

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF BIOMEDICAL ENGINEERING

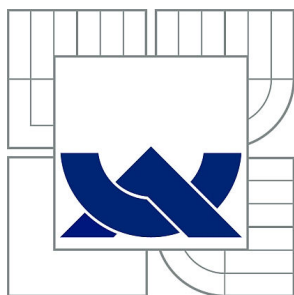
3D REKONSTRUKCE MAKROMOLEKULÁRNÍCH KOMPLEXŮ POMOCÍ
KRYOELEKTRONOVÉ MIKROSKOPIE

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

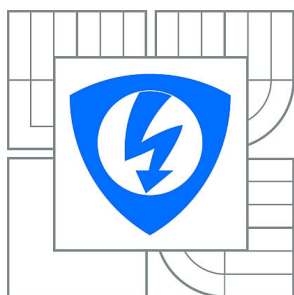
AUTOR PRÁCE
AUTHOR

Mgr. RADIM SKOUPÝ

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF BIOMEDICAL ENGINEERING

3D REKONSTRUKCE MAKROMOLEKULÁRNÍCH KOMPLEXŮ POMOCÍ KRYOELEKTRONOVÉ MIKROSKOPIE

3D RECONSTRUCTION OF MACROMOLECULAR COMPLEXES BY CRYOELECTRON
MICROSCOPY

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

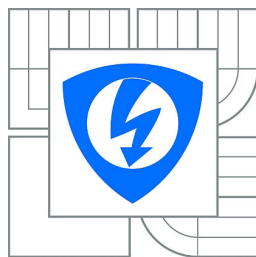
AUTOR PRÁCE
AUTHOR

Mgr. RADIM SKOUPÝ

VEDOUCÍ PRÁCE
SUPERVISOR

RNDr. Daniel Němeček, Ph.D.

BRNO 2015



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav biomedicínského inženýrství

Diplomová práce

magisterský navazující studijní obor
Biomedicínské inženýrství a bioinformatika

Student: Mgr. Radim Skoupý

ID: 136487

Ročník: 2

Akademický rok: 2014/2015

NÁZEV TÉMATU:

3D rekonstrukce makromolekulárních komplexů pomocí kryoelektronové mikroskopie

POKYNY PRO VYPRACOVÁNÍ:

1) Proveďte literární rešerši metod pro zpracování obrazu a rekonstrukce 3D struktury biologických makromolekul zobrazených pomocí transmisní elektronové mikroskopie. 2) Seznamte se s funkcí transmisních elektronových mikroskopů a různých metod měření cryoEM experimentálních dat. 3) Navrhněte optimální metodiku pro měření a zpracování cryoEM datasetů heterogenních komplexů s nízkou symetrií jako je EcoR124I. 4) Vypočtěte 3D rekonstrukce studovaných biomolekulárních komplexů v různých konformačních a funkčních stavech. 5) Proveďte diskusi získaných výsledků a navrhněte mechanismus funkce studovaných komplexů.

DOPORUČENÁ LITERATURA:

[1] FRANK, J. (2006) "Three-Dimensional Electron Microscopy of Macromolecular Assemblies", Edition 2, Oxford University Press, Inc., New York, ISBN: 978-0195182187.

[2] JENSEN G.J. (2010) Methods in Enzymology, Volume 481-483, First Edition, Academic Press, Elsevier, Inc, San Diego, ISBN: 978-0-12-374906-2.

Termín zadání: 9.2.2015

Termín odevzdání: 22.5.2015

Vedoucí práce: RNDr. Daniel Němeček, Ph.D.

Konzultanti diplomové práce:

prof. Ing. Ivo Provazník, Ph.D.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Diplomová práce se zabývá zpracováním dat z transmisní elektronové mikroskopie a jejich analýzou (tomografie, jednočásticová analýza). Hlavním úkolem práce je stanovit 3D strukturu studovaného enzymu. Jako testovací vzorek s nízkou symetrií používá restriční endonukleázu EcoR124I.

KLÍČOVÁ SLOVA

kryoEM, restriční endonukleáza, EcoR124I, tomografie, jednočásticová analýza

ABSTRACT

Master thesis deals with the processing of data from transmission electron microscopy and their analysis (tomography, single particle analysis). The main aim of this work is to determine the 3D structure of the studied enzyme. As a test sample with low symmetry is used restriction endonuclease EcoR124I.

KEYWORDS

cryoEM, restriction endonuclease, EcoR124I, tomography, single particle analysis

SKOUPÝ, Radim *3D rekonstrukce makromolekulárních komplexů pomocí kryoelektronové mikroskopie*: diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ, 2015. 62 s. Vedoucí práce byl RNDr. Daniel Němeček, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „3D rekonstrukce makromolekulárních komplexů pomocí kryoelektronové mikroskopie“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

(podpis autora)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu RNDr. Danielu Němečkovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

(podpis autora)

OBSAH

1 Úvod	12
Úvod	12
2 Kryoelektronová mikroskopie	13
2.1 Formování obrazu a kontrast v kryoEM	14
2.1.1 Rozptylový kontrast	15
2.1.2 Fázový kontrast	15
2.2 Přístrojové vybavení	16
3 Restrikční nukleázy	17
3.1 EcoR124I	18
3.2 Příprava vzorku	20
4 Tomografie	23
4.1 Princip	23
4.2 Sběr dat	23
4.3 3D rekonstrukce	25
4.3.1 Nelineární anisotropická difuze	25
4.3.2 Sběr částic	26
4.3.3 Průměrování subtomogramů	26
4.4 Výsledek zpracování 1. datasetu	27
4.5 Výsledek zpracování 2. datasetu	28
5 Jednočásticová analýza	31
5.1 DDD a CCD kamery	35
5.1.1 Zarovnání subsnímků	37
5.2 Filtrace, úprava histogramu	37
5.2.1 Gaussovský filtr	37
5.2.2 Pásmová propust	38
5.2.3 Prahování koeficientů vlnkové transformace	38
5.3 Korekce přenosové funkce kontrastu	40
5.3.1 Manuální korekce	41
5.3.2 Automatická korekce	41
5.3.3 Amplitudová korekce a Wienerova filtrace	41
5.4 Sběr částic	43
5.4.1 Algoritmy využívající templát	43
5.4.2 Algoritmy hledající charakteristické vlastnosti	44

5.4.3	Aplikovaný postup	44
5.4.4	Extrakce částic	45
5.5	Zatřídění nalezených částic – závisle na referenci	47
5.6	Zatřídění nalezených částic - nezávisle na referenci	48
6	Závěr	51
	Literatura	52
	Seznam symbolů, veličin a zkratk	55
	Seznam příloh	56
A	Pracovní postupy	57
A.1	Tomografie	57
A.2	Jednočásticová analýza	58
B	Software	59
B.1	Bsoft	59
B.2	RELION	60
C	Konferenční abstrakt	61
C.1	CryoEM reconstruction of small nucleoprotein complexes: Application to the EcoR124I restriction-modification enzyme	61

SEZNAM OBRÁZKŮ

2.1	Schématické zobrazení sítky pro TEM s připraveným vzorkem typu tenký film pro kryoEM.[1]	14
2.2	Hlavní přístrojové vybavení využitě během přípravy a snímání vzorků. Titan Krios (vlevo). Vitrobot (vpravo).	16
3.1	Příklady restričních endonukleáz spolu s jejich rozpoznávacími sekvencemi.	17
3.2	(A) EcoR124I+DNA (zavřený stav) EM v negativním kontrastování. (B) EcoR124I bez DNA (otevřený stav) EM v negativním kontrastování. [8]	19
3.3	Stanovení stechiometrie komplexu pomocí afinitní chromatografie.	20
3.4	Výsledek vylučovací chromatografie skladovaného purifikovaného komplexu R_1M_2S naznačuje jeho stabilitu během uskladnění.	21
3.5	Komplex EcoR124I zobrazený pomocí negativního kontrastování (2% uranylacetát) (nahore). Stejný komplex zobrazený v kryomikroskopu Tecnai TF20 při urychlovacím napětí 120kV s koncentrací vzorku 0.3 mg/ML (uprostřed). Tentýž komplex zobrazený pomocí kryomikroskopu Titan Krios při urychlovacím napětí 300kV s koncentrací vzorku 10 mg/mL. (dole).	22
4.1	Princip kryoelektronové tomografie. (A) Vzorek je nakláněn v tzn. tiltovací sérii a postupně zaznamenáván. (B) Jednotlivé projekce se pro představu umísťují na virtuální plášť válce se středem ve vzorku. (C) 3D struktura se dopočítává zvolenou rekonstrukční metodou. [14]	24
4.2	Výběr částic pro výpočet zprůměrovaného tomogramu enzymu EcoR124I. V pravém horním rohu je pohled na jednu z částic ve všech třech rovinách.	27
4.3	Výsledek první objemové rekonstrukce enzymu EcoR124I. Na jejích řezech jsou patrné dvě podjednotky místo tří. Z tohoto důvodu není model dostatečný a musí být provedena další rekonstrukce.	28
4.4	Druhý vypočítaný tomogram, ze kterého byly extrahovány částice pro další zpracování.	29
4.5	Jednotlivé řezy druhé objemové rekonstrukce enzymu EcoR124I, na jejichž řezech jsou patrné všechny tři podjednotky.	30
5.1	Určení struktury neobaleného ikosahedrálního viru pomocí kryoEM. (A) Vizualizace celé struktury. (B) Vybraný region demonstrující téměř atomové rozlišení, které je možné získat u vzorků s velkou symetrií. [2]	31

5.2	Automaticky zpracovaná struktura pomocí SPA. (A) Obrázky sítky pro kryoEM postupně v narůstajícím zvětšení zleva doprava až po zvětšení, při kterém jsou rozpoznatelné jednotlivé molekuly. (B) Vzorový projekční obraz zmraženého zavodněného vzorku purifikovaného proteinového komplexu GroEL. Ve čtverečkách jsou zvýrazněny komplexy s různou orientací vůči ose svazku. (C) 3D rekonstrukce sestavená z přibližně 28000 jednotlivých projekcí s výsledným rozlišením 7 Å. Iniciální model byl spočítán ze 2000 GroEL částic. Zpřesňování iniciální rekonstrukce byla provedena téměř automatickou procedurou implementovanou v softwarovém balíku FREALIGN. (D) Demonstrace, že dosažené rozlišení je dostatečné k zobrazení α -helixů ilustrovaná superpozicí mapy hustoty a odpovídajícího polypeptidového řetězce určeného pomocí rentgenové krystalografie (PDB ID: 3E76). [2]	32
5.3	Surový mikrogram enzymu EcoR124I „otevřený“ (FEI Titan Krios, urychlovací napětí 300 kV, defokus $-4\mu\text{m}$) nahoře. Stejný mikrogram po filtraci pásmovou propustí s mezními frekvencemi 20 a 300 Å (uprostřed) a pomocí dyadické vlnkové transformace s metodou stanovení prahu NeighShrinkSure (dole).	34
5.4	Základní rozdíl ve skladbě CCD detektoru elektronů (vlevo) a DDD využívající CMOS architekturu (vpravo).[18]	36
5.5	Srovnání metod VisuShrink, SureShrink a NeighShrinkSure pro určení prahu během filtrace užitím DWT. [20]	40
5.6	Průběh prvních pěti CTF zero vzhledem k aplikovanému defokusu při 300 kV a koeficientu sférické aberace $C_S = 3\text{mm}$. Pro defokus $-4\mu\text{m}$ je první CTF zero (černá křivka) 28 Å (vlevo). Výsledek CTF korekce na jednom z mikrogramů. V levé polovině je vypočítaný model. V pravé části pak fourierova transformace mikrogramu (vpravo).	43
5.7	Jeden z filtrovaných mikrogramů s vyznačenými nalezenými částicemi (zelené koužky).	46
5.8	Nalezené a extrahované částice seřazené vzhledem k podobnosti s referencí.	47
5.9	Čištění datasetu pomocí odstraňování malých a špatně zarovnaných skupin. Příklad nekvalitních částic tvořících si v malém počtu svoji vlastní skupinu (nahoře) a kvalitní skupiny vyznačující se dobrým zarovnáním – nízkým pozadím (vpravo dole).	49
5.10	Výsledek sedmého běhu iterací již naznačuje vnitřní strukturu částic (červeně ohraničené třídy). Upraveno transformací LUT.	50

A.1	Blokové schéma při zpracování tomografických dat v programovém balíku Bsoft včetně používaných příkazů.[15]	57
A.2	Blokové schéma jendočásticová analýzy v programovém balíku Bsoft včetně využívaných příkazů.[15]	58

SEZNAM TABULEK

5.1	Časová náročnost výpočtu fourierovy transformace v závislosti na zvolené velikosti boxu. Ve 3D případě je čas nejhoršího 20 krát delší než nejlepšího.	46
5.2	Počet zůstávajících částic po jednotlivých iteracích třídění do tříd. . .	48

1 ÚVOD

Tato diplomová práce se věnuje 3D rekonstrukci makromolekulárních komplexů pomocí kryoelektronové mikroskopie (cryoEM). Trojrozměrná mapa EM hustoty studovaného komplexu proteinů a nukleových kyselin pak může být využita pro určení molekulární architektury a struktury komplexu a identifikace strukturního mechanismu jeho funkce. V případě, že jsou známy atomové modely částí komplexu z krystalografie či nukleární magnetické rezonance lze rekonstruovanou EM mapu využít jako obálku pro fitování těchto fragmentů a určení jejich interakcí v kompletním komplexu.

