuje pouze s interakcemi atomových orbitalů stejného typu lokalizovaných na stejném atomu. Tyto hrubé aproximace nahradila metoda INDO (*Intermediate Neglect of Differential Overlap*) zahrnutím interakcí odlišných typů atomových orbitalů. Dalšími semi-empirickými metodami jsou např. NDDO, MNDO, PM3 nebo SAM1, založené na MNDO.

2.3 Empirické metody

Empirické metody výpočtu parciálních atomových nábojů byly vyvinuty v reakci na velkou výpočetní náročnost QM metod. V porovnání s QM metodami dosahují velmi přesných výsledků, a to ve výrazně kratším čase. Empirické metody se dělí na dvě hlavní skupiny, a to metody pracující s topologií molekuly (jinak řečeno s její 2D strukturou) a metody pracující s prostorovým uspořádáním molekuly. Metody zastupující obě uvedené skupiny, jmenovitě metoda PEOE a metoda EEM, jsou popsány v odstavcích níže.

2.3.1 PEOE

Metoda PEOE (*Partial Equalization of Orbital Electronegativity*), známá také pod jménem autorů jako metoda Gasteiger-Marsili, byla poprvé publikována v roce 1980 [c]. Metodu lze aplikovat pouze na systémy obsahující σ vazby a nekonjugované π vazby, později však byla autory rozšířena i o výpočet systémů konjugovaných π vazeb. V rámci metody není uvažována 3D struktura molekuly, pracuje se pouze s její topologií.

Koncept elektronegativity atomových orbitalů, na němž je metoda založena, vychází z Mullikenovy definice elektronegativity χ_A atomu A

$$\chi_A = \frac{1}{2}(I_A + E_A) \tag{2.2}$$

Dle Mullikena je elektronegativita atomu určena hodnotami elektronových afinit E_A a ionizačních potenciálů I_A jeho valenčních stavů. PEOE připisuje na základě hodnot I_A a E_A elektronegativitu každému orbitalu valenčního stavu atomu. Elektronegativita χ_{iv} orbitalu iv na atomu i

$$\chi_{iv} = a_{iv} + b_{iv}Q_i + c_{iv}Q_i^2 \tag{2.3}$$

je ovlivněna náboji ostatních orbitalů a tedy i celkovým nábojem příslušného atomu Q_i. Koeficienty a_{iv}, b_{iv} a c_{iv} jsou empirické parametry vypočtené z ionizačních potenciálů a elektronových afinit neutrálního, kationtového a aniontového stavu příslušného orbitalu. (za PEOE vložit tabulku parametrů abc, viz Gasteiger s. 330)

Při vzniku vazby dochází vlivem elektronegativity atomů k přesunu elektronů od elektropozitivnějšího atomu směrem k elektronegativnějšímu. Interagují spolu příslušné atomové orbitaly a dochází k částečné ekvalizaci (vyrovnání) jejich nábojů. Množství přeneseného náboje mezi atomy A a B v k-té iteraci výpočtu, kde atom B má vyšší elektronegativitu, je definováno jako

$$Q^{\langle k \rangle} = \frac{\chi_B^{\langle k \rangle} - \chi_A^{\langle k \rangle}}{\chi_A^+} \cdot \alpha^k \tag{2.4}$$

kde χ_A^+ označuje elektronegativitu kationtu atomu A. Iniciální výpočet elektronegativity orbitalu (2.3) pracuje s formálním nábojem atomu. Po výpočtu příspěvků přenesených nábojů (2.4) všech vazebných partnerů atomu je náboj daného atomu přepočítán a použit v další iteraci. Množství přeneseného náboje mezi dvěma atomy se v každé iteraci výpočtu snižuje a vypočtené hodnoty nábojů atomů postupně konvergují. Přibližně po šesté iteraci dochází k ustálení výpočtů.

Díky přesnosti a rychlosti výpočtů byla PEOE implementována do většiny programů pro molekulové modelování jako základní metoda výpočtu atomových nábojů. Reziduální elektronegativita atomů, získaná na základě PEOE, se prokázala vhodnou pro popis indukčního efektu v molekulách.

2.3.2 EEM

Autoři Mortier, Ghosh a Shankar publikovali metodu elektronové ekvalizace (*Electronegativity Equalization Method*, EEM) v r. 1986. Teoretický základ metody vychází z teorie funkcionálu hustoty, na němž je vystavěn matematický aparát pro výpočet atomových nábojů. EEM využívá referenční náboje získané kvantově-mechanickými metodami, pomocí nichž parametrizuje vlastní výpočty. Díky nízké výpočetní náročnosti ($\Theta(N^3)$, kde N je počet atomů systému) a poměrně přesným výsledkům se stala metoda hojně využívanou a byla použita např. pro výpočet parciálních nábojů zeolitů, organických molekul nebo polypeptidů.

