

Analýza vlivu expozice na difrakční data a model struktury proteinu

Jan Stránský^{1,2}, Leona Švecová^{1,2}, Petr Kolenko^{1,2}, Jan Dohnálek²

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

²Biotechnologický ústav, Akademie věd České republiky, v. v. i.
jan.stransky@fjfi.cvut.cz

Abstrakt

Expoziční čas je důležitým parametrem pro úspěšné řešení, upřesnění a analýzu struktur bílkovin. Vysoká expozice způsobuje radiační poškození, nízká pak snižuje rozlišení difrakčních dat, rozmývá detaily v mapách elektronové hustoty a v některých případech může úplně znemožnit řešení fázového problému experimentálními metodami. V příspěvku z minulého roku bylo ukázáno, že graf závislosti poměru intenzity a chyby měření intenzity na intenzitě může být pomocným nástrojem při určování správné expozice již před sběrem kompletních difrakčních dat. V této případové studii je ukázán vliv různých expozičních časů na difrakční data a kvalitu krystalové struktury bílkoviny.

Klíčová slova: Proteinová krystalografie; Expozice; Radiační poškození; Zpracování dat.

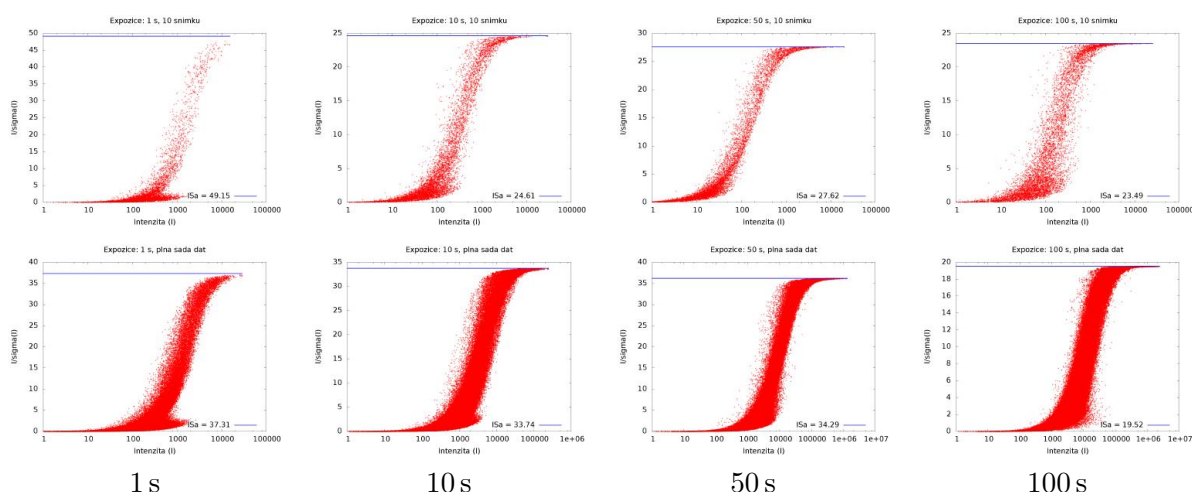
Úvod

Expozice je důležitý parametr difrakčního experimentu a může rozhodovat o úspěšnosti či neúspěšnosti řešení struktury z naměřených dat. Nové detektory záření však ztěžují odhad optimální expozice na základě vizuální inspekce snímků. Pomocným nástrojem pro odhad expozice může být graf závislosti $I/\sigma(I)$ na I [1] a byl proto vyvinutý skript pro rychlé vykreslení tohoto grafu [2]. Správnost expozice lze pak odhadovat na základě rozložení hodnot $I/\sigma(I)$ silných reflexí vzhledem k limitní hodnotě ISa [3]. Graf lze vykreslit již pro několik málo snímků, které je zapotřebí nasbírat pro určení strategie měření, a odhad expozice tak lze učinit před měřením kompletní sady difrakčních snímků. Pro porozumění vlivu různých expozičních časů na difrakční data a výsledný model struktury proteinu byla naměřena data s různou expozicí. Dále byla naměřena data pro analýzu vlivu oscilačního úhlu na graf závislosti $I/\sigma(I)$ na I . Vliv tohoto parametru na difrakční data byl analyzován dříve [4].