Základní zobrazení v transmisním elektronovém mikroskopu (TEM) je 2D projekce zobrazených 3D molekul. Jednotlivé 2D projekce jsou následně extrahovány z naměřených mikrografů a v iterativním procesu jim jsou přiřazeny orientace v 3D rozměru. Kombinací mnoha jednotlivých projekcí se známými orientacemi je pak možné rekonstruovat 3D mapu zobrazených molekul pomocí inverzní Radonovy transformace. Typicky přiřazení orientací a 3D rekonstrukce jsou opakovány v iterativním procesu, dokud přiřazení orientací a kvalita 3D rekonstrukce konvergují ke stálým hodnotám.

Záměrem této práce je optimalizovat metodiku pro zpracování cryoEM datasetů relativně malých heterogenních komplexů s nízkou symetrií. Jako vzorový komplex, na němž je postup zpracování testován byl použit enzym EcoR124I, který je svou velikostí (450 kDa) na hranici efektivního zpracování s použitím standardních metod.

Použili jsme datasety z EM negativního kontrastu i z cryoEM a porovnali kvalitu dat naměřených na standardním mikroskopu FEI Tecnai F20 a FEI Titan Krios. Dále byly testovány různé filtry pro zlepšení poměru signál/šum a automatické sbírání projekcí neglobulární struktury EcoR124I.

2 KRYOELEKTRONOVÁ MIKROSKOPIE

Kryoelektronová mikroskopie je moderní zobrazovací metoda umožňující studium buněčné architektury, virů a proteinových komplexů téměř v atomovém rozlišení. Při urychlovacím napětí 300 kV je vlnová délka elektronů asi 0,02 Å. Odpovídajícího limitu rozlišení však není u biologických makromolekul možné dosáhnout z důvodu ztráty jejich strukturální integrity (zlaté částice je například možné zobrazovat na atomovém rozlišení přímo). Při dopadu svazku na vzorek dochází k rozpadu chemických vazeb a tvorbě volných radikálů, které vedou k dalším poškozením. Z tohoto důvodu se využívají velmi nízké dávky na úrovni jednotek elektronů na čtvereční Å.[2]

Vzhledem k používání nízké dávky a nízkému kontrastu způsobenému složením proteinových molekul, které jsou převážně tvořeny lehkými prvky, je výsledný poměr signál/šum (SNR) také nízký. To lze překonat zprůměrováním jinak totožných obrazů, ve kterých se náhodné složky šumu navzájem vykompenzují. Pro dosažení rozlišení 3 - 4 Å je nutné zprůměrování 10^6 částic. Pro rozlišení 20 Å je jich potřeba několik tisíc. Například při 10 Å rozlišení mohou již být viditelné dlouhé α -helixy, pod 5 Å pak jednotlivé pásy β -skládaných listů. [1]

Druhou možností je využití metody negativního kontrastování, při které se dá dosáhnout rozlišení 20 až 40 Å. Metoda spočívá v ukotvení částic v roztoku soli těžkého kovu (uranyl acetát, wolframové nebo molybdenové soli). Sůl vyplní prostor okolo částice, ale neproniká dovnitř, a tak vzniká velmi dobrý kontrast proteinových obálek. [2]

Velkým problémem je však příprava vzorku před samotným zobrazením probíhajícíím ve vysokém vakuu. Při využití kryoEM je možné pozorovat zavodněné biologické vzorky při jejich nativním obsahu vody a to díky mrazové fixaci.

Voda se stabilizuje pomocí vitrifikace kapalné fáze, která spočívá v rychlém a hlubokém podchlazení preparátu tak, aby voda nestačila zkrystalizovat a přeměnila se v amorfni led. Takto připravený vzorek je pak možné vysušit pomocí mrazové sublimace nebo pozorovat přímo pomocí kryoEM. [3]

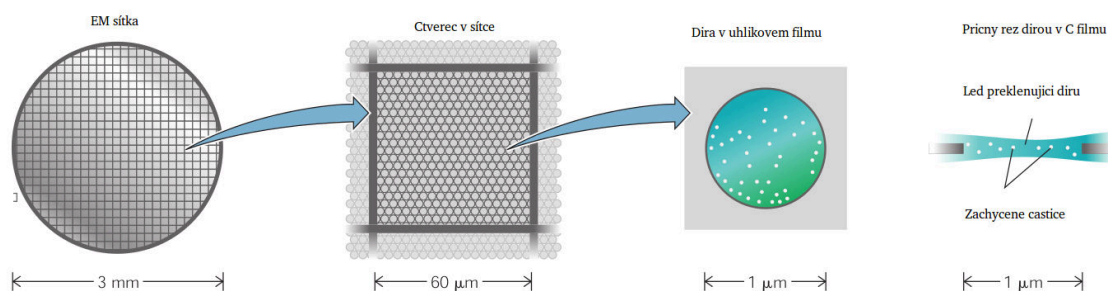
Výhody kryo fixace oproti chemické jsou:

- možnost pozorovat buňky, organely, molekuly v jejich přirozeném prostředí
- nedochází ke změnám (denaturaci) enzymů a antigenů
- vzorek je imobilizován ve zlomcích sekundy - což umožňuje sledovat i dynamické děje [3]

K tvorbě amorfniho ledu z čisté vody je třeba ji ochladit pod -150°C rychlostí $10^6\text{ }^{\circ}\text{C s}^{-1}$. V případě biologických vzorků obsahujících směs organických a anorganických vodných roztoků oddělených membránami postačuje ochlazení pod -90°C rychlostí 10^4 až $10^5\text{ }^{\circ}\text{C s}^{-1}$. Jako chladicí médium se používá kapalný etan chlazený

dusíkem.[3]

Využitelné jsou dva typy vzorků – tenké filmy nebo řezy. Tenký film (typ z něhož budou v další části práce zpracovávána data) se připravuje nanesením suspenze na sítku pokrytou uhlíkovou blankou s otvory. Po odsátí přebytečné tekutiny se provádí mrazová fixace.



Obr. 2.1: Schématické zobrazení sítky pro TEM s připraveným vzorkem typu tenký film pro kryoEM.[1]

Podle velikosti zkoumaných vzorků je možné rozdělit paletu metod spadajících pod kryoEM na dvě kategorie:

Elektronová mikroskopie buněk a organel

- Ultrastrukturální tomografie
- Detekce specifických molekul – imunoznačení
- Stínování těžkými kovy
- Repliky
 - Mrazové lámání
 - Mrazové leptání

Elektronová mikroskopie molekul

- Negativní kontrastování
- Jednočásticová analýza, kapitola 5
- Elektronová tomografie molekul, kapitola 4

2.1 Formování obrazu a kontrast v kryoEM

Důsledkem interakce dopadajícího primárního elektronového svazku se vzorkem je nerovnoměrná distribuce intenzity elektronů na jeho spodním povrchu, což však bohužel nepostačuje k dosažení dostatečného kontrastu, a tak je nutné provést vnější zásah (změna defokusu, změna tloušťky vzorku, konvergence dopadajícího svazku, umístění a centrování clon), který ho zvětší. V TEM můžeme využít tři základní druhy kontrastu:

1. amplitudový kontrast (rozptylový a difrakční) - používá se malá objektivová clona vymezující přímý a difraktovaný svazek
2. fázový kontrast (interferenční kontrast) - používá se velká objektivová clona
3. Z-kontrast - možné jen v mikroskopu umožňujícím skenování (STEM)

V kryoEM se využívá kombinace amplitudového - rozptylového kontrastu a fázového kontrastu v nestejném poměru asi 0,1 (rozptylový) ku 0,9 (fázový).

2.1.1 Rozptylový kontrast

Při elastickém rozptylu elektrony interagují s elektrostatickým potenciálem jader atomů díky němuž mění svůj směr - kontrast je tedy závislý na tloušťce a hustotě vzorku. Rozptýlené elektrony jsou pak zachyceny clonou a podílejí se tak na vzniku kontrastu ve světlém poli. Stejně tak je zachycena i většina neelasticky rozptýlených elektronů, zbytek se však podílí na vytváření pozadí obrazu. Části vzorku s vyšší hustotou (větší tloušťkou) jsou ve světlém poli tmavé (na detektor nedopadá část svazku zachyceného clonou).

2.1.2 Fázový kontrast

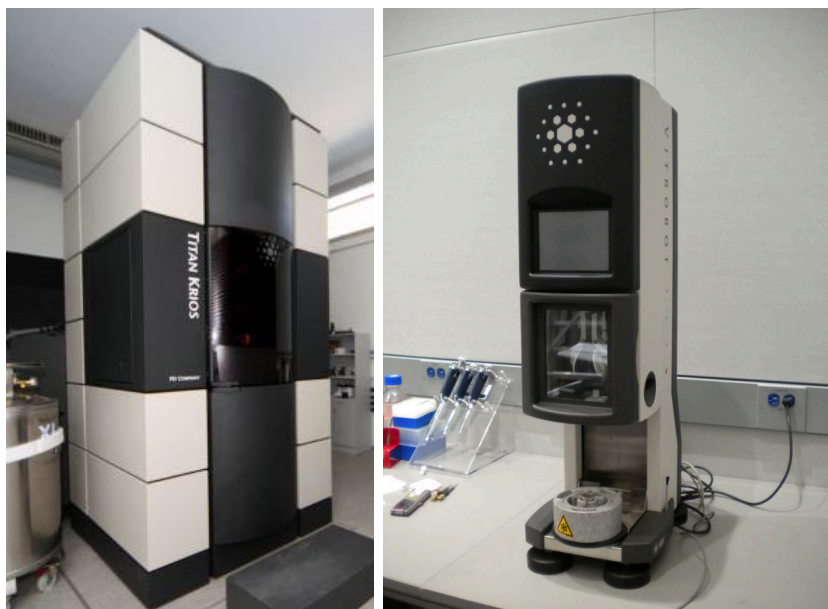
V tomto případě se využívá velká objektivová clona, kterou prochází více svazků, jejichž interferencí vzniká obraz. Difraktované svazky mají oproti primárnímu svazku fázový posuv $\pi/2$, a tak při přesném zaostření objektivu nemohlo k interferenci vůbec dojít. To se kompenzuje částečným rozostřením objektivu o přesně definovanou hodnotu nad vzorek, čímž získáme přídatný fázový posuv $\pi/2$. Toto rozostření se nazývá "defokus" a má rozměr jednotek μm . Se zvyšujícím se defokusem se zvětšuje kontrast na úkor rozlišení, které je rozostřením samozřejmě poznamenáno. Během procesu zobrazení je důležité najít vhodný poměr mezi těmito protichůdnými požadavky.[3]

Přenosová funkce kontrastu (CTF) popisuje, jak je pozorovaný objekt zobrazován a měněn prostřednictvím zobrazovacích nedokonalostí mikroskopu. CTF je bohužel velmi vzdálena od ideálního průběhu o přenosu jedna na všech prostorových frekvencích. Všechny prostorové frekvence totiž nejsou přenášeny ani zdaleka stejně. Hlavními kritérii jsou koeficient sférické aberace C_s (dán mikroskopem) a zvolený defokus (dán uživatelem). CTF jednoho mikrografu obsahuje jak kladné, tak záporné příspěvky k signálu, a tak je základní úpravou jejich překlopení do kladné poloviny. Doplnění informací ve frekvenčních oblastech s nízkým přenosem se provádí nasnímáním datasetu při různých hodnotách defokusu, a tím i při rozdílných CTF. Chybějící části se pak vzájemně doplní při průměrování projekcí z různých mikrografů..