Výchozím konceptem metody je Sandersonův princip ekvalizace elektronegativity. Dle něj je každému atomu přiřazena stejná elektronegativita jako je elektronegativita ostatních atomů molekuly. Podle rovnice

$$\overline{\chi} = \chi_1 = \chi_2 = \chi_3 = \ldots = \chi_N \tag{2.5}$$

kde N je počet atomů, se elektronegativita každého atomu rovná průměrné elektronegativitě molekuly $\overline{\chi}$. Sandersonův postulát je potvrzen principy DFT.

Další základním principem metody je princip zachování náboje. Celkový náboj molekuly (Q) odpovídá součtu dílčích atomových nábojů q_i.

$$\sum_i q_i = Q \tag{2.6}$$

Třetí základní princip představuje efektivní elektronegativita χ_i atomu *i*. Jelikož metoda pracuje s prostorovým uspořádáním molekuly, započítává se při určování elektronegativity atomu *i* jeho molekulové okolí. Sumace ve vzorci reprezentuje elektrostatickou interakci atomu *i* s dalšími *N* atomy *j* molekuly v závislosti na jejich vzdálenosti R_{ij}.

$$\chi_i = A_i + B_i \cdot q_i + \kappa \sum_{j \neq i}^{N} \frac{q_j}{R_{ij}} \tag{2.7}$$

V rovnici kromě nábojů q_i, q_j interagujících atomů vystupují empirické parametry A_i, B_i a κ. Parametry A_i a B_i zahrnují elektronegativitu χ_i^0 a tvrdost η_i^0 neutrálního izolovaného atomu a korekce $\Delta\chi_i$, $\Delta\eta_i$, které upravují výslednou elektronegativitu χ_i atomu na základě jeho interakce s prostředím molekuly.

$$A_i = \chi_i^0 + \Delta\chi_i \tag{2.8}$$

$$B_i = 2(\eta_i^0 + \Delta\eta_i) \tag{2.9}$$

Cílem metody je empiricky nalézt hodnoty korekcí pro definované atomové typy a zajistit tak znovupoužitelnost uvedených parametrů.

Řešení systému s *N* atomy vede po kombinaci vztahů 2.5, 2.6 a 2.7 na systém *N*+1 lineárních rovnic o *N*+1 neznámých (q_1, q_2, q_3, ... , q_N, $\overline{\chi}$).

$$\begin{pmatrix} B_1 & \frac{\kappa}{r_{1,2}} & \cdots & \frac{\kappa}{r_{1,n}} & -1 \\ \frac{\kappa}{r_{1,2}} & B_2 & \cdots & \frac{\kappa}{r_{2,n}} & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\kappa}{r_{1,n}} & \frac{\kappa}{r_{2,n}} & \cdots & B_n & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \\ \overline{\chi} \end{pmatrix} = \begin{pmatrix} -A_1 \\ -A_2 \\ \vdots \\ -A_n \\ Q \end{pmatrix} \tag{2.10}$$

Po předchozí parametrizaci (sekce 2.3.3) získáváme z matice hodnoty parciálních atomových nábojů, které jsou následně porovnány s referenčními QM náboji. Proces uzavírá statistické vyhodnocení vypočtených dat.

Na principu EEM byly později vyvinuty další empirické metody jako např. *Atom-Bond Electronegativity Equalization Method* (ABEEM) nebo *General Bond Electronegativity Equalization Method* (GBEEM).

2.3.3 Parametrizace

Parametrizace je základním nástrojem empirických modelů, jejichž cílem je reprodukce experimentálních dat. Parametrizovat lze silová pole pro výpočty molekulové mechaniky nebo také *ab initio* výpočty zahrnující korelační energii elektronů. Cílem parametrizace je nalezení hodnot parametrů, po jejichž integraci do empirického modelu je dosaženo co nejlepší shody experimentálních a empirických výpočtů.

Parametrizace empirických metod pro výpočet parciálních atomových nábojů se skládá z následujících kroků:

1. Výběr tréninkové sady molekul obsahující atomy, které v dostatečné míře reprezentující atomové typy, jež chceme parametrizovat

2. Výpočet parciálních atomových nábojů tréninkové sady pomocí QM
3. Parametrizace tréninkové sady na základě nábojů získaných v kroku 2
4. Výpočet nábojů testovací sady molekul kvantově-mechanickou a parametrizovanou empirickou metodou
5. Statistické vyhodnocení parametrizace