Metody

Krystalizace

Protein (lysozym) v prášku byl rozpuštěn v demineralizované vodě na koncentraci 100 mg/ml. Krystalizační experiment byl připraven krystalizačním robotem Gryphon (Art Robbins Instruments) do 96 jamkových desek v uspořádání sedící kapky. Pro krystalizaci byla použita sada podmínek Morpheus (Molecular Dimensions). Objem rezervoáru byl 70 μ l, kapky byly složeny ze směsi roztoku proteinu a krystalizačního roztoku v poměrech 2:1, 1:1 a 1:2 při



Obrázek 1: Grafy závislosti $I/\sigma(I)$ na I pro vybrané dílčí měření. V prvním řádku data pro prvních 10 snímků (7°), v druhém řádku pro kompletní sadu difrakčních dat dílčího měření. Modrou linií je vyznačena hodnota ISa , limitní hodnota $I/\sigma(I)$ pro velké hodnoty I [3].

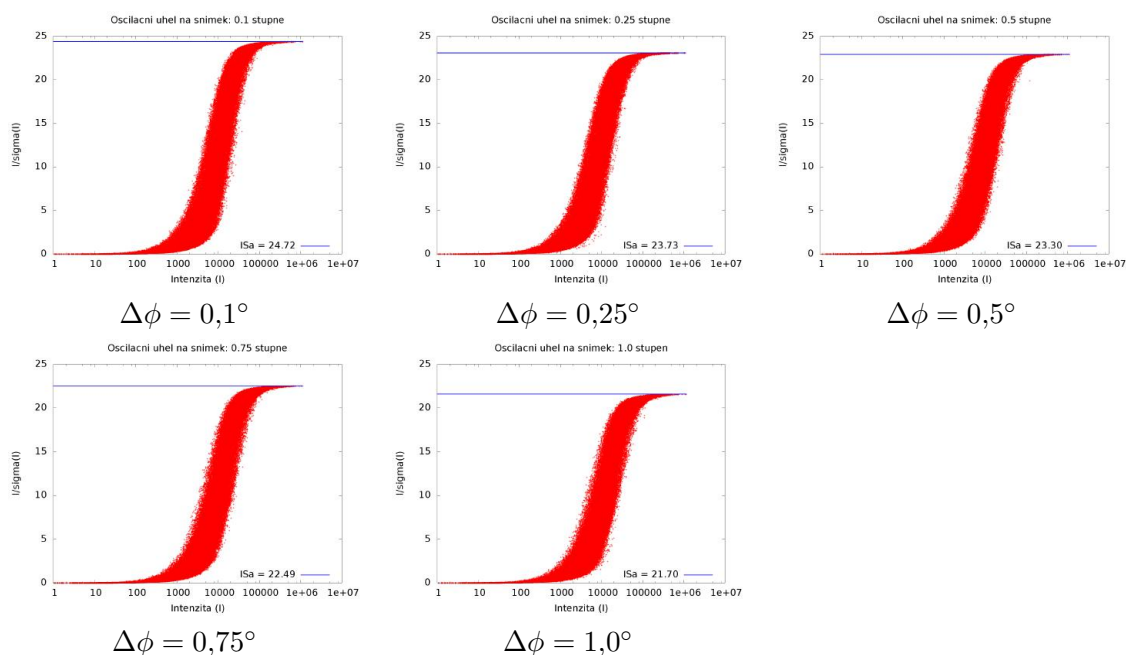
finálním objemu kapek $0,3 \mu\text{l}$. Deska byla následně uložena do krystalizačního hotelu RI-1000 (Formulatrix), kde probíhalo automatické snímkování. Krystal použitý pro difrakční experiment vyrostl v podmínce $0,2 \text{ M}$ mravenčanu sodného, $0,2 \text{ M}$ octanu amonného, $0,2 \text{ M}$ citronanu sodného, $0,2 \text{ M}$ vinnanu sodno-draselného, $0,2 \text{ M}$ šťavelanu sodného, 1 M směsi pufrů Tris a BICINE při $\text{pH } 8,5$, 40% (v/v) 1,2-ethandiol a 20% (w/v) PEG 8000.

