

Predikce krystalových struktur se zapojením experimentálních dat

Milan Kočí¹, Jan Drahokoupil¹

¹Katedra inženýrství pevných látek, Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické v Praze

Abstrakt

V této práci se věnujeme studiu energetických ploch potenciálních krystalických struktur a algoritmu „Optimalizace hejnem částic“. Pro daný problém, optimalizace souřadnic paracetamolu ve známé základní buňce, je pro tento algoritmus navržen optimální počet částic. Je zde ukázáno, že i nekvalitní prášková difrakční data v kombinaci s energetickým predikováním potenciálních struktur mohou vést k vyřešení krystalické struktury dané látky.

Klíčová slova: Predikce krystalových částic; Optimalizace hejnem částic; Prášková rentgenová difrakce.

Úvod

Predikování krystalických struktur na základě energetického hodnocení neslouží jenom k nalezení jediné „správné“ krystalické struktury, ale umožňuje nalézt řadu potenciálních kandidátů struktury, které mohou být, za určitých podmínek (teplota, tlak, proces vzniku), také stabilní. Samotné výpočty energie mohou být časově náročné nebo nemusí být dost přesné na to, aby určily výhodnost konkrétní struktury, proto mohou být do výpočtů započítávána také naměřená experimentální data.

Často používanými experimentálními daty jsou data ssNMR (pevnolátková nukleární magnetická rezonance). Výsledky ssNMR mohou být informace o počtu nezávislých molekul v základní buňce, konformacích molekul, tautomerech molekul a o chemickém chování vazeb molekul, které mohou zmenšit prohledávaný fázový prostor, a tak urychlit dobu hledání krystalické struktury látky [1].

Dalšími experimentálními daty, které se používají v kombinaci s predikcí krystalových struktur, jsou prášková rentgenová difrakční data (PXR data). V [2] byla navržena metoda FFPASS, ve které se krystalová struktura hledá pomocí genetického algoritmu, kde se místo uvažování absolutních hodnot míry shody a energie počítá pouze s pořadími struktur v generaci. Prášková rentgenová difrakční data propojená s energií vedou také k relativnímu prohloubení globálního minima společné funkce, toto minimum by mělo být snáze nalezeno [3].

Výsledkem predikce krystalových struktur je seznam krystalových struktur seřazený podle jejich klesající energie. Experimentální data lze použít na výběr konkrétní struktury z tohoto seznamu. V práci [4] autoři použili naměřená data ssNMR na identifikaci správného polymorfu látky ze seznamu predikovaných struktur.

V této práci rozvíjíme metodiku predikce krystalových struktur. Jsou zde validovány predikované struktury z [5] pomocí PXR dat a je zde optimalizován algoritmus optimalizace hejnem částic, který se používá pro urychlení hledání krystalových struktur.

Teorie & Počítačové nastavení

Úkolem predikce krystalových struktur je předpovědět ze znalosti modelu struktury látky možné krystalové struktury, ve kterých se látka může nacházet. Nejprve je potřeba vygenerovat si kandidáty krystalových struktur látky, tj. určit prostorovou grupu, parametry mřížky a rozložení a orientaci částic v základní buňce. Následuje optimalizace kandidátů, při které jsou hledány struktury s minimální hodnotou energie. V tomto kroku se používají různé algoritmy globální optimalizace, které urychlují hledání minim energie. Tato optimalizace se může opakovat víckrát, z důvodu získání přesnějších výsledků. Před uveřejněním seznamu struktur je potřeba provést závěrečnou lokální optimalizaci struktur, aby výsledná struktura měla opravdu minimální hodnotu energie.

Algoritmem, požívaným v této práci, je optimalizace hejnem částic (PSO). V tomto algoritmu myšlené hejno částic prochází fázovým prostorem určeným optimalizovanými stupni volnosti krystalové struktury, vzájemně se ovlivňuje a hledá strukturu s minimální hodnotou energie. Každá částice j je definovaná svou polohou \vec{x}_j a rychlostí \vec{v}_j , které se vyvíjí podle

$$\vec{v}_j(t+1) = w\vec{v}_j(t) + \alpha \cdot (\vec{m}_j(t) - \vec{x}_j(t)) + \alpha \cdot (\vec{M}(t) - \vec{x}_j(t)), \quad (1)$$

$$\vec{x}_j(t+1) = \vec{x}_j(t) + \vec{v}_j(t+1), \quad (2)$$

kde t je časový krok, w je parametr, který určuje závislost následující rychlosti na současné rychlosti, α je parametr udávající závislost na polohách předchozích nejlepších hodnot, \vec{m}_j je poloha nejlepší hodnoty nalezené částicí j a \vec{M} je poloha nejlepší hodnoty nalezené všemi částicemi.

