

SKRIPT „baSHELIXir“ USNADŇUJE ŘEŠENÍ MAKROMOLEKULÁRNÍCH STRUKTUR POMOCÍ PROGRAMŮ SHELX C/D/E

PETR KOLENKO

*Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické, Břehová 7, 115 19 Praha 1
petr.kolenko@fffi.cvut.cz*

Došlo 18.3.19, přijato 15.7.19.

Klíčová slova: rtg. difrakce, experimentální fázování, optimalizace

Úvod

Stanovení počátečních fází představuje jeden z hlavních problémů v procesu řešení prostorové struktury nových makromolekul a jejich komplexů. V mnoha případech je rovněž obtížné určit prostorovou grupu krystalu přímo při zpracování naměřených difrakčních dat. Fázový problém však nemůže být úspěšně vyřešen v nesprávné prostorové grupě.

Pro účely experimentálního stanovení počátečních fází makromolekulárních struktur existuje mnoho výpočetních nástrojů. Tyto nástroje jsou dostupné například v programových balících CCP4 (cit.¹) nebo PHENIX (cit.²). Žádný z nich však neumožňuje automatické testování různých prostorových grup tak, jak je to například umožněno ve výpočetních nástrojích pro metodu molekulárního nahrazení. Programy MOLREP (cit.¹), PHASER (cit.²), BALBES (cit.³) a MORDA (cit.⁴) umožňují automatické hledání správného řešení molekulárního nahrazení v rámci zvolené Laueho třídy, nebo mezi enantiomorfními prostorovými grupami. Pouze programový balík PHENIX (cit.²) umožňuje pro případ experimentálního fázování hledat řešení u enantiomorfních prostorových grup současně, například $P4_12_12$ a $P4_32_12$.

Jeden z nejrozšířenějších nástrojů pro účely experimentálního fázování je programový balík SHELX C/D/E (cit.⁵). Programy umožňují snadnou a širokou parametrizaci výpočetních úloh a jsou jednoduše začlenitelné do dalších výpočetních nástrojů. Na základě těchto programů bylo navrženo například grafické rozhraní HKL2MAP (cit.⁶). Další uživatelské rozhraní pro SHELX C/D/E je dostupné v rámci programového balíku CCP4 a to i v jeho online variantě¹. Tato rozhraní však neumožňují automatizování výpočetních procedur, neboť jednotlivé běhy programů je potřeba spouštět odděleně a sledování vlivu změny jednotlivých parametrů na kvalitu výsledného řešení není přímočaré.

Cílem tohoto příspěvku je představit skript baSHELIXir, který umožňuje automatizaci výše zmíněných analýz při řešení experimentálního fázování makromolekulárních struktur pomocí programového balíku SHELX C/D/E.