Difrakční experiment

Difrakční data byla měřena na difraktometru D8 Venture (Bruker) se čtyřkruhovým goniometrem, zdrojem záření MetalJet (Excillum) a detektorem Photon 2 (Bruker). Data byla měřena při teplotě 100 K . Pro sběr kompletních dat byla zapotřebí 4 měření s různým nastavením goniometru a rotační osou ω , celkem 739° při oscilaci $0,7^\circ$ na snímek v bezzávěrkovém režimu. Kompletní data byla naměřena pro expoziční časy 1 s , 10 s , 50 s a 100 s . Na jiném krystalu byla naměřena data s rotační osou ϕ s oscilačními úhly na jeden snímek $0,1^\circ$, $0,25^\circ$, $0,5^\circ$, $0,75^\circ$ a $1,0^\circ$ při konstantní expozici 10 s na 1° rotace.

Zpracování difrakčních dat a řešení struktury

Difrakční data byla zpracována programem *XDS* [5] prostřednictvím nástroje *xds kappa*. Jednotlivé sady difrakčních dat byly seškálovány dohromady programem *XSCALE* [5]. Byla zvolena prostorová grupa $P4_32_12$ s mřížkovými parametry $a = b = 77,0 \text{ \AA}$, $c = 38,3 \text{ \AA}$. Fázový problém byl vyřešen metodou SAD programy *SHELXC/D/E* [6] na sadě dat, která vznikla spojením všech naměřených snímků. Ruční stavění modelu molekuly v programu *COOT* [7] a upřesňování v recipročním prostoru programem *REFMAC5* [8] bylo provedeno oproti datům naměřených při expozici 50 s . Na finálním modelu pak bylo provedeno upřesnění v programu *REFMAC5* oproti datům naměřených s expozicemi 1 s , 10 s a 100 s . Při finálním kroku upřesňování v recipročním prostoru byly použity anisotropní B-faktory.



Obrázek 2: Grafy závislosti $I/\sigma(I)$ na I pro různé oscilační úhly jednoho snímku ($\Delta\phi$) při stejné expozici na 1° měření. Grafy jsou vykresleny pro kompletní dílčí měření. Modrou linií je vyznačena hodnota ISa , limitní hodnota $I/\sigma(I)$ pro velké hodnoty I [3].