Výpočty energie byly prováděny pomocí programu Materials Studio modulem na výpočet energie molekulární mechanikou Forcite s nastavením: silové pole COMPASS, kvalita výpočtů Ultra-fine – odpovídá konvergenci Ewaldovy sumace 10^{-5} kcal.mol $^{-1}$ a omezení sumace ne vazebných energií 18,5 Å.

Výpočty R_{wp} byly prováděny pomocí programu Materials Studio modulem Reflex Rietveldovou metodou za podmínek: rozsah difrakčního záznamu 5° až 45° 2θ ; krok difrakčního záznamu $0,015^\circ$ 2θ ; Braggova-Brentanova geometrie; zero-shift 0,013; profilové funkce Pseudo-Voightovy; profilové parametry $U = 0,02192$; $V = -0,01152$; $W = 0,00788$; $N_A = 0,1615$; $N_B = -0,00075$.

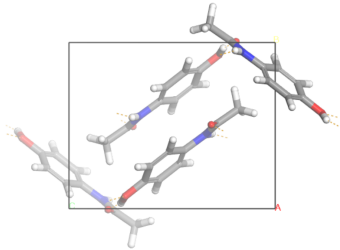
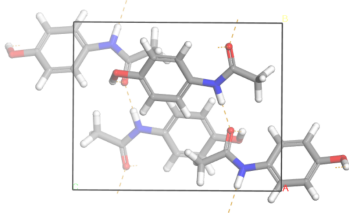
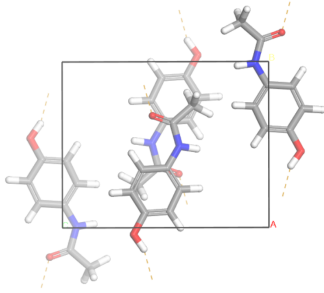
Výsledky a diskuze

Ověření správnosti predikovaných struktur

V práci [5] byla provedena predikce krystalové struktury paracetamolu [6]. Výsledkem byl seznam tří struktur, které mohli odpovídat reálné krystalové struktuře paracetamolu. Vypočítali jsme hodnotu faktoru R_{wp} pro strukturu nasimulovanou ze struktury [6] a predikované struktury. Výpočty potvrdili, že struktura s nejnižší hodnotou energie ze seznamu predikovaných struktur má také nejnižší hodnotu R_{wp} .

V další fázi jsme ověřili, že by tato validace predikovaných struktur fungovala i pro případ nanokrystalického rozšiřování difrakčních linií. Opět jsme počítali hodnotu faktoru R_{wp} , tentokrát jsme ovšem započítávali příspěvek velikosti krystalitů prášku. Velikost krystalitů byla 5 nm. Výpočty opět potvrdili, že struktura s nejnižší hodnotou energie má

Tabulka 1: Predikované struktury z práce [5] a jejich ohodnocení pomocí měř shod práškových difrakčních záznamů

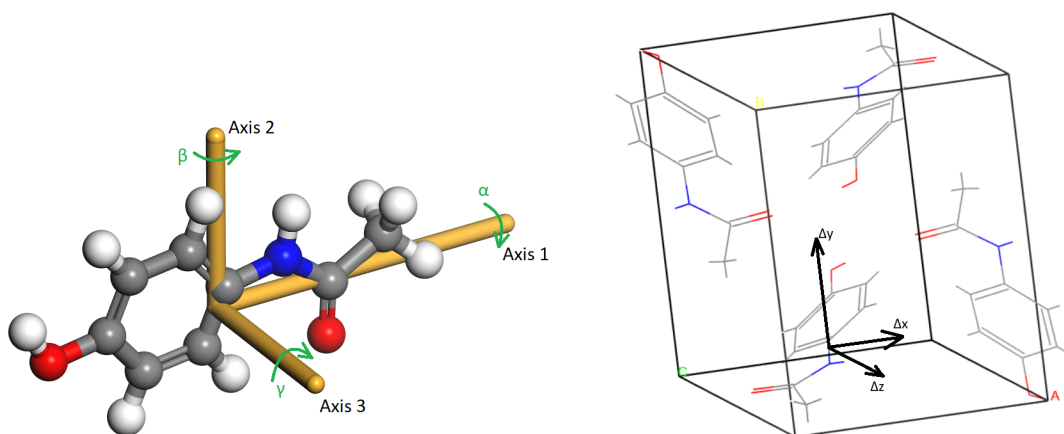
	Struktura	Energie [eV/atom]	R_{wp}^{ideal}	R_{wp}^{broad}
1.		-126,509	12,8	5,1
2.		-126,495	90,5	20,5
3.		-126,495	93,5	23,1