2.2 Přístrojové vybavení

Titan Krios 2.2 (vlevo) je transmisní elektronový mikroskop nové generace navrhnutý k zobrazování velkých makromolekulárních proteinových komplexů a nukleových kyselin. Tyto mikroskopy s paralelním ozařováním vzorků, stabilními elektromagnetickými čočkami a novou generací počítačem řízených stolků představují nástroj k získání dat na téměř atomárním rozlišení. Důležitou inovací jsou inovované poloautomatické nástroje pro 2D elektronovou krystalografii, jednočásticovou analýzu a dvojosou tomografii. Mikroskop je též teplotně a mechanicky stabilizován, což spolu s možností automatického vkládání až 12 sítěk umožňuje získat velké data-sady. Energie primárního svazku je nastavitelná v rozmezí 80 - 300 kV v závislosti na druhu zkoumaného vzorku. Funkci klasického fluorescenčního stínítka zastává vysokorychlostní digitální kamera. Hlavním zdrojem dat je pak přímý detektor elektronů (DDD) s čtyřikrát lepší senzitivitou a schopností zaznamenávat několik obrazů za sekundu. [4]

Při přípravě a zmrazování vzorků byl použit Vitrobot, taktéž od firmy FEI 2.2 (vpravo). Vitrobot umožňuje opakované, přesně definované a vysoce kvalitní zmrazení vzorku a tím tvorbu amorfního ledu důležitého k zachování původní struktury. Pracovní komora poskytuje přesnou kontrolu prostředí - teploty a vlhkosti, což předchází vzniku artefaktů tvořících se při mrazení v otevřené atmosféře. [5, 6]



Obr. 2.2: Hlavní přístrojové vybavení využívané během přípravy a snímání vzorků. Titan Krios (vlevo). Vitrobot (vpravo).

3 RESTRIKČNÍ NUKLEÁZY

Nukleázy jsou původem bakteriální enzymy štěpící DNA (fosfodiesterickou vazbu v nukleových kyselinách) ve specifických místech.

K restriční analýze DNA se používají bakteriální restriční endonukleázy (RE), které spolu s metylázami vytvářejí v bakteriích ochranný systém před cizorodou DNA vnikající do buňky přes horizontální genový transfer (HGT). To je velmi důležité při rozšiřování rezistence na antibiotika. Metylován je při tom adenin nebo cytosin. Při analýze DNA se naštípané úseky dále oddělují například gelovou elektroforézou. Rozpoznávací sekvence se u většiny enzymů pohybuje v rozpětí 4 - 6 bp. V DNA se čtyřnukleotidové rozpoznávací místo vyskytuje přibližně každých 4^4 tj. 256 nukleotidů, šestinukleotidové místo každých 4^6 tj. 4096 a osminukleotidové každých 4^8 tj. 65536 nukleotidů. Frekvence štěpení je ovlivněna i relativním obsahem G + C v DNA, obsahem repetitivních sekvencí a degenerací rozpoznávacích sekvencí. Příklady specifických míst jsou na obr. 3.1. Většina RE pracuje při 37° v mírně alkalickém prostředí (pH 7,2 - 7,6). Všechny RE vyžadují přítomnost iontů Mg^{2+} . [7]

Název	Bakteriální zdroj	Rozpoznávací sekvence a štěpení
<u>EcoR I</u>	<u>Escherichia coli RY 13</u>	5'---G <u>AATTC</u> ---3' 3'---CTTAA G---5'
<u>Hind III</u>	<u>Haemophilus influenzae Rd</u>	5'---A <u>AGCTT</u> ---3' 3'---TTCGA <u>A</u> ---5'
<u>Pst I</u>	<u>Providencia stuartii 164</u>	5'---CTGCA G---3' 3'---G <u>ACGTC</u> ---5'
<u>Taq I</u>	<u>Thermus aquaticus YT1</u>	5'---T <u>CGA</u> ---3' 3'---AGC <u>T</u> ---5'
<u>Alu I</u>	<u>Arthrobacter luteus</u>	5'---AG <u>CT</u> ---3' 3'---TC <u>GA</u> ---5'

Obr. 3.1: Příklady restričních endonukleáz spolu s jejich rozpoznávacími sekvencemi.

3.1 EcoR124I

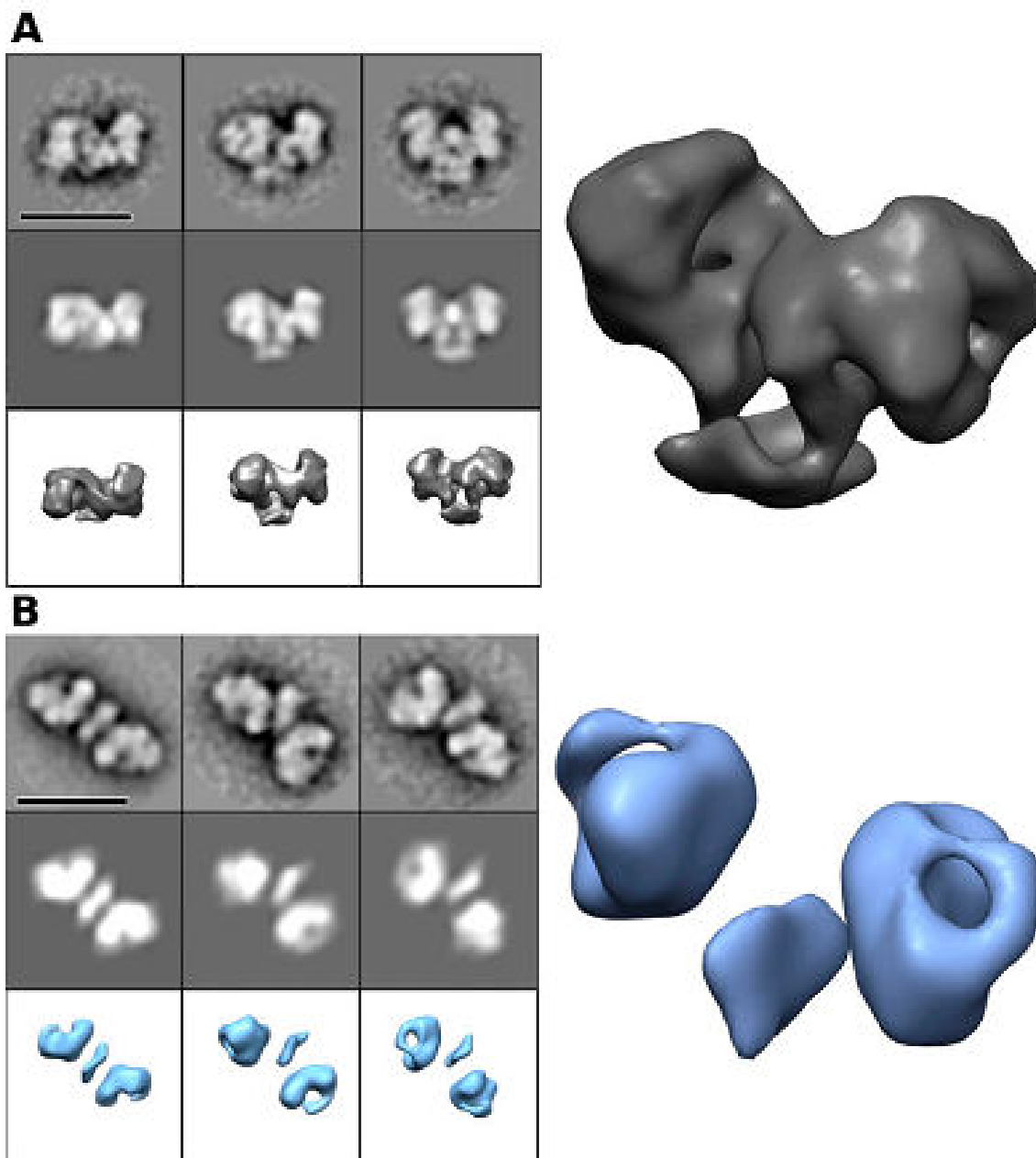
Poprvé byly DNA restrikční/modifikační (RM) enzymy typu I purifikovány již před více než 40 lety. Typ I RM enzymů byl nalezen ve více než polovině bakteriálních genomů. Jejich velikost je asi 440 kDa.[8]

Pro hlubší znalost HGT je tedy nutná znalost struktury a funkce RM systémů. RM I jsou molekulární stroje vyhledávající specifické, typicky asymetrické krátké sekvence DNA. Enzym tvoří tři spolupracující podjednotky – dvěma motorovými podjednotkami, které posouvají dvoušroubovici DNA skrz stacionární jednotku - samotný enzym. Tyto helikázové motory spotřebovávají jednu molekulu ATP na posun o jeden pár bazí. [9]

RM I enzymy jsou komplexní struktury složené ze dvou HsdR restrikčních podjednotek (R), dvou HsdM modifikačních (M) podjednotek a jedné HsdS k sekvenci specifické (S) podjednotky. HsdS se skládá ze dvou DNA rozpoznávacích domén (TRDs - target recognition domains), jedna pro každou část dvojité rozpoznávací sekvence, napojených na dva α helixy v antiparalelním uspořádání. HsdS se vyznačuje dvojitou pseudosymetrií stejně tak jako celý enzym. Dva HsdM leží na každé straně HsdS a tvoří tak M_2S_1 metyltransferázové (MTázové) jádro. HsdM obsahuje metyltransferázovou katalytickou doménu. Dvě HsdR se nacházejí na obou stranách MTázy a obsahují nukleázovou doménu a ATP-hydrolyzující DNA translokační motorovou doménu. MTáza specificky metyluje adeninové báze v cílové sekvenci, tedy nově replikované hostitelské DNA. Zatímco HsdR se zapojuje jen v případě detekce nemodifikované cílové sekvence na příchodí DNA. Rychlost pohybu enzymu po DNA helixu je 1000 bp za sekundu.[8, 10]

EcoR124I má cílovou sekvenci GAAnnnnnnRTCG, kde n je jakákoliv báze a R je purin. [11]

Enzym EcoR124I se vyskytuje ve dvou formách – s navázanou DNA a bez ní. Strukturální změny jsou vidět na obr. 3.2, při kterém bylo dosaženo rozlišení 2,1 nm. DNA není na těchto snímcích vidět. Srovnání „otevřeného“ a „zavřeného“ stavu ukazuje velký pohyb podjednotek v rámci komplexu. [8]

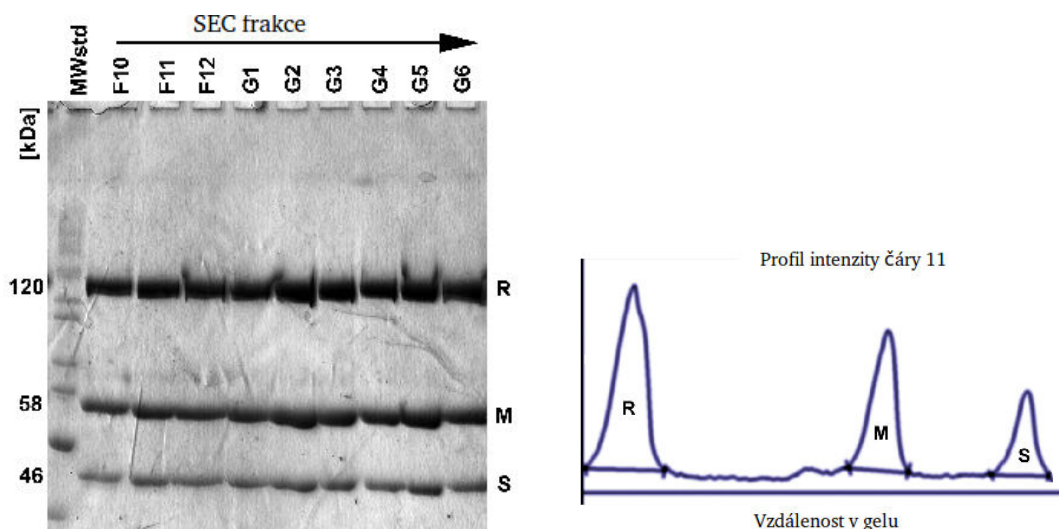


Obr. 3.2: (A) EcoR124I+DNA (zavřený stav) EM v negativním kontrastování. (B) EcoR124I bez DNA (otevřený stav) EM v negativním kontrastování. [8]

3.2 Příprava vzorku

Enzym byl připraven skupinou Prof. RNDr. Rüdiger Ettricha, PhD. z Ústavu nanobiologie a strukturní biologie (ÚNSB) na Nových Hradech a v Českých Budějovicích.