Výsledky a diskuze

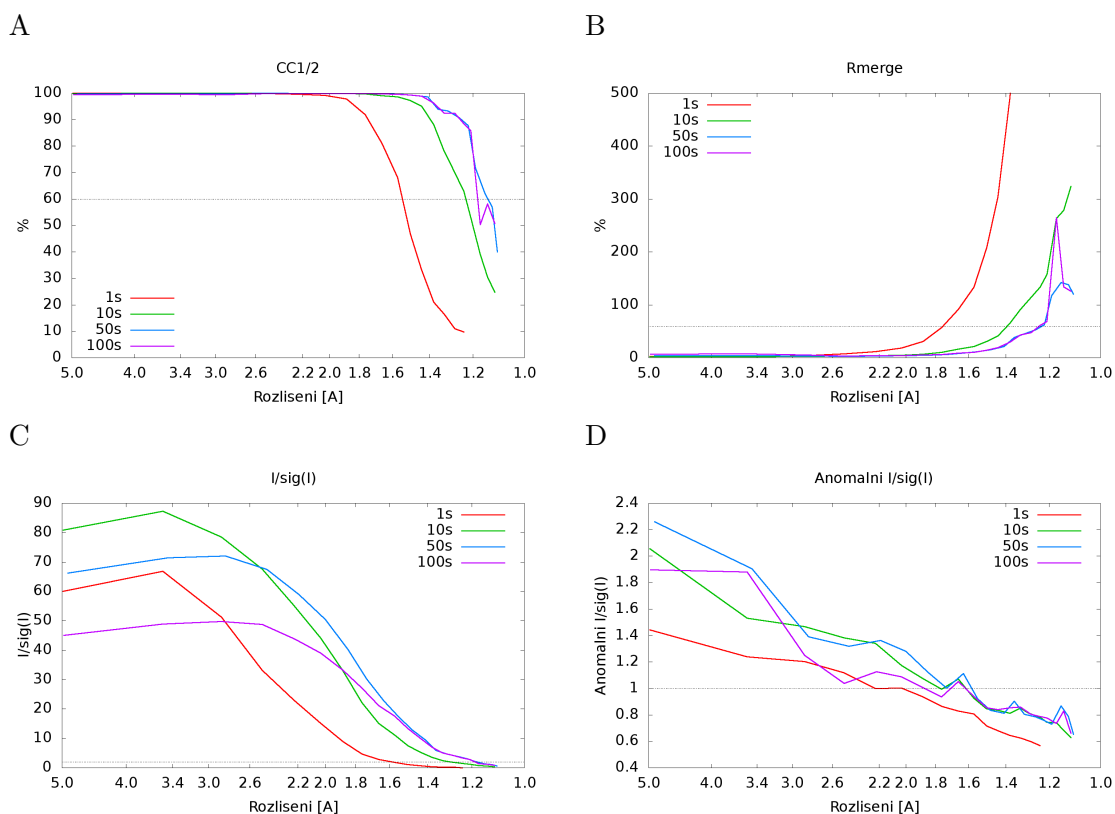
Grafy závislosti $I/\sigma(I)$ na I

Pro všechny změřené dílčí sady difrakčních dat byly vykresleny grafy závislosti $I/\sigma(I)$ na I . Pro kompletní data (spojená ze 4 dílčích měření) není možné graf vykreslit, protože chybový model a výpočet $\sigma(I)$ je prováděn při zpracování dílčích měření. Pro vybranou geometrii (tzn. měření se stejným nastavením goniometru, ale různé expozice) jsou grafy zobrazeny na Obrázku 1, společně s grafy vytvořenými pro data z prvních 10 snímků (7° rotace podle ω). V grafech pro expozici 1 s se téměř žádné reflexe nepřibližují limitní hodnotě ISa , pro expozici 10 s už se některé reflexe této limitě blíží. Pro hodnoty 50 s a 100 s je poblíž limitní hodnoty již velké množství reflexí, se zvyšující se expozicí pak přibývá reflexí, které i přes zvyšující se intenzitu I nezvyšují $I/\sigma(I)$ a tedy nepřinášejí nové informace. Při porovnání grafů pro prvních 10 snímků a pro celé dílčí měření se ukazuje, že celkový charakter grafu se zachovává.

Pro různé volby oscilačního úhlu při zachovávající se expozici na 1° se graf závislosti $I/\sigma(I)$ na I nemění (Obrázek 2).

Statistiky difrakčních dat a parametry struktury.