nejnižší hodnotu R_{wp} . V tabulce 1 jsou predikované struktury a k nim přiřazené hodnoty R_{wp} , kde R_{wp}^{ideal} odpovídá výpočtům bez započítání rozšiřování difrakčních linií a R_{wp}^{broad} odpovídá výpočtům se započtením rozšiřování difrakčních linií.

Trendy funkcí *Energie* a R_{wp}

Závislost funkcí *Energie* a R_{wp} na stupních volnosti krystalové struktury může být komplexní. Pro zjednodušení a získání představy o jejich trendech jsme vynesli jejich dvou-dimenzionální mapy. Jako stupně volnosti jsme brali vychýlení molekuly z rovnovážné polohy Δx , Δy a pootočení molekuly kolem vlastní osy, značíme γ . Na obrázku 1 jsou nakresleny stupně volnosti molekuly paracetamolu v krystalové struktuře.

Výsledkem jsou dvou-dimenzionální mapy, viz obrázky 2 a 3. Na funkci *Energie* je značný rozdíl mezi rotací molekuly a jejím posunem – zatímco při pohybu ve směru Δx se hodnota *Energie* mění minimálně, při rotaci molekuly o γ dochází k velkým změnám hodnot. Dále jsou hodnoty *Energie* uvedeny v logaritmickém měřítku z důvodu, že pro různé body mapy se hodnoty liší v řádech až 10^{18} kcal/mol. V lineárním zobrazení by byly



Obrázek 1: Rotační (vlevo) a translační stupně volnosti (vpravo)

změny hodnot *Energie* nerozlišitelné.

Funkce R_{wp} má dvě hluboká minima. Hlubší z nich odpovídá globálnímu minimu funkce a leží v rozích grafů. Druhé hluboké minimum odpovídá rotaci molekuly o 180° úhlu γ . Při translaci molekuly z minima funkce dochází ke zmenšování hloubky minim.

Optimalizace počtu částic algoritmu optimalizace hejnem částic (PSO)

Provedli jsme systematickou analýzu hodnot minim funkce R_{wp} , které našel algoritmus PSO, v závislosti na parametru počtu částic. Spustili jsme PSO pro počet částic N roven 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096 a 10000, omezili jsme počet výpočtů hodnoty funkce na $\lceil 10000 : N \rceil \cdot N$. Algoritmus jsme spustili desetkrát, abychom získali statistický výsledek.

V grafu 4 jsou průměrné nalezené hodnoty funkce. Je patrné, že závislost hodnoty minima na počtu částic má své minimum. Toto minimum jsme získali proložení hodnot funkcí