Při přípravě byla využita afinitní chromatografie (Ni-NTA) a gelová filtrace k purifikaci R_2M_2S komplexu. Kvantitativní analýza SDS-PAGE gelu (obr. 3.3) ukázala průměrnou stechiometrii 1.6(R):2.0(M):1.1(S) (obr. 3.3). To naznačuje, že jen zhruba polovina složených komplexů obsahuje obě HsdR podjednotky. Tento jev je způsoben výrazně rozdílnou vazebnou afinitou HsdR podjednotek vzhledem k M_2S komplexu. [12]



Obr. 3.3: Stanovení stechiometrie komplexu pomocí afinitní chromatografie.

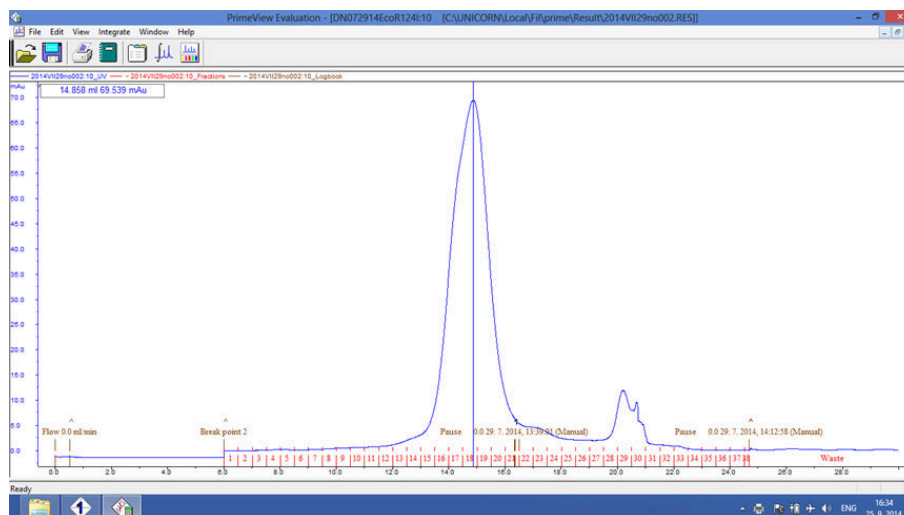
Prvotní dataset byl proveden v negative-stain (2% urany acetát) obr. 3.5 nahoře. Mikrografy odpovídají struktuře komplexu Eco R124I v jeho otevřené konformaci, jak ji publikovali autoři článku [8].

Dále byl komplex zmražen ponořením do tekutého etanu a nasnímán pomocí kryoEM. Dva datasety byly vytvořeny s 2 a 4 μm defokusem na mikroskopu FEI Tecnai F20 (3.5 uprostřed). Na mikrografu jsou jasně zřetelné částice o velikosti 120 Å, což odpovídá disociovaným částem komplexu. Kompletní $R_2M_2S_1$ komplexy jsou zdůrazněny šipkou. Ani na těchto snímcích však není jednoduché jednotlivé částice identifikovat. [12]

Po vícenásobném převážení vzorku i vzhledem k jeho stáří ¹ nebylo jisté, jestli je možné využít i nově zprovozněný mikroskop FEI Titan Krios. Pro zjištění stability byla provedena vylučovací chromatografie (obr. 3.4), která prokázala stabilitu

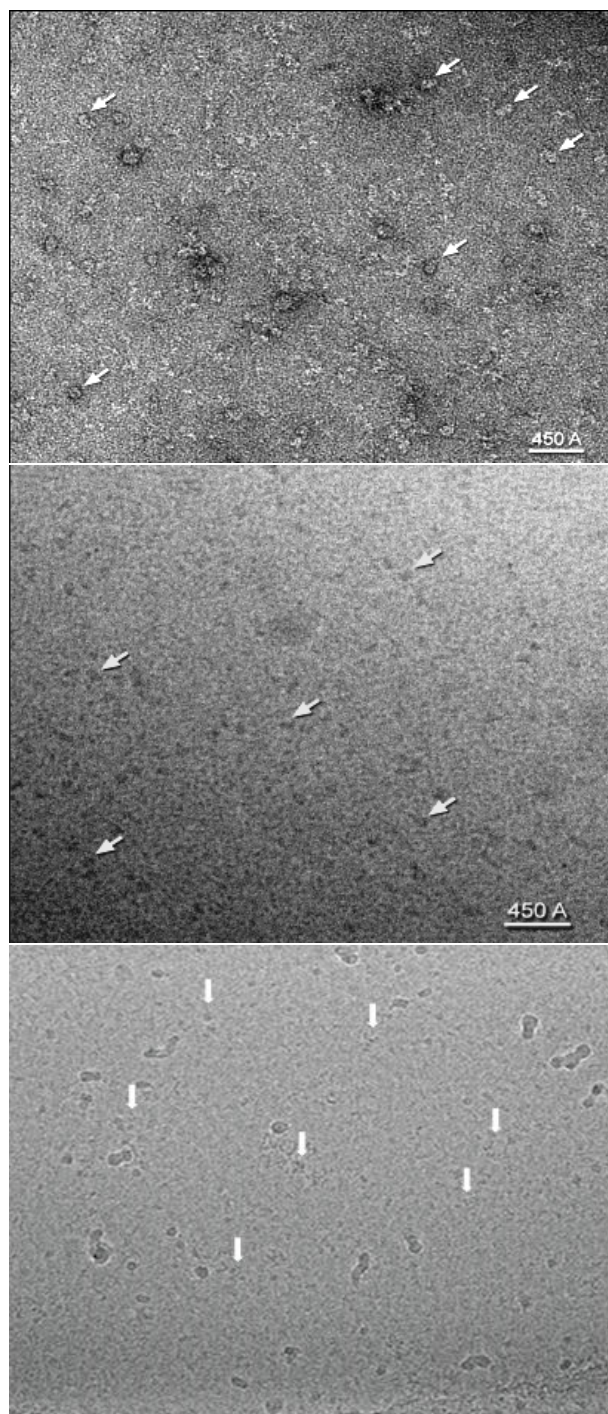
¹ Asi 100 μL roztoku s komplexem bylo skladováno při teplotě -80°C

celého komplexu s malým ramenem na levé straně hlavního píku poukazujícím na obsah R_1M_2S částic. Všechny dále zpracovávané mikrografy jsou tak již z tohoto mikroskopu (obr. 3.5 dole).



Obr. 3.4: Výsledek vylučovací chromatografie skladovaného purifikovaného komplexu R_1M_2S naznačuje jeho stabilitu během uskladnění.

Při snímání byla vyzkoušena technika "focal pairs", která spočívá v nasnímání mikrografu s malým defokusem ($2\ \mu\text{m}$) obsahujícím informace na hranici rozlišení $4\ \text{\AA}$ a následným snímáním stejné oblasti s vyšším defokusem ($6\ \mu\text{m}$), kdy je výrazně zvýšen kontrast částic vůči pozadí za cenu nižšího rozlišení v mikrografu. Druhý mikrograf s vysokým kontrastem pak lze využít pro identifikaci projekcí zobrazených molekul a pak přenést jejich souřadnice na první mikrograf s malým defokusem a s vysokým rozlišením. Zjistili jsme, že tento přístup není použitelný na mikroskopu FEI Titan Krios s přímým detektorem, kdy během prvního snímání byl vzorek zatížen dávkou $70\ e^-/\text{\AA}^2$ a během druhého snímání stejnou dávkou tak, že poškodila vzorek natolik, že částice nebylo možné v mikrografu identifikovat. Tento přístup se tedy pro takto malé ($450\ \text{kDa}$) biologické molekuly neosvědčil a museli jsme použít data naměřená pouze s jednou exozicí.



Obr. 3.5: Komplex EcoR124I zobrazený pomocí negativního kontrastování (2% uranylacetát) (nahore). Stejný komplex zobrazený v kryomikroskopu Tecnai TF20 při urychlovacím napětí 120kV s koncentrací vzorku 0.3 mg/ML (uprostřed). Tentýž komplex zobrazený pomocí kryomikroskopu Titan Krios při urychlovacím napětí 300kV s koncentrací vzorku 10 mg/mL. (dole).

4 TOMOGRAFIE

Jeden z možných přístupů k získání 3D struktury makromolekulárních komplexů z dat kryoEM je elektronová tomografie. Metodou je možné vytvářet rekonstrukce částic od velikosti jednotlivých enzymů, přes viry až celé bakterie a eukariotické buňky, je-li využita filtrace neelasticky vychýlených elektronů. Tato metoda je užitečná v zobrazování jen jedenkrát se vyskytujících heterogeních entit v rozlišení 50 - 100 Å, což pokrývá oblast mimo možnosti světelné mikroskopie. [2]

4.1 Princip

Elektronová tomografie spočívá v nasnímání série mikrografů s rozdílným úhlem jejich normály vůči primárnímu svazku. Tím vzniká dataset obsahující trojrozměrné informace o zobrazovaném vzorku promítnuté do jednotlivých 2D snímků.

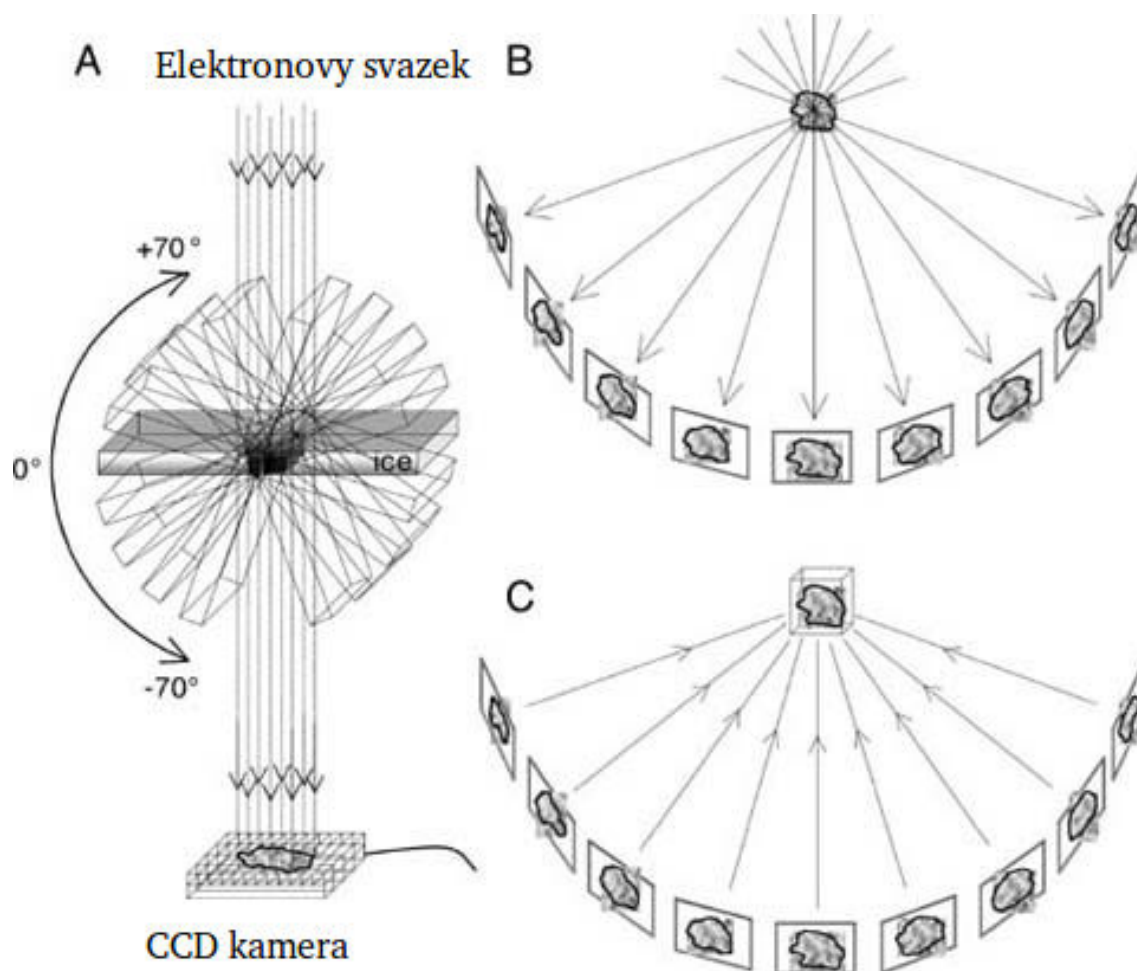
Vzorek musí být dostatečně tenký, aby jím dopadající elektrony prošly. Typicky by tloušťka neměla přesáhnout 0,5 μm pro urychlovací napětí 300 kV. Individuální tomogramy však nedosahují takového rozlišení, aby je bylo možné využít při popisu molekulárních komplexů. Poměr signál/šum (SNR) je na to příliš malý. To se obchází vytvořením mnoha subtomogramů identické částice a jejich následným průměrováním. Tím je možné posunout mez rozlišení na 20 - 30 Å. [2]

4.2 Sběr dat

Mikrografy se zaznamenávají v co největším rozsahu náklonů (typicky mezi -70° až 70°). Schéma je na obr. 4.1. Limitující je zvětšování délky trajektorie elektronového svazku skrz vzorek. Tomogram tak postrádá informace z klínového úseku 60° což vede k distorzi výsledného objektu. Vysoké úhly též zvláště na tlustších vzorcích vedou k nárůstu neelastického rozptylu a vícenásobným elastickým srážkám, což snižuje podíl koherentních elektronů použitelných při tvorbě obrazu. Pro zvýšení kontrastu se využívá jejich energiová filtrace. [13] Dávka (počet elektronů na \AA^2) není díky změnám geometrie konstantní – má kosinové rozložení.¹ Další aspekt vyplývající z naklánění vzorku je změna defokusu aplikovaného na jedno místo v čase, nebo ve stejném čase na různá místa. Oba efekty vedou k nutnosti provedení normalizace dat během jejich zpracování.