Kritériem pro stanovení difrakčního limitu dat použitých pro škálování byl korelační faktor mezi polovičními sadami ($CC_{1/2}$) v nejvyšší slupce dosahující hodnoty alespoň 0,6. Statistiky jsou uvedeny v Tabulce 1, závislost vybraných statistik na rozlišení je vynesena do Obrázku 3. Statistiky difrakčních dat se zlepšují se zvyšující se expozicí, mezi statistikami



Obrázek 3: Závislost statistik na rozlišení po průměrování difrakčních dat. *A* - korelace polovičních sad ($CC_{1/2}$), *B* - R_{merge} , *C* - $I/\sigma(I)$ a *D* - $I/\sigma(I)$ rozdílů Friedlových párů (anomální signál). Čárkované je vyznačen limit obvykle uznávaný jako parametr důvěryhodných dat.

pro měření při expozičních 50 s a 100 s však již nejsou žádné významné rozdíly; výjimkou jsou hodnoty $I/\sigma(I)$, které se v nízkých rozlišeních při zvyšování expozice nad 10 s snižují. Podobnou závislost lze pozorovat i pro $I/\sigma(I)$ anomálních rozdílů.

Pro smysluplnější porovnání shody modelu struktury s difrakčními daty byly po finálním upřesnění vypočteny strukturní statistiky (R -faktory, odchylky od geometrie, střední B -faktor) i pro rozlišení 1,5 Å, tedy na stejných sadách reflexí. Velký rozdíl mezi R a R_{free} pro expozici 1 s ukazuje, že na těchto datech již není vhodné používat upřesňování anisotropních B -faktorů. Rozdíly v R -faktorech a odchylkách od geometrie jsou mezi měřeními při 10 s, 50 s a 100 s minimální a mohou být způsobeny různým nastavením váhy difrakčních dat při upřesňování. Rozdíly mezi 50 s a 100 s (např. zvyšující se B -faktor) mohou být také způsobeny radiačním poškozením, jelikož měření s expozicí 100 s probíhalo jako poslední.

Závěr

Graf závislosti $I/\sigma(I)$ na I se ukazuje být vhodným pomocníkem při odhadu expoziční doby pro měření difrakčních dat, protože graf je možné vykreslit již pro několik počátečních snímků a jeho charakter je podobný jako graf pro kompletní data. Volba oscilačního úhlu při konstantní expozici na 1° nemá na podobu grafu vliv. Různé expoziční doby pak

Tabulka 1: Statistika zpracování difrakčních dat a parametry struktury

Expoziční čas Slupka rozlišení	1 s		10 s			
	Celkem	Vnější	Celkem	1,5 Å	Vnější	
Dolní limit rozlišení [Å]	10,9	1,53	10,9	1,53	1,22	
Horní limit rozlišení [Å]	1,5	1,5	1,2	1,5	1,2	
R_{merge}	0,161	2,512	0,085	0,339	1,992	
R_{meas}	0,170	2,659	0,090	0,359	2,149	
R_{pim}	0,052	0,869	0,030	0,117	0,803	
Počet pozorování	382 820	16 479	645 979	16 454	24 513	
Počet nezávislých reflexí	18 982	922	36 572	925	1 767	
Střední $(I)/\sigma(I)$	21,7	1,2	26,9	8,5	1,4	
CC(1/2)	0,999	0,608	0,999	0,982	0,657	
Kompletnost [%]	99,7	100	99,8	100	100	
Multiplicita	20,2	17,9	17,7	17,8	13,9	
R_{work}	0,1157		0,1247	0,1137		
R_{free}	0,1938		0,1716	0,1583		
s. s. o. ¹ vazeb [Å]	0,0234		0,0223	0,0223		
s. s. o. ¹ úhlů [°]	1,9769		2,0080	2,0079		
Střední B-faktor [Å ²]	16,21		13,69	13,80		
Expoziční čas Slupka rozlišení	50 s			100 s		
	Celkem	1,5 Å	Vnější	Celkem	1,5 Å	Vnější
Dolní limit rozlišení [Å]	10,9	1,53	1,16	10,9	1,53	1,16
Horní limit rozlišení [Å]	1,14	1,5	1,14	1,14	1,5	1,14
R_{merge}	0,069	0,167	1,906	0,080	0,157	1,842
R_{meas}	0,074	0,177	2,091	0,085	0,166	2,023
R_{pim}	0,025	0,059	0,853	0,029	0,055	0,831
Počet pozorování	689 819	16 070	23 376	676 460	15 835	22 767
Počet nezávislých reflexí	42 491	925	2 088	42 538	924	2 082
Střední $(I)/\sigma(I)$	27,5	17	1,4	23,2	18,1	1,5
CC(1/2)	0,998	0,995	0,683	0,998	0,996	0,684
Kompletnost (%)	99,7	100	99,5	99,7	100	99,3
Multiplicita	16,2	17,4	11,2	15,9	17,1	10,9
R_{work}	0,1327	0,1242		0,1355	0,1274	
R_{free}	0,1698	0,1616		0,1723	0,1649	
s. s. o. ¹ vazeb [Å]	0,0207	0,0207		0,0212	0,0212	
s. s. o. ¹ úhlů [°]	1,9151	1,9156		1,9308	1,9310	
Střední B-faktor [Å ²]	13,58	13,70		14,21	14,32	