Experimentální část

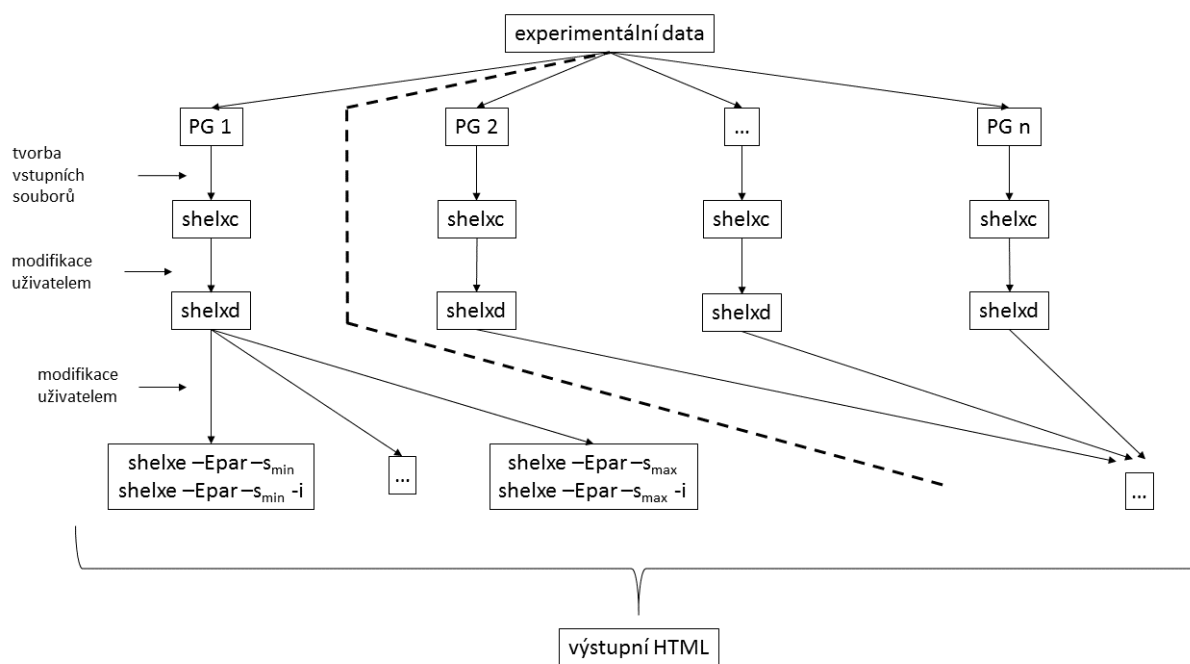
Skript baSHELIXir byl navržen v jazyku Bash s využitím programů GNUplot a SHELX C/D/E. Tento minimalistický přístup ohledně závislostí na dalších programech zajišťuje funkčnost na většině linuxových operačních systémů.

V prvním kroku programu je proveden pokus o načtení parametrů dat ze vstupních souborů (např. parametry základní buňky a vlnová délka primárního svazku). V následujícím kroku skript vytvoří adresáře pro veškeré prostorové grupy, ve kterých uživatel zadal hledat řešení fázového problému. Do těchto adresářů jsou dále nakopirována experimentální data a jsou v nich vytvořeny vstupní soubory pro výpočty procedury SHELX C. Po těchto přípravných krocích jsou v jednotlivých adresářích provedeny výpočty procedur SHELX C a D. Následně jsou provedeny výpočty procedury SHELX E na základě uživatelem zadaných parametrů, nebo je provedeno automatické hledání optimálního parametru obsahu solventu. Tato procedura je provedena pro obě enantiomorfní řešení a spočívá v mnoha cyklech procedury SHELX E s měnícím se parametrem obsahu solventu. Po úspěšném dokončení všech výpočetních procedur je vytvořena HTML stránka, která shrnuje výsledky jednotlivých výpočtů v různých prostorových grupách. Tato stránka poskytuje jednak všechny vstupní soubory a výstupní informace o průběhu jednotlivých výpočtů, ale rovněž grafické znázornění výsledků těchto procedur. Schéma jednotlivých výpočetních kroků skriptu baSHELIXir je uvedeno na obr. 1. Příklady grafických výstupů jsou uvedeny na obr. 2.

Skript baSHELIXir umožňuje automatizované spouštění programů softwarového balíku SHELX C/D/E se systematicky se měnícími vstupními parametry a usnadňuje tak řešení fázového problému pomocí metod:

- anomálního rozptylu s využitím jedné vlnové délky – SAD (Single-wavelength Anomalous Dispersion),
- anomálního rozptylu s využitím více vlnových délek – MAD (Multi-wavelength Anomalous Dispersion),
- fázování pomocí indukovaného radiačního poškození – RIP (Radiation Damage-Induced Phasing),
- izomorfního nahrazení s využitím anomálního rozptylu – SIRAS (Single Isomorphous Replacement with Anomalous Scattering).

Skript dále umožňuje měnit parametry výpočtů jako například počet disulfidových vazeb, minimální vzdálenost těžkých atomů a difrakční limit. Veškeré původní vstupní, výstupní soubory, záznamy z výpočtů i skripty pro vytvoření grafických výstupů pomocí programu GNUplot jsou ponechány v adresářích pro případ následného opětovného využití (například tvorba kvalitnějších obrázků pro publikaci).