Cílem EEM je na základě znalosti parciálních nábojů odvozených z QM určit hodnoty parametrů A_i, B_i a κ (viz 2.10) pro každý atom molekuly. Pro zjednodušení výpočtu jsou odpovídající řádky matice (atomy molekuly) sloučeny pod jeden atomový typ, pro který jsou počítány výše zmíněné parametry. Ty jsou na základě znalosti dílčích atomových nábojů (q_1, q_2, q_3, ... , q_N) a elektronegativity molekuly $\overline{\chi}$ získány z rovnice 2.7 upravené na tvar

$$A_i + B_i \cdot q_i = \overline{\chi} - \kappa \sum_{j \neq i}^{N} \frac{q_j}{R_{ij}} \tag{2.11}$$

Pro každou hodnotu parametru κ ze zvoleného intervalu jsou hledány vhodné hodnoty parametrů A_ia B_i. Každému atomovému typu je pak přiřazena trojice hodnot A_i, B_i a κ, která nejlépe reprodukuje náboje získané QM výpočty.

2.3.4 Atomové typy

Atomové typy jsou nástrojem parametrizace empirických metod výpočtu parciálních atomových nábojů. Definice atomových typů souvisí s charakteristickými chemickými vlastnostmi atomů (hybridizace, vazebný partner, nejvyšší řád vazby), které tyto atomové typy popisují. Každý atomový typ je definován takovou charakteristikou atomu, na základě které atom vykazuje odlišné chemické vlastnosti od jiných atomů, a je tedy vhodné pro něj definovat samostatný atomový typ. Názorným příkladem může být separace atomu uhlíku do tří atomových typů na základě jeho hybridizace. Prostorové uspořádání orbitalů uhlíku je v hybridizacích sp^3, sp^2, sp diametrálně odlišné a předurčuje tak tvorbu odlišných typů vazeb s vazebnými partnery.

Napříč parametrizacemi empirických metod zaznamenáváme různé úrovně návrhu atomových typů, od triviálních klasifikací definujících atomové typy na základě protonových čísel až po komplexní rozdělení zahrnující nabité funkční skupiny či příslušnost k větším atomovým celkům (aromatické systémy, postranní řetězce aminokyselin). Detailní dělení atomových typů nalézáme zejména v publikacích orientujících se na parametrizaci metod pro výpočet parciálních nábojů komplexních celků, např. polypeptidů.

2.4 Statistické pojmy

2.4.1 Průměrná a maximální absolutní odchylka

Průměrná absolutní odchylka (ang. *Mean Absolute Error*, MAE) je dána aritmetickým průměrem absolutních hodnot rozdílů hodnot x_i a y_i příslušných náhodných veličin.

Po odstranění absolutních hodnot by vzorec popisoval tzv. *Mean Bias Error* (MBE).

$$MAE(X,Y) = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{2.12}$$

Maximální absolutní odchylka popisuje největší rozdíl nalezený mezi hodnotami x_i a y_i náhodných veličin X a Y.

$$ABSMAX(X,Y) = max|x_i - y_i| \tag{2.13}$$

2.4.2 RMSD

Veličina RMSD (z anglického *Root Mean Square Deviation*, někdy uváděná též jako *Root Mean Square Error*) popisuje míru odlišnosti dvojic odpovídajících hodnot (x_i,y_i) náhodných veličin X a Y napříč datovým souborem. Je definována jako odmocnina ze střední kvadratické chyby (*Mean Square Deviation*, MSD). Stejně jako rozptyl je tato veličina kvůli kvadrátu rozdílu hodnot x_i a y_i citlivá na odlehlé a chybné hodnoty, které se promítají do vyšších výsledných hodnot RMSD srovnávaných datových sad.

$$RMSD(X,Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.14}$$

2.4.3 Pearsonův korelační koeficient

Pro kvantifikaci funkčního vztahu dvou sledovaných veličin užíváme tzv. Pearsonův korelační koeficient (*Pearson Correlation Coefficient*, PCC). PCC popisuje míru linearity závislosti veličiny Y na veličině X (lineární korelaci), pro popis jiných typů závislostí (např. kvadratických) není vhodný. Je definován jako

$$r(X,Y) = \frac{\sum_{i=1}^{n}((x_i - \overline{x})(y_i - \overline{y}))}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{2.15}$$

kde hodnoty x_i, ..., x_n, y_i, ...,y_n jsou i-té prvky dvourozměrného náhodného vektoru o velikosti n realizovaného dvěma náhodnými veličinami X a Y, a kde $\overline{x}$, $\overline{y}$ jsou aritmetické průměry naměřených hodnot veličin X a Y.