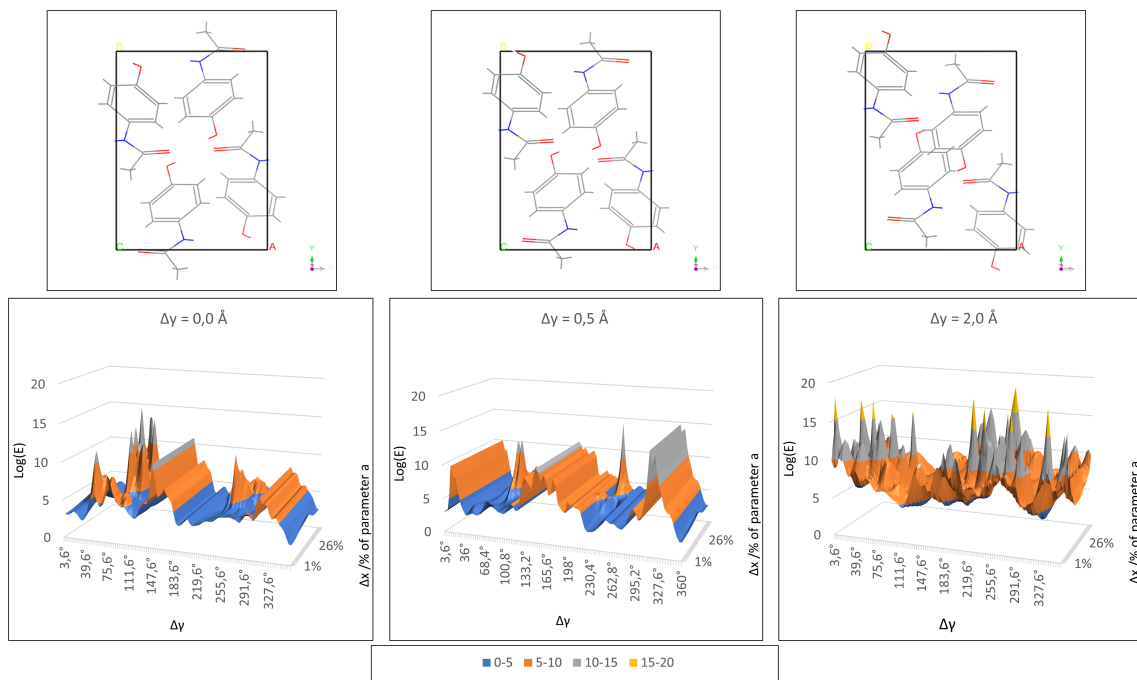
$$y(N) = D_e \left[\left(1 - e^{-a(N-N_0)} \right)^2 + 1 \right], \quad (3)$$

kde N je počet částic, D_e a a jsou parametry a N_0 je minimální hodnota počtu částic. Jako optimální nám vyšla hodnota $N=196$.

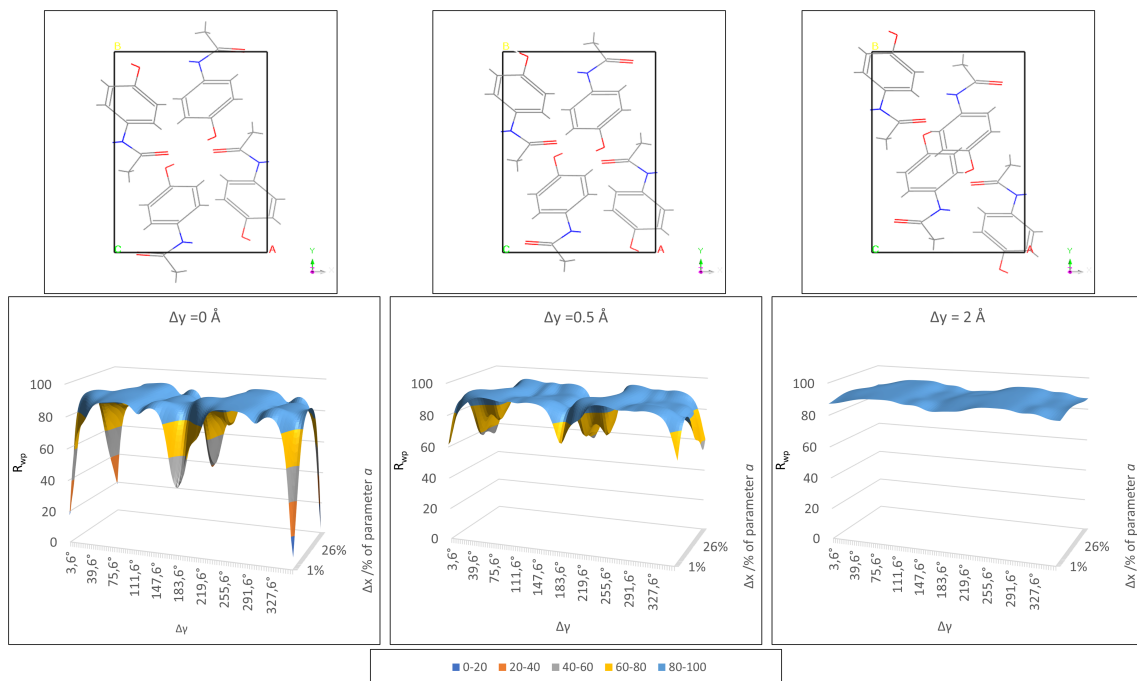
Nalezená optimální hodnota $N=196$ je vhodná pro tento systém, její využití pro jiné systémy by mohlo být otázkou do dalších výzkumů. Domníváme se, že přesná hodnota $N=196$ není příliš podstatná, ale důležité je, že by počet částic měl být v řádu nižších stovek, toto se domníváme z důvodu, že minimum hodnot je relativně široké a zaujímá oblast okolo nižších stovek počtu částic.