Obr. 1. Schéma výpočetních procesů zahrnutých ve skriptu baSHELIXir. Vstupní soubory jsou tříděny do adresářů příslušným řešením v jednotlivých prostorových grupách (PG). Následují výpočty pomocí programů SHELX C a D. Hledání optimálního parametru obsahu rozpouštědla může být zahrnuto v závěrečných výpočtech pomocí programu SHELX E. Závěrem je vytvořena HTML stránka s odkazy na vstupní a výstupní soubory, která je doplněna grafickým výstupem znázorňujícím úspěšnost řešení fázového problému

Pro demonstraci funkčnosti skriptu bylo vybráno šest příkladů řešení fázového problému. Data k těmto příkladům jsou ve většině případů distribuována spolu s programy SHELX C/D/E (cit.⁵) a PHENIX (cit.²). Příklady aplikace skriptu baSHELIXir zahrnují hledání správné prostorové grupy, automatické načtení parametrů základní buňky a vlnové délky a automatické i uživatelem zadané hledání optimálního parametru obsahu solventu.

Skript baSHELIXir mimo jiné umožňuje uživatelům měnit rozlišení, na kterém jsou výpočty prováděny, počet disulfidových můstků, stanovit minimální vzdálenosti me-

zi těžkými atomy a mnoho dalších parametrů, které jsou vyžadovány programy SHELX C/D/E.