¹ Střední směrodatná odchylka od ideálních hodnot

mají významný dopad na měřená difrakční data. Testovací případ ukazuje, že se zvyšující se dobou expozice se zvyšuje rozsah použitelného rozlišení a zlepšují se indikátory kvality zpracování dat. Při zvyšování expozice nad určitou hodnotu (zde z 50 s na 100 s na snímek) už k žádnému zlepšení nedochází. Při zvyšování expozice naopak docházelo ke snižování odhadu $I/\sigma(I)$ pro nízké rozlišení, což může být způsobeno i radiačním poškozením, jelikož experiment probíhal na jednom krystalu. Zvyšování expozice nad 10 s také nemělo významný vliv na indikátory shody modelu struktury s měřením. Měření při expozici 1 s je podexponované a z krystalu je možné získat více informací při správném nastavení expozice. Naproti tomu měření při 100 s oproti nižším expozicím již žádné nové informace nepřináší. Rozsah expozic pro optimální měření je široký a konkrétní volba by tedy měla být přizpůsobena i dalším požadavkům s ohledem na radiační poškození a cíle měření.

Reference

- [1] T. Weinert, aj., Fast native-SAD phasing for routine macromolecular structure determination. *Nature Methods* **12**(2): 131–U163, 2015.
- [2] J. Stránský, J. Dohnálek, Zanedbávaný parametr difrakčního experimentu: expoziční čas. V *Sborník příspěvků 5. studentské vědecké konference fyziky pevných látek*, roč. 1, 22–27, ČVUT v Praze, 2015.
- [3] K. Diederichs, Quantifying instrument errors in macromolecular X-ray data sets. *Acta Crystallographica Section D - Biological Crystallography* **66**(6): 733–740, 2010.
- [4] M. Mueller, M. Wang, C. Schulze-Briese, Optimal fine φ -slicing for single-photon-counting pixel detectors. *Acta Crystallographica Section D - Biological Crystallography* **68**(1): 42–56, 2012.
- [5] W. Kabsch, XDS. *Acta Crystallographica Section D - Biological Crystallography* **66**(2): 125–132, 2010.
- [6] G. M. Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D - Biological Crystallography* **66**(4): 479–485, 2010.
- [7] P. Emsley, aj., Features and development of coot. *Acta Crystallographica Section D - Biological Crystallography* **66**: 486–501, 2010.
- [8] G. Murshudov, A. Vagin, E. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D - Biological Crystallography* **53**(3): 240–255, 1997.

Poděkování

Tato práce byla podpořena Ministerstvem školství, mládeže a tělovýchovy granty LQ1604 NPU II, LG14009 a CIISB - LM2015043, Evropským fondem regionálního rozvoje č. CZ.1.05/1.1.00/02.0109, institucionální podporou Biotechnologického ústavu AV ČR, v. v. i. RVO: 86652036 a Grantovou agenturou ČVUT č. SGS16/246/OHK4/3T/14. Dále bychom chtěli poděkovat Jiřímu Pavlíčkovi z Centra molekulární struktury, BIOCEV za podporu při přípravě vzorků a měření.