Závěr

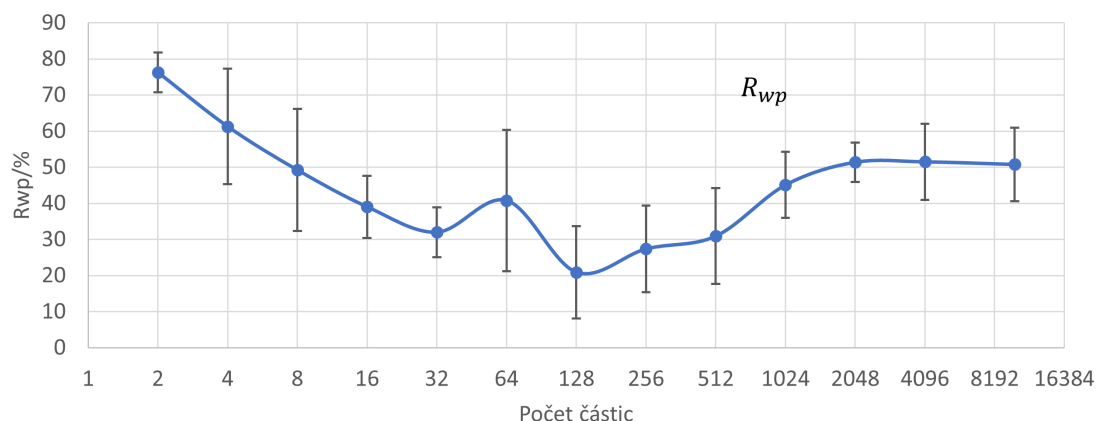
Ověřili jsme, že lze použít prášková rentgenová difrakční data na výběr krystalové struktury. Kvalitní data dosahovala lepších výsledků, ale použití dat s širokými difrakčními liniemi bylo také úspěšné. Získané trendy funkcí *Energie* a R_{wp} lze využít ke zrychlení predikce krystalových struktur. Dalšího vylepšení lze dosáhnout nastavením vhodných hodnot para-



Obrázek 2: Závislost *Energie* na stupních volnosti molekuly paracetamolu



Obrázek 3: Závislost R_{wp} na stupních volnosti molekuly paracetamolu



Obrázek 4: Závislost průměrné hodnoty nalezených minim na změnách nastavení parametru počtu částic algoritmu PSO

metrů optimalizačního algoritmu, jako se nám povedlo nastavením počtu částic algoritmu PSO na nízké stovky.

Reference

- [1] F. Bravetti et al. Solid-State NMR-Driven Crystal Structure Prediction of Molecular Crystals: The Case of Mebendazole. *Chemistry – A European Journal* **28**(6): e202103589, 2022.
- [2] B. Meredig, C. Wolverton. hybrid computational–experimental approach for automated crystal structure solution. *Nature Materials* **12** (2): 123-127, 2013.
- [3] Tsujimoto et al. Search for common minima in joint optimization of multiple cost functions. *Computer Physics Communications* **241**: 92-97, 2019.
- [4] G. A. Valdivia-Berroeta, K. Sarpal, N. C. Gonnella. OPTICS: Operant Probability Theory in Crystal Solutions, Application of ssNMR, and Probability Theory in Polymorph Identification. *Crystal Growth and Design* **21** (10): 5904-5913, 2021.
- [5] M. Kočí. Predikce krystalových struktur. *Bakalářská práce*. FJFI, ČVUT, Praha, 2021.
- [6] G. Nichols, C. S. Frampton. Physicochemical Characterization of the Orthorhombic Polymorph of Paracetamol Crystallized from Solution. *Journal of Pharmaceutical Sciences* **87** (6): 684-693, 1998.

Poděkování

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT č. SGS22/183/OHK4/3T/14.