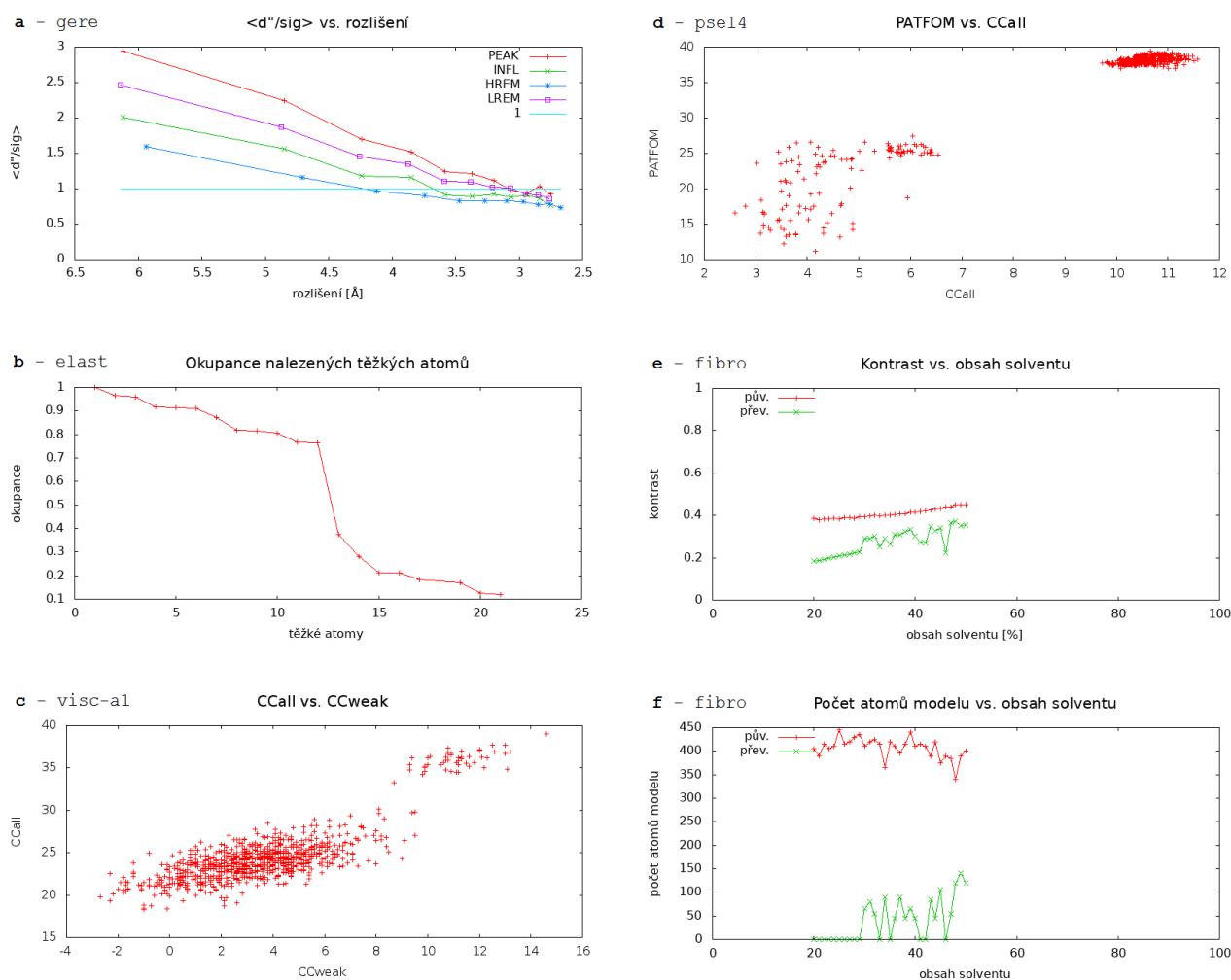
Výsledky a diskuse

Funkčnost skriptu baSHELIXir byla testována na šesti případech metody anomálního rozptylu. Ve všech případech byl fázový problém úspěšně vyřešen. Parametry výpočtů včetně testovaných funkcí skriptu jsou uvedeny v tab. I. Příklady grafických výstupů jsou uvedeny na obr. 2.

Tabulka I

Seznam testovacích projektů. Příklady byly vybrány tak, aby demonstrovaly automatické hledání správné prostorové grupy (visc-a1, elast, p9, pse14, gere), manuální zadání prostorové grupy (fibro), pevně definovaný obsah solventu (visc-a1, p9, pse14) a automaticky (gere, elast) či uživatelsky definovanou optimalizaci parametru obsahu solventu

Projekt	Metoda fázování	Bravaisova mřížka (příp. prostorová grupa)	Rozlišení [Å]	Obsah solventu	Lit.
visc-a1	SAD	tP	2,2	50 %	5
fibro	SAD	$P4_12_12$	2,5	20–50 %, krok 1 %	5
elast	SAD	oP	1,9	automaticky	5
p9	SAD	tI	2,2	60 %	2
pse14	SAD	oP	2,8	45 %	10
gere	MAD	mC	3,3	automaticky	5



Obr. 2. Příklady výstupních grafů skriptu baSHELIXir po úspěšném řešení fázového problému. a) Odhad anomálního signálu při různých vlnových délkách pomocí programu SHELX C (gere⁵). b) Závislost okupance jednotlivých těžkých atomů – výstup z programu SHELX D (elast⁵). c–d) Charakteristické grafy běžně užívané pro charakterizaci úspěšného řešení fázového problému (visc-al⁵, pse14¹⁰). e) Výsledný kontrast při jednotlivých výpočtech programu SHELX E pro obě enantiomerní řešení s proměnlivým parametrem obsahu solventu (fibro⁵). f) Závislost počtu modelovaných atomů pomocí automatické stavby modelu programem SHELX E na parametru obsahu solventu pro obě enantiomerní řešení (fibro⁵)

Řešení v různých prostorových grupách

Stanovení prostorové grupy při zpracování difrakčních dat není vždy jednoduchou záležitostí. Přitom úspěšné řešení fázového problému není možné bez znalosti správné prostorové grupy. K tomuto účelu bylo již vyvinuto několik výpočetních nástrojů, například AIMLESS (cit.⁷) nebo PHENIX.XTRIAGE (cit.²). Dalším nástrojem je kontrola systematických absencí samotným uživatelem. V některých případech je však obtížné vyhodnotit tyto parametry například z důvodu chybějících experimentálních dat (nevhodná orientace krystalu během měření, nízké rozlišení, šum). Proto je v mnoha případech prostorová grupa určena až na základě úspěšně vyřešeného fázového

problému a stabilního upřesňování takto získaného modelu. U výpočetních nástrojů pro experimentální fázování není dostupná možnost analýzy různých prostorových grup v rámci Laueho třídy jako u metody molekulárního nahrazení.