PCC nabývá hodnot z intervalu $\langle -1, 1 \rangle$, přičemž hodnoty koeficientu blízké číslu -1 nebo 1 indikují silnou lineární korelaci mezi pozorovanými veličinami. Linearita vztahu je dobře pozorovatelná v grafu (viz obr x.x.x), kde jsou dvojice hodnot (x_i,y_i) znázorněny jako body v dvourozměrné soustavě souřadnic. Interpretace hodnoty k Pearsonova korelačního koeficientu je následující:

- pokud je k kladné, pak veličiny X a Y vykazují kladnou korelaci (pokud se hodnota veličiny Y zvětšuje, pak hodnota X roste)
- pokud je k záporné, pak veličiny X a Y vykazují zápornou korelaci (v závislosti na zvětšující se hodnotě veličiny Y hodnota X klesá)
- pokud je k rovno 0, pak veličiny X a Y nejsou lineárně korelované

Kapitola 3

Metody

3.1 Vstupní data (Formát souborů)

3.2 (SMILES/SMARTS - jestli je použiju)

3.3 RDKit

3.4 NumPy

3.5 MatPlotLib? Co mi Ondra ukazoval jeho program, tak ten grafy taky tvoří

3.6 Ondrův program

Kapitola 4

Implementace

4.1 Klasifikátory (Co jiného?)

Kapitola 5

Výsledky a diskuse

5.1 Výsledky parametrizace

Jedním z cílů této bakalářské práce je otestovat, jak jemná či hrubá klasifikace atomových typů je postačující pro parametrizaci výpočetních metod poskytujících ve srovnání s QM metodami kvalitní výsledky. V souvislosti s parametrizací je též testováno, zda jsou některé charakteristiky atomu (např. hybridizace, nejvyšší řád vazby, vazebný partner) pro definici atomových typů vhodnější než charakteristiky jiné.

5.1.1 Klasifikátor1

5.1.2 Klasifikátor2

5.1.3 Klasifikátor3

5.2 Srovnání navržených klasifikátorů

Kapitola 6

Závěr

Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr.

Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr. Zde můžete napsat závěr.

Příloha

Sem můžete přidat přílohu. Pokud chcete "Přílohy", tak upravte definici záhlaví v souboru sci.muni.thesis.sty, viz příkaz `\HlavickaPriloha`**.**

Seznam použité literatury

[1] ATKINS, P. W. & DE PAULA, J. *Atkins' physical chemistry*. 9th ed. Oxford: Oxford University Press, c2010. ISBN 978-0-19-954337-3.

[2] LEACH, A. R. *Molecular modelling: principles and applications*. 2nd ed. New York: Prentice Hall, 2001. ISBN 0-582-38210-6.

[3] GASTEIGER, J. & ENGEL, T. *Chemoinformatics: A Textbook*. Weinheim: Wiley-VCH, c2003. ISBN 978-3-527-30681-7.

[4] RAPPE, A. K. & GODDARD, W. A. „Charge equilibration for molecular dynamics simulations". *The Journal of Physical Chemistry*. 1991, **95**(8), 3358-3363. DOI: 10.1021/j100161a070. ISSN: 0022-3654. URL: http://pubs.acs.org/doi/abs/10.1021/j100161a070

[5] LYNE, P. D. „Structure-based virtual screening: an overview". *Drug Discovery Today*. 2002, **7**(20), 1047-1055. DOI: 10.1016/S1359-6446(02)02483-2. ISSN: 13596446. URL: http://linkinghub.elsevier.com/retrieve/pii/S1359644602024832

[6] GHAFOURIAN, T., DEARDEN, J. C. & KATRITZKY, A. R. „The Use of Atomic Charges and Orbital Energies as Hydrogen-bonding-donor Parameters for QSAR Studies: Comparison of MNDO, AM1 and PM3 Methods". *Journal of Pharmacy and Pharmacology*. 2000, **52**(6), 1027-1044. DOI: 10.1211/0022357001774435. ISSN: 00223573. URL: http://doi.wiley.com/10.1211/0022357001774435

[7] KARELSON, M., LOBANOV, V. S. & KATRITZKY, A. R. „Quantum-Chemical Descriptors in QSAR/QSPR studies". *Chemical Reviews*. 1996, **96**(3), 1027-1044. DOI: 10.1021/cr950202r. ISSN: 0009-2665. URL: https://pubs.acs.org/doi/10.1021/cr950202r

[8] WANG, B., LI, S. L. & TRUHLAR. D. G. „Modeling the Partial Atomic Charges in Inorganometallic Molecules and Solids and Charge Redistribution in Lithium-Ion Cathodes". *Journal of Chemical Theory and Computation*. 2014, **10**(12), 5640-5650. DOI: 10.1021/ct500790p. ISSN 1549-9618. URL: http://pubs.acs.org/doi/10.1021/ct500790p

[9] CELÝ, J. *Základy kvantové mechaniky pro chemiky: I. Principy*. Brno: Rektorát UJEP Brno, 1986

[10] JEAN, Y., VOLATRON, F. & BURDETT, J. K. *An introduction to molecular orbitals.* New York: Oxford University Press, 1993. ISBN 0-19-506918-8.