Při snímání vzorku EcoR124I byl využit rozsah -66° až $+66^\circ$ s krokem 2° což dává výslednou sérii 66 mikrografů.

¹Poměr dávky I při nulovém náklonu a dávky I_{60} při náklonu 60° je přibližně 1,4 .



Obr. 4.1: Princip kryoelektronové tomografie. (A) Vzorek je nakláněn v tzn. tiltovací sérii a postupně zaznamenáván. (B) Jednotlivé projekce se pro představu umísťují na virtuální plášť válce se středem ve vzorku. (C) 3D struktura se dopočítává zvolenou rekonstrukční metodou. [14]

4.3 3D rekonstrukce

Pro praktické provedení rekonstrukce je využit programový balík Bsoft. Podrobný popis jednotlivých bloků i příkazů uvedených v dalších podkapitolách je na adrese [http://lsbr.niams.nih.gov/bsoft/#záložka Usage/Workflows/Tomography](http://lsbr.niams.nih.gov/bsoft/#záložka%20Usage/Workflows/Tomography). [15]

Prvním krokem je normalizace všech zpracovávaných mikrogramů, po které následuje sledování posunu zlatých značek napříč jednotlivými snímky jako funkce jejich náklonu. Využívají se částičky koloidního zlata, které jsou pro elektrony velmi denzní a tak mají velký kontrast vůči svému okolí. U kryoEM se značky přimíchávají do suspenze vzorku před jeho zmrazením. Přesnost tohoto sledování určuje výslednou kvalitu tomogramu a tak hraje důležitou roli při zpracování. Celá procedura se provádí iterativně v zobrazovacím programu *bshow*, kde se ručně opravují špatně sledované značky a následným přepočítáním nových koordinát pomocí příkazu *btrack*. Úkolem je snížit reziduální chybu na co nejnížší úroveň.

Samotná rekonstrukce je pak realizována skriptem *tomrec*, kterému se jako parametry předávají: název zpracovávaného datasetu ve formátu pif, požadované rozlišení v Å, velikost dat, tloušťka tomogramu v pixelech, měřítko, přeškálování na průměrnou hodnotu a směrodatnou odchylku, průměr značek pro jejich odstranění a název výstupního souboru. Skript je volně stažitelný ze stránek Bsoftu. Rekonstrukce je prováděna ve fourierově prostoru pomocí následujících tří funkcí:

1. *bmgft* - generuje fourierovu transformaci všech mikrogramů
2. *btomrec* - počítá řezy ve fourierově prostoru a jejich zpětnou transformaci do prostorové oblasti
3. *bzfft* - shlukuje řezy a provádí zpětnou transformaci výsledného 3D objemu

4.3.1 Nelineární anisotropická difuze

Při filtrování vypočtených tomogramů se využívá nelineární anisotropická difuze (Non-linear Anisotropic Diffusion - NAD). Základní myšlenka je analogická k fyzikální difúzi. Jde o nehomogenní rozložení šedotónových hodnot v prostoru, které je možné upravovat a zvyšovat tak jeho homogenitu – vyhlazovat šum. Základní myšlenkou je kontrola velikosti a směru difusního toku v závislosti na okolní struktuře. Například difusní tok napříč hranami bude malý kvůli jejich zachování, zatímco oblasti bez výrazných hran budou vyhlazeny velkým difusním tokem.

Matematicky je difusní tok dán parciální diferenciální rovnicí,

$$\mathbf{j} = -D\nabla I(\mathbf{x}, t) \quad (4.1)$$

kde $I(\mathbf{x}, t)$ je rozložení šedotónových hodnot v obraze v závislosti na čase. Gradient ∇I tak způsobuje tok \mathbf{j} kompenzující tento gradient. D je pak difusivita prostředí.

Změna intenzity je pak dána vztahem

$$\frac{\delta}{\delta t} I(\mathbf{x}, t) = -\text{div} \mathbf{j} = \text{div}(D \nabla I). \quad (4.2)$$

Za předpokladu konstantního D je lineární anisotropická difuze ekvivalentní gaussově filtraci. Pro zachování hran v signálu je třeba substituovat D funkcí

$$g(|\nabla I|^2) = \frac{1}{1 + |\nabla I|^2 / \kappa^2}. \quad (4.3)$$

To změnilo rovnici na nelineární, kde je tok j měněn difusivitou D závislou na okolním obraze. Tok okolo hran je tak redukován, což je sice zachovává, ale na druhou stranu nezahluje šum v jejich okolí. Řešením je přidání dalšího vhodného toku J_σ probíhajícího paralelně s hranami. Ten se vypočítává konvolucí ∇I^T s lokální směrodatnou odchylkou σ . Strukturní tensor se stanovuje pro každý voxel zvlášť po kterém následuje výpočet vlastních hodnot μ_i a vektorů \mathbf{v}_i této matice charakterizující lokální strukturní vlastnosti obrazu v sousedství o velikosti σ . V závislosti na těchto vlastnostech může být difusní tok rozložen na komponenty ve směru vlastních vektorů \mathbf{v}_i . Porovnáním jednotlivých koeficientů je možné určit směr probíhající hrany a zvolit tak odpovídající směr filtrace, který provede požadované vyhlazení v homogenních oblastech a přitom zanechá hrany v obraze. [16]

4.3.2 Sběr částic

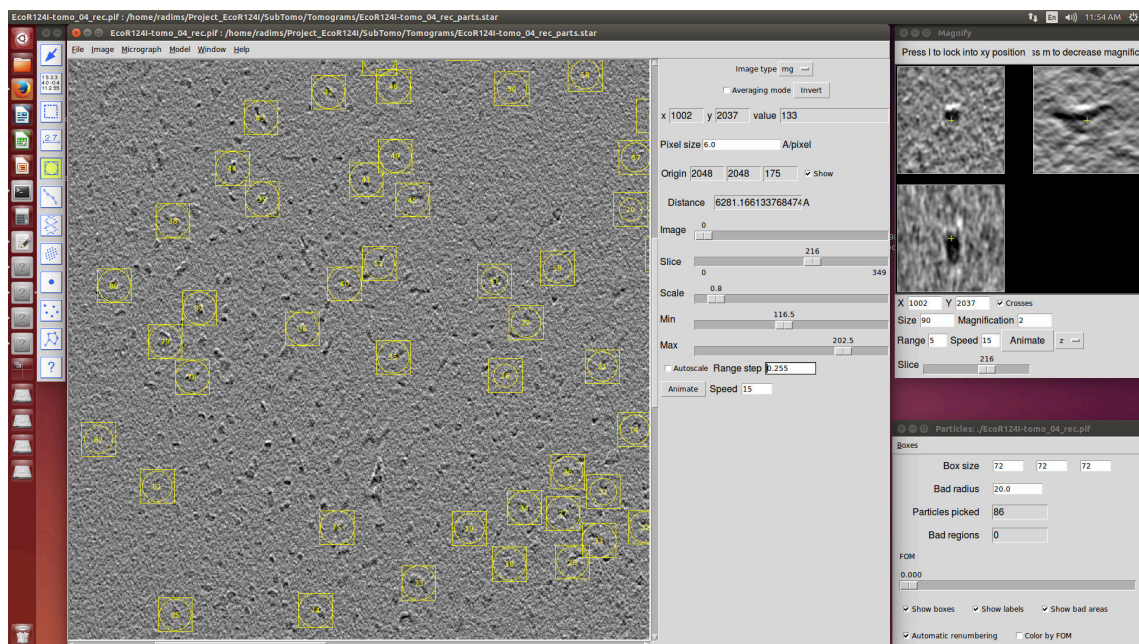
Dalším krokem je sběr částic v *bshow*, uložení jejich koordinát do souboru *EcoR124I_parts.star* (celkově 182 částic) a extrakce prováděná příkazem

```
bpick -verb 7 -recon -extract 72 -back -norm -split -center -resol 50,500 -sampling 6,6,6
-base EcoR124I_ -path Parts/TM4 -extension mrc -out Parts/EcoR124I-tomo_07_rec_parts.star
Tomograms/EcoR124I-tomo_07_rec_parts.star
```

4.3.3 Průměrování subtomogramů

Na obrázku obr. 4.2 jsou vidět sebrané částice označené žlutým prostorovým boxem o hraně 72 pixelů. Posledním krokem je výběr jedné kvalitní částice a její použití jako reference pro subtomogram alignments (zarovnání a zprůměrování částic v rámci celého tomogramu). To významně zlepší SNR a také se odstraní problém chybějící části dat. Každá jednotlivá částice má totiž svoji odlišnou orientaci, a proto i svoji chybějící část. Jejich zprůměrováním se zmenšuje část dat, která chybí, a rekonstrukce se tak přibližuje originálu. Prakticky se provádí příkazem:

```
bcat -ver 7 -out Parts/Parts_TM4.pif Parts/TM4/*.mrc
-> zvolená částice EcoR124I_001048.mrc
-> cp Parts/TM4/EcoR124I_001048.mrc Refs/
-> bmask -ver 7 -data float -create 72,72,72 -radius 0,35 -origin 36,36,36 EcoR124I_mask0.map
```



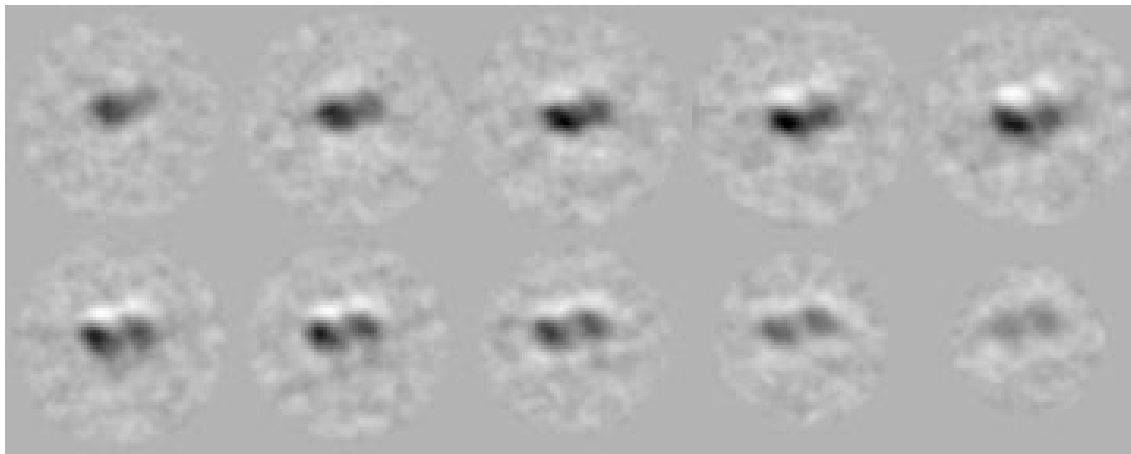
Obr. 4.2: Výběr částic pro výpočet zprůměrovaného tomogramu enzymu EcoR124I. V pravém horním rohu je pohled na jednu z částic ve všech třech rovinách.

3D zarovnání je iterativní proces s postupnou proměnou zadávaných parametrů od globálního (global) zarovnání ke směrovému (directional) a nakonec zjemňujícímu (refine). Ty jsou uloženy do textového souboru `alignxx.cnfg`, kde `xx` je pořadové číslo iterace. Samotné zarovnání je provedeno skriptem `spt-paliave` dostupným na stránkách Bsoft.

Výsledky cryoelektronové tomografie enzymu EcoR124I dále slouží jako iniciální model v jednočásticové analýze. S jejich pomocí je prováděno hledání částic v datech (particle picking) a slouží i jako iniciální model při výsledné rekonstrukci. Během práce na projektu byly vytvořeny dva iniciální modely.

4.4 Výsledek zpracování 1. datasetu

Výsledná objemová rekonstrukce prvního datasetu neposkytla kýžené výsledky (obr. 4.3). Na řezech jsou patrné dvě podjednotky – jedna chybí. Tudíž není její následné využití vhodné a bylo nutné provést nové zpracování s pečlivějším výběrem zahrnutých částic. Do extrahovaných subtomogramů pravděpodobně zasahovaly sousední částice, které změnilы výsledek rekonstrukce.



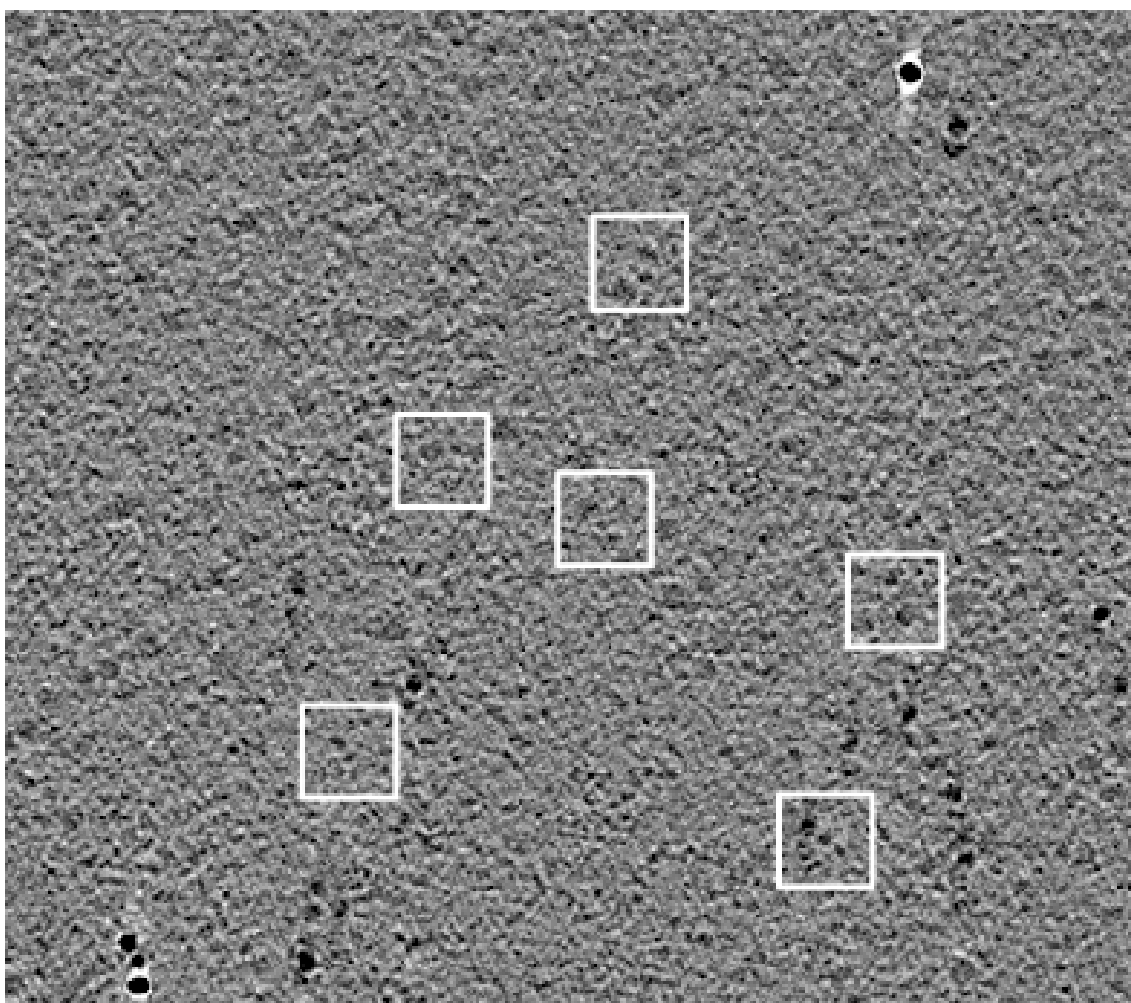
Obr. 4.3: Výsledek první objemové rekonstrukce enzymu EcoR124I. Na jejích řezech jsou patrné dvě podjednotky místo tří. Z tohoto důvodu není model dostatečný a musí být provedena další rekonstrukce.

4.5 Výsledek zpracování 2. datasetu

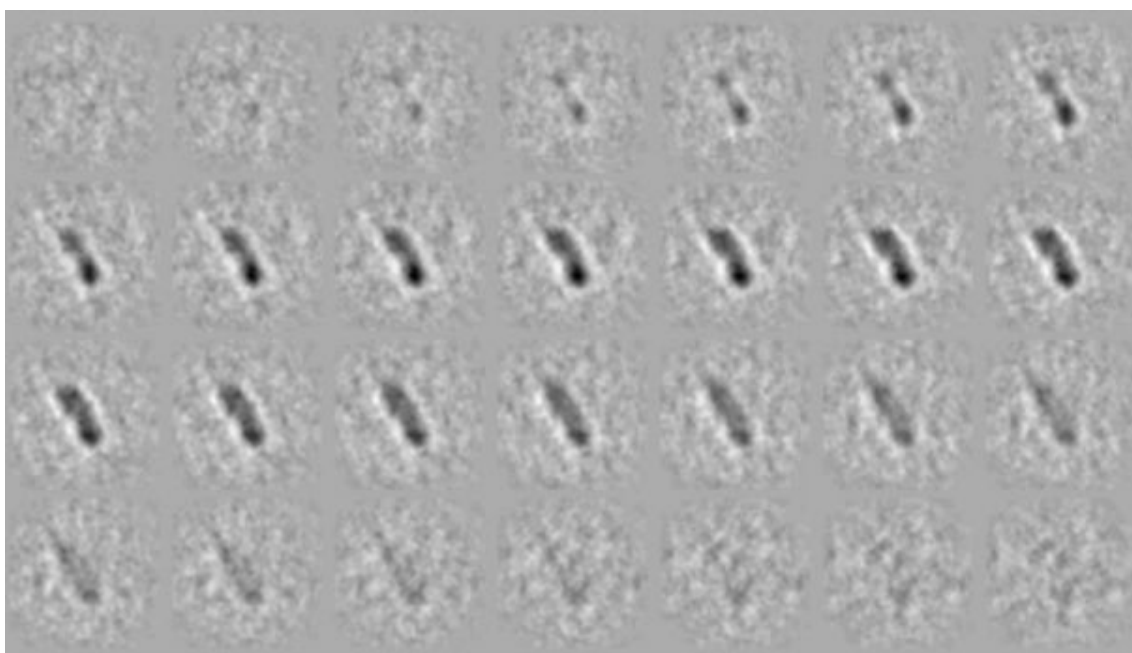
Druhý zpracovaný dataset již dává lepší výsledky. Na jednotlivých řezech jsou rozeznatelné všechny tři podjednotky (obr. 4.5), což umožňuje její využití jak při hledání částic (particle picking) tak i jako iniciálního modelu v jednočásticové analýze.

Jako reference bylo použito jen 40 % nejlépe zarovnaných částic. To snížilo rozmazání celé struktury vzniklé vlivem nesprávně zarovnaných částic. Prakticky provedeno příkazem:

```
bpartsel -ver 4 -top 40 -out Avg05/Align05_top40.star Align05/TM4/EcoR124I_001*.star
brecadd -ver 4 -rescale 0,1 -select 1 -average Avg05/Sel40_avg.mrc
-fom Avg05/Sel40_var.mrc Avg05/Align05_top40.star
```



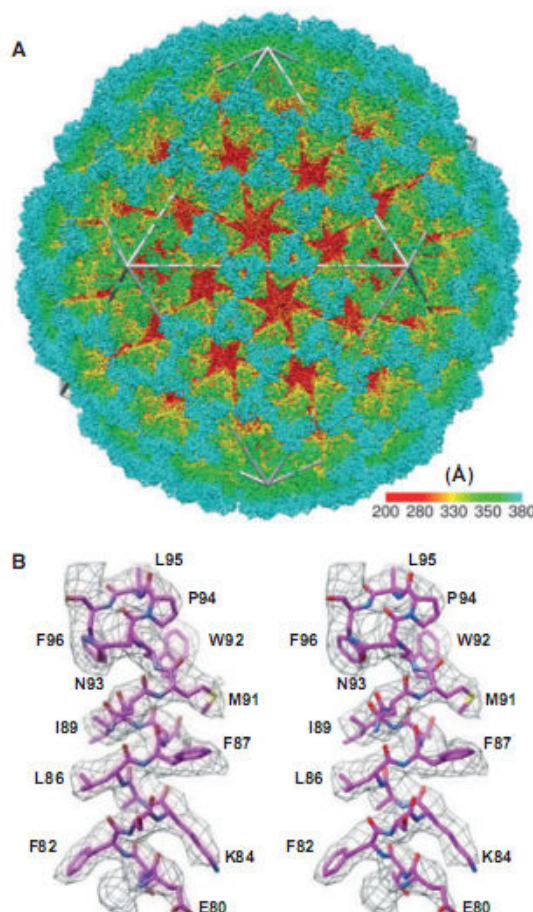
Obr. 4.4: Druhý vypočítaný tomogram, ze kterého byly extrahovány částice pro další zpracování.



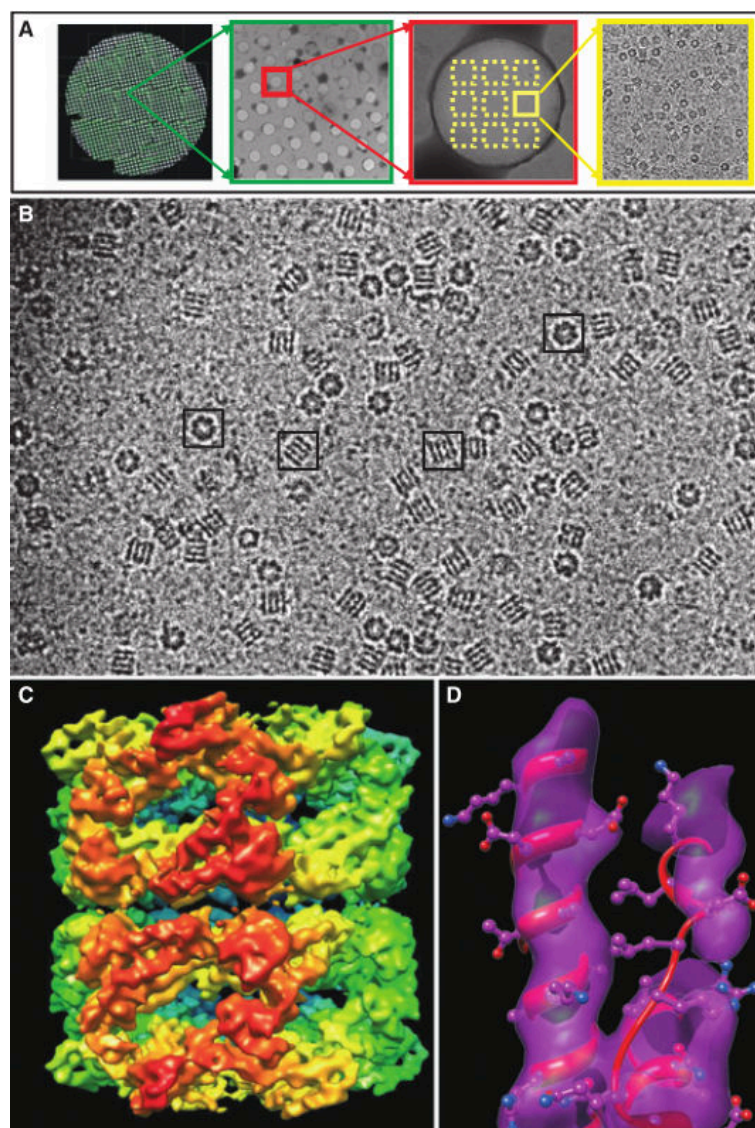
Obr. 4.5: Jednotlivé řezy druhé objemové rekonstrukce enzymu EcoR124I, na jejíchž řezech jsou patrné všechny tři podjednotky.

5 JEDNOČÁSTICOVÁ ANALÝZA

Obecně využívaným označením pro jednočásticovou analýzu je v mikroskopické obci anglický výraz single particle analysis (SPA). Tato technika generuje 3D rekonstrukci struktury z velkého počtu 2D projekčních obrazů identických kopií proteinových komplexů v rozdílných orientacích. Samotné zpracování je možné v mnoha programových balících např. EMAN, SPIDER, IMAGIC. Pro nafitování atomových struktur získaných z krystalografie nebo NMR se využívá např. UCSF Chimera navržená jak pro tuhé tak i flexibilní fitování. Přehled využitých programových balíčků včetně jejich možného využití je v příloze B. V případě komplexů s vysokou symetrií je možné dosáhnout téměř atomárního rozlišení – 3 Å (obr. 5.9), při nižších stupních symetrie pak 7 Å (obr. 5.2) [2]



Obr. 5.1: Určení struktury neobaleného ikosahedrálního viru pomocí kryoEM. (A) Vizualizace celé struktury. (B) Vybraný region demonstrující téměř atomové rozlišení, které je možné získat u vzorků s velkou symetrií. [2]



Obr. 5.2: Automaticky zpracovaná struktura pomocí SPA. (A) Obrázky sítky pro kryoEM postupně v narůstajícím zvětšení zleva doprava až po zvětšení, při kterém jsou rozpoznatelné jednotlivé molekuly. (B) Vzorový projekční obraz zmraženého zavodněného vzorku purifikovaného proteinového komplexu GroEL. Ve čtverečích jsou zvýrazněny komplexy s různou orientací vůči ose svazku. (C) 3D rekonstrukce sestavená z přibližně 28000 jednotlivých projekcí s výsledným rozlišením 7 Å. Iniciální model byl spočítán ze 2000 GroEL částic. Zpřesňování iniciální rekonstrukce byla provedena téměř automatickou procedurou implementovanou v softwarovém balíku FREALIGN. (D) Demonstrace, že dosažené rozlišení je dostatečné k zobrazení α -helixů ilustrovaná superpozicí mapy hustoty a odpovídajícího polypeptidového řetězce určeného pomocí rentgenové krystalografie (PDB ID: 3E76). [2]

Kritická část celého zpracování je správné automatické rozpoznání jednotlivých částic v individuálních mikrografech. V případě vzorku EcoR124I je tato část na surových datech velmi obtížná (obr. 5.3 nahoře). Jejich výběr se provádí buďto automaticky, nebo ručně – tato metoda je ale při obrovských počtech nutných částic velmi časově náročná. Model vytvořený pomocí tomografie (kap. 4) byl využit pro vytvoření banky projekcí s jejíž pomocí byly hledány částice ve vyfiltrovaných mikrografech (příklad filtrace na obr. 5.3 uprostřed a dole). Před jejich výběrem je nutné provést korekci transformační funkce kontrastu (CTF – contrast transfer function) popsanou v kapitole 5.3.

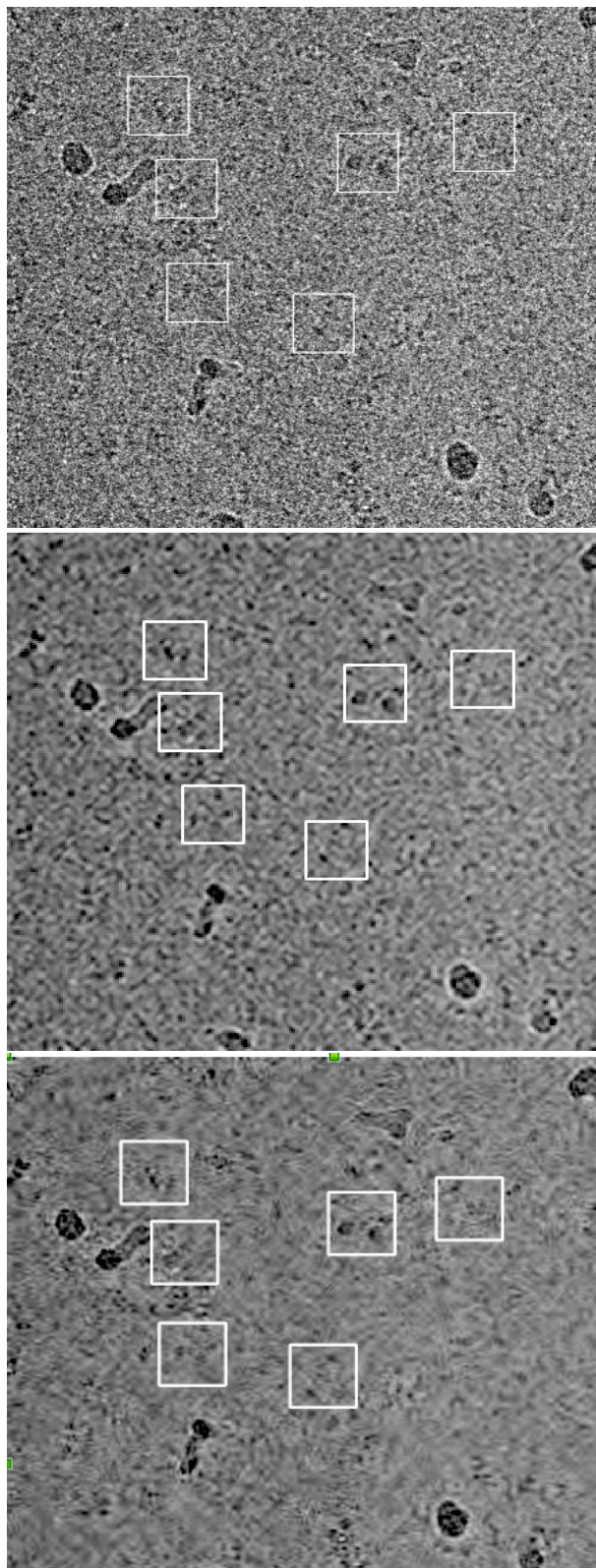
Dále se postupuje jejich rozčleněním do příslušných skupin. Využívají se statistické přístupy jako analýza hlavních komponent, multivarietní analýza nebo kovarianční analýza. Částice se rozdělují do jednotlivých tříd na základě jejich stupně podobnosti. Příbuzné projekce se pak zprůměrují a vznikají charakteristické projekce studovaného komplexu s mnohem vyšším SNR. Pro generování 3D rekonstrukce je potřeba znát relativní orientace všech zahrnutých částic. Při výpočtu se využívá centrální řezový teorém, který pro 3D objekt říká, že fourierova transformace každé 2D projekce je totožná s odpovídajícím řezem 3D spektra. Jako iniciální model lze využít výsledek tomografické rekonstrukce.[2] Podrobný popis jednotlivých kroků je v následujících kapitolách.

Během zpracování se pořád dokola opakují dva druhy souborů. Obrazové, které obsahují samotné obrazové informace (formát mrc, mrsc), a parametrické textové soubory s metadaty (formát star). Ty také obsahují informace o prováděných výpočtech vzhledem ke svému mateřskému mikrografu (koeficienty CTF korekce, souřadnice částic při jejich sběru atd.) Tvorba STAR souborů se provádí přes grafický interface balíku RELION, ve kterém se vybírají jen dostatečně kvalitní snímky (selected_micrographs.star). Vyřazují se ty neostře, obsahující velké kontaminace nebo s popraskaným QUANTIFOILEM¹. Ve zpracovávaném datasetu zbylo po jejich prohlídce 1413 mikrografů. Odkazy na jednotlivé mikrografy se zapisují do textového souboru ve tvaru:

```
data_

loop_
_rlnVoltage #1
_rlnDefocusU #2
_rlnDefocusV #3
_rlnDefocusAngle #4
_rlnSphericalAberration #5
_rlnDetectorPixelSize #6
_rlnCtfFigureOfMerit #7
_rlnMagnification #8
```

¹Děrovaný uhlíkový film kryjící sítku pro TEM s definovanou velikostí děr, jejich tvarem a uspořádáním



Obr. 5.3: Surový mikrograf enzymu EcoR124I „otevřený“ (FEI Titan Krios, urychlovací napětí 300 kV, defokus $-4\mu\text{m}$) nahoře. Stejný mikrograf po filtraci pásmovou propustí s mezními frekvencemi 20 a 300 Å (uprostřed) a pomocí dyadické vlnkové transformace s metodou stanovení prahu NeighShrinkSure (dole).

```

_rlnAmplitudeContrast #9
_rlnImageName #10
_rlnCoordinateX #11
_rlnCoordinateY #12
_rlnNormCorrection #13
_rlnMicrographName #14
_rlnGroupNumber #15
_rlnOriginX #16
_rlnOriginY #17
_rlnAngleRot #18
_rlnAngleTilt #19
_rlnAnglePsi #20
_rlnAutopickFigureOfMerit #21
_rlnClassNumber #22
_rlnLogLikeliContribution #23
_rlnNrOfSignificantSamples #24
_rlnMaxValueProbDistribution #25
_rlnParticleSelectZScore #26
300.000000 19417.189453 19478.000000 15.300000 2.700000 14.000000 0.092750
46667.000000 0.100000 000009@Particles/micrographs/mg_0459_sel_particles_autopick_bin3.mrcs
3277.000000 1658.000000 0.499617 micrographs/mg_0459_sel.mrc 142 -7.398345
-4.398345 0.000000 0.000000 -161.991725 0.408557 1 12119.563824 3
0.321869 0.232409

300.000000 25912.630859 26225.710938 29.500000 2.700000 14.000000 0.071830
46667.000000 0.100000 000023@Particles/micrographs/mg_0577_sel_particles_autopick_bin3.mrcs
1522.000000 1984.000000 0.482759 micrographs/mg_0577_sel.mrc 37 7.601655
8.601655 0.000000 0.000000 88.008275 0.327057 1 12199.151452 1
0.393871 0.158423

```

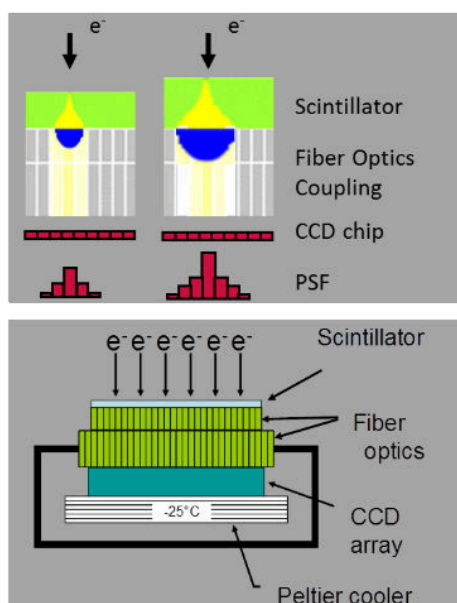
5.1 DDD a CCD kamery

Využitý mikroskop je vybaven kamerou Falcon II CMOS Direct Electron Detection. Hlavní odlišnost je již v základní technologii využívající CMOS architekturu, která umožňuje mnohem rychlejší akvizici dat oproti CCD. Nevýhodou CCD kamer je využití scintilátoru k převodu elektronů na fotony a čtení po jednotlivých řádcích.

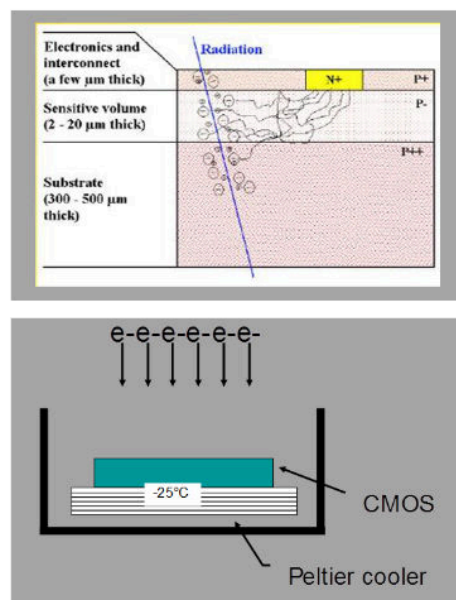
Přímá detekce elektronů u CCD není bohužel možná kvůli radiačnímu poškození a brzké saturaci signálu. Elektrony, dopadající na scintilátor s energií 300 keV, zanechávají stopu o FWHM okolo 30 μm , což limituje výsledné rozlišení. Další faktor omezující rozlišení je zpracování vzniklého světla vláknovou optikou nebo soustavou čoček. [17]

Přímá detekce zvyšuje kvantovou účinnost detekce (detector quantum efficiency - DQE) zmenšuje PSF a zvyšuje kvalitu dat včetně SNR. Mnohem rychlejší zpracování dané vlastní čtecí elektronikou pro každý pixel umožňuje zaznamenávat v rámci jednoho mikrogramu více subsnímů, což umožňuje jejich zpracování a úpravu i po skončení akvizice a tím zvýšení výsledné kvality zobrazení. [18] Srovnání základních komponent obou detektorů je na obr. 5.4

CCD: multi stage conversion of electron energy via fiber or lens optics



CMOS: direct conversion of electron energy without fiber or lens optics



Obr. 5.4: Základní rozdíl ve skladbě CCD detektoru elektronů (vlevo) a DDD využívající CMOS architekturu (vpravo).[18]

5.1.1 Zarovnání subsnímků

Kamera Falcon II zaznamenává jednotlivé snímky jako sérii sedmi subsnímků. Výsledný obraz vzniká jejich následným složením. To dává možnost provést před finálním složením zarovnání (alignment) subsnímků a tím omezit vliv driftu obrazu během snímání. Další výhodou je možnost vybrat a použít pro další zpracování jen data, na kterých jsou zachyceny zkoumané molekuly ještě nepoškozené aplikovanou dávkou.

Při zpracování dat vzorku EcoR124I byly kvůli zachování vysokého rozlišení (nutná nízká absorbovaná dávka) zarovnávány první tři části akvizice. Příklad příkazů provádějící zarovnání subsnímků mikrografu mg_1205.mrc :

```
#PBS -q short
#PBS -l nodes=1:ppn=1
#PBS -l mem=2gb
#PBS -l walltime=00:30:00
#PBS -j oe
source /storage/brno2/home/radims/.tcshrc

cp Frames/FoilHole_15843980_Data_15847842_15847843_20150419_0805_frames.mrc
micrographs/mg_1205_movie.mrcs
bsplit -images -first 0 -digits 1 micrographs/mg_1205_movie.mrcs:mrc micrographs/tt_1205.mrc
bmg -number 3 -Volt 300 -Cs 2.7 -Ampl 0.05 -out micrographs/movie_1205.star
micrographs/tt_1205_0.mrc micrographs/tt_1205_1.mrc micrographs/tt_1205_2.mrc
bseries -ver 1 -align 2 -sampling 1.0745 -resol 300,5 -image micrographs/mg_1205_algn.pif
micrographs/movie_1205.star
bimg -average micrographs/mg_1205_algn.pif micrographs/mg_1205_sel.mrc
bmg -number 7 -Volt 300 -Cs 2.7 -Ampl 0.05 -out micrographs/movie_1205.star micrographs/tt_1205_?.mrc
bseries -ver 1 -align 2 -sampling 1.0745 -resol 300,5 -image micrographs/mg_1205_algn.pif
micrographs/movie_1205.star
bimg -average micrographs/mg_1205_algn.pif micrographs/mg_1205_all.mrc
rm micrographs/movie_1205.star
rm micrographs/mg_1205_algn.pif
rm micrographs/tt_1205_?.mrc
```

5.2 Filtrace, úprava histogramu

Surová data jsou bohužel příliš zašuměná a rozeznávání jednotlivých částic je velmi obtížné. To vedlo k potřebě provést filtraci k jejich zvýraznění. První dva uvedené způsoby filtrace (gaussovský filtr a pásmová propust) byly realizovány v balíku Bsoft. Pro výpočet filtrace pomocí vlnkové transformace byly zarovnané snímky ve formátu mrc přepsány do formátu tif, který je možné načíst jako vstupní soubor v programovacím prostředí MATLAB, kde byl samotný výpočet realizován.

5.2.1 Gaussovský filtr

Nejjednodušší zkoušenou variantou byl gaussovský filtr odstraňující vysokofrekvenční šum za cenu vyhlazování hran v obraze. Gaussovský filtr spadá do skupiny lineárních

filtrů, kde nová hodnota pixelu vzniká z jeho hodnoty a hodnot okolních pixelů po započítání váhovací funkce. Ta je invariantní vůči posunu po filtrovaném obraze a zachovává si svůj tvar. Filtrace se provádí konvolucí filtrovaného obrazu s filtračním jádrem. V případě gaussovského filtru je jím jádro 2D gaussovy funkce v její diskretní verzi. Bohužel zvýšení rozlišitelnosti hledaného komplexu EcoR124I v mikrofotografech nebylo dostatečné, a tak nebyl tento typ filtru dále používán.

5.2.2 Pásmová propust

Zkoušená pásmová propust je lineární filtr propouštějící jen signál o určitých frekvencích. Prakticky byla realizována programem *bfilter*. Rozsah zkoušených mezních frekvencí byl na horní hranici 300 – 400 Å, na dolní hranici pak 10 – 30 Å. Vyfiltrovaný mikrofotograf s mezními frekvencemi 20 Å a 300 Å je vidět na obr. 5.3 uprostřed. Výsledky však nesplňovaly kladené nároky.

5.2.3 Prahování koeficientů vlnkové transformace

Vlnková transformace používá pro rozklad signálu bázi časově omezených funkcí nazývaných vlnky. To umožňuje lokalizovat nejen přítomnost dané frekvence, ale i lokalizovat její pozici. Vlnková transformace je ve své spojitě podobě s konečnou energií definována funkcí

$$w(s, p) = \int_{-\infty}^{\infty} f(t) \cdot \frac{1}{\sqrt{s}} \overline{\psi\left(\frac{t-p}{s}\right)} dt \quad (5.1)$$

kde $f(t)$ je analyzovaný signál, ψ je vlnka, $\overline{\psi}$ je komplexně sdružená vlnka, s měřítko, p poloha umístění vlnky v čase, t čas a $\frac{1}{\sqrt{s}}$ normalizuje energii vlnky při změně měřítka. V případě symetrické vlnkové funkce se postupně hledá míra podobnosti (korelace) mezi vstupním signálem a měnící se vlnkou v závislosti na velikosti extrahovaných detailů.

Inverzní transformace se pak vypočítává jako superpozice vlnkových koeficientů

$$f(t) = \sum_{s,p} w(s, p) \cdot \psi(s, p, t). \quad (5.2)$$

Pro diskretní systémy je množství vlnkových koeficientů generované CWT zbytečně vysoké a proto se využívá diskretní vlnková transformace využívající jen koeficienty s měřítky $s = 2^j$, kde $j = 0, 1, 2, \dots, n$. Vhodnou dvojkovou závislostí parametrů s a p lze vytvořit z vyhovující vlnky ψ ortonormální bázi s $s = 2^j, p = k \times 2^j, j, k \in \mathbb{Z}$. Způsob výpočtu dopředné i zpětné transformace je pak analogický k jeho spojitě formě.

Prakticky se výpočet provádí pomocí Mallatova algoritmu rychlé vlnkové transformace, který využívá soubor kvadraturních zrcadlových filtrů typu dolní a horní propust. Tento způsob výpočtu se nazývá dyadickou vlnkovou transformací (DWT), při které dojde po každé filtraci k podvzorkování s faktorem 2. Druhým používaným řešením je nařazení impulzní charakteristiky filtru nulami. Tento způsob je nazýván stacionární vlnkovou transformací (SWT). Dekompozice se opakuje tak dlouho, jak je potřeba pro další zamýšlené zpracování nebo dokud je délka signálu větší než jedna. V případě dvojrozměrného signálu se filtrace provádí po řádcích a následně po sloupcích. Výsledkem jsou čtyři pásma v jednom rozkladovém kroku. Rekonstrukce se pak provádí pomocí banky zrcadlových filtrů s interpolací. [19]

Po dekompozici jsou užitečné složky signálu obsaženy v koeficientech s velkou amplitudou. To dává možnost omezit šum vhodným prahováním koeficientů. Vzhledem k zamezení vzniku nespojitostí v obrazu bylo zvoleno měkké prahování dané vzorcem

$$w' = \begin{cases} \operatorname{sgn}(w) \cdot (|w| - th), & |w| \geq th \\ 0, & |w| < th \end{cases} \quad (5.3)$$

Stanovení prahu th je pak možné vícero způsoby. Vyzkoušeny byly:

VisuShrink threshold

Je univerzální práh pro všechny stupně dekompozice $th = \sigma\sqrt{2\log m}$ kde σ je směrodatná odchylka šumu a m je množství pixelů v obraze.[20]

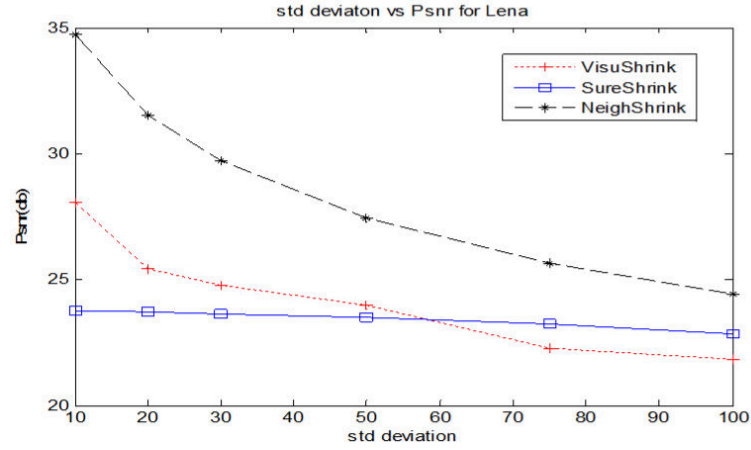
SureShrink threshold

Cílem tohoto prahu je minimalizovat MSE (mean square error) v rámci každého pásma. Prah je definován jako $t^* = \min(t, \sigma\sqrt{2\log m})$ kde t značí hodnotu minimalizující Steinův estimátor, σ je směrodatná odchylka šumu a m je velikost obrazu. [20]

NeighShrinkSure threshold

Tento typ prahu počítá samostatný práh pro každé pásmo a navíc během výpočtu bere v potaz velikost koeficientů v okolí vyšetřovaného bodu a lokálně tak upravuje práh i v rámci jednoho pásma. Pro případ okna o velikosti 3×3 se středem ve měněném koeficientu o souřadnicích (i,j) je tento práh upravován pomocí funkce $\Gamma_{ij} = [1 - \frac{T_u^2}{S_{ij}^2}]$, kde T_u^2 je univerzální práh a S_{ij}^2 je druhá mocnina sumy všech vlnkových koeficientů ve zvoleném okně. [20]

Na základě srovnání metod vypočítávajících práh užitých při filtraci Leny v závislosti na směrodatné odchylce šumu v obraze uvedeném v článku [19] jsem se



Obr. 5.5: Srovnání metod VisuShrink, SureShrink a NeighShrinkSure pro určení prahu během filtrace užitím DWT. [20]

rozhodl využít pro filtraci mikrogramů metodu NeighShrinkSure. Tato metoda dávala nejlepší výsledky i při přímém porovnání testovacího souboru zpracovávaných mikrogramů. Po filtraci následovala úprava histogramu a uložení obrazu do nového souboru formátu tif. Po nahrání zpět na disk metacentra byl proveden přepis zpět do formátu mrc, který umožňuje další zpracování.

Celkově byla filtrace provedena na 1806 mikrogramech použitých při konečném výpočtu struktury studovaného enzymu. Při velikosti jednoho z nich 32 MB to dělá asi 60,5 GB dat. Výpočetní náročnost tohoto kroku byla přibližně 30 hodin na 16 jádrech.

5.3 Korekce přenosové funkce kontrastu

Každý reálný objekt je v průběhu zobrazení modifikován v závislosti na vlastnostech zobrazovací soustavy. Původní obraz konvoluje s PSF mikroskopu což způsobuje, že výsledný snímek neodpovídá reálnému rozložení density vzorku. Ve frekvenční oblasti je tento vztah možné zapsat rovnicí

$$F\{PSF(\vec{r})\} = CTF(\vec{R}) \cdot E(\vec{R}) \quad (5.4)$$

kde \vec{r} a \vec{R} jsou rozložení prostorových frekvencí a