Skript baSHELIXir je navržen především tak, aby navazoval na zpracování difrakčních dat pomocí programu XDS (cit.⁸). Výstupní soubory z tohoto programu mají standardizovaný formát, což umožňuje jednoduché načtení parametrů základní buňky a vlnové délky primárního svazku. Podobně jako u XDS jsou ve skriptu baSHELIXir prostorové grupy seřazeny do skupin podle Bravaisových mříží. Uživatel však může libovolně učinit výběr prostorových grup dle vlastního uvážení v libovolném pořadí. Ně-

keré procesy mohou obsahovat spoustu výpočtů. Uživatel tak může upřednostnit některé prostorové grupy indikované nástroji jako AIMLESS a PHENIX.XTRIAGE, průběžně kontrolovat výsledky a případně přerušit běh programu, pokud již bylo výsledné řešení nalezeno. Funkčnost obou metod výběru prostorových grup je uvedena v tab. I.

Přestože některé Bravaisovy mříže obsahují pouze jedinou prostorovou grupu (např. P1 nebo C2), další vlastnosti skriptu baSHELIXir (např. optimalizace parametru obsahu solventu) činí skript nadále zajímavý pro některé druhy výpočtů.

Optimalizace parametru obsahu solventu

U obtížných případů může parametr obsahu solventu hrát důležitou úlohu při řešení fázového problému⁹. Hledání optimální hodnoty tohoto parametru je pomocí skriptu baSHELIXir prováděno nezávisle na hledání správné prostorové grupy.

Parametr obsahu solventu hraje důležitou úlohu také u metody molekulárního nahrazení. Počet hledaných molekul uvnitř asymetrické jednotky bývá obvykle odhadován na základě velikosti asymetrické jednotky základní buňky a odhadované velikosti jednotlivých komponent. Znalost parametru obsahu solventu u metody experimentálního fázování umožňuje vylepšení počátečních fází. Nicméně užití nesprávně stanoveného parametru obsahu solventu může vést k poškození počátečních fází. Podobně jako u hledání správné prostorové grupy, ani tento parametr není přesně znám, dokud není struktura vyřešena. Z toho důvodu bylo do běhu skriptu začleněno testování tohoto parametru. Tato analýza může být spuštěna automaticky a to s počáteční úrovní 20% obsahu solventu a postupným zvyšováním po 5% krocích až do úrovně 80 %. Pro potřeby detailnější analýzy je umožněno tyto tři parametry (nejnižší úroveň, šířka kroku a nejvyšší úroveň) libovolně měnit. Tato procedura však vyžaduje největší množství výpočetního času, není paralelizována, a proto je v rámci urychlení výpočtů doporučeno u snadných případů nevyužívat této možnosti.

Příklad molekuly fibronektinu (tab. I, obr. 2e) ukazuje, že rozdíl mezi hodnotou kontrastu (standardně uváděná hodnota) enantiomorfních řešení není dostatečně rozsáhlý k tomu, aby na první pohled přesvědčil uživatele, že fázový problém je úspěšně vyřešen. Nejen z tohoto důvodu byl do monitorovacího výstupu zahrnut údaj o množství namodelovaných atomů v rámci automatické stavby modelu v proceduře SHELX E. Pro pevně stanovený parametr obsahu solventu byla do závěrečné HTML stránky přidána textová řádka, pro automatickou optimalizaci parametru obsahu solventu byl vytvořen graf závislosti počtu modelovaných atomů na úrovni obsahu solventu (obr. 2f).

Dostupnost a spouštění skriptu

Skript baSHELIXir je dostupný na webové adrese <http://kmlinux.fjfi.cvut.cz/~kolenpe1/baSHELIXir>. Skript je spouštěn ve formě příkazu doplněného parametry po-

třebnými k výpočtům. Udrží filozofii a notaci svého hlavního výpočetního nástroje – SHELX C/D/E.

Závěr

Možnosti skriptu baSHELIXir a jeho úspěšnost řešení experimentálních fází nepřesahují možnosti programu SHELX C/D/E, ale nabízí širokou automatizaci výpočtů. Snadná parametrizace pomocí příkazové řádky rovněž umožňuje případnou vyšší úroveň automatizace, například hromadné použití na soubor mnoha zpracovaných difrakčních dat. Z důvodu předpokládaného využití v rámci komunity strukturních biologů i ze zahraničí je skript a jeho webová dokumentace psány v anglickém jazyce.

Tato práce byla podpořena grantem MŠMT ČR (CZ.02.1.01/0.0/0.0/16_019/0000778) a také Grantovou agenturou Českého vysokého učení technického v Praze (SGS16/246/OHK4/3T/14). Uživatelé jsou vyzváni k návrhům dalšího vylepšení skriptu baSHELIXir.

LITERATURA

1. Winn M. D. a 17 spoluautorů: Acta Crystallogr., Sect. D: Biol. Crystallogr. 67, 235 (2011).
2. Adams P. D. a 17 spoluautorů: Acta Crystallogr., Sect. D: Biol. Crystallogr. 68, 539 (2012).
3. Long F., Vagin A. A., Young P., Murshudov G. N.: Acta Crystallogr., Sect. D: Biol. Crystallogr. 64, 125 (2008).
4. Vagin A., Lebedev A.: Acta Crystallogr., Sect. A: Found. Crystallogr. 71, 19 (2015).
5. Sheldrick G. M.: Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 479 (2010).
6. Pape T., Schneider T. R.: J. Appl. Crystallogr. 37, 843 (2004).
7. Evans P. R.: Acta Crystallogr., Sect. D: Biol. Crystallogr. 67, 282 (2011).
8. Kabsch W.: Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 125 (2010).
9. Stránský J., Dohnálek J.: *Sborník příspěvků 4. studentské vědecké konference fyziky pevných látek, June 23. – 27. 2014, Nové Hradky – Czech Republic*, str. 67, Praha 2014.
10. Zahradník J., Kolářová L., Pařízková H., Kolenko P., Schneider B.: Fish Shellfish Immunol. 79, 140 (2018).

P. Kolenko (Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague): **baSHELIXir: A Bash Script for Automated Experimental Phasing Using Shelx C/D/E**

Estimation of initial phases represents one of the major problems in the structure determination process of novel macromolecules by single-crystal X-ray analysis. In this

study, a simple command-line tool with minimal software dependencies for a wide analysis of phasing information using the SHELX C/D/E software package, baSHELiXir, is presented. The novelty of the tool is in the parallel analysis in multiple space groups as well as in the screening for the optimal solvent content parameter. All results and graphs, including the input, output and log files, are summarized on an HTML page. The tool has been thoroughly tested on a number of available cases.

Keywords: X-ray diffraction, experimental phasing, optimization

Acknowledgement

This work was supported by the MEYS CR (grant No. CZ.02.1.01/0.0/0.0/16_019/0000778) and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/246/OHK4/3T/14. Users are welcome to make suggestions for further improvement.



72. Sjezd chemiků v Praze 6.–9. září 2020



Sekce

- Analytická chemie
- Anorganické materiály
- Ekonomika a řízení chemického průmyslu
- Fyzikální chemie a elektrochemie
- Chemické vzdělávání a historie chemie
- Chemie životního prostředí
- Jaderná chemie
- Organické materiály
- Polymery
- Popularizace chemie
- Průmyslová chemie
- Termická analýza
- Toxikologie a lékařská chemie

<http://sjezd72.csch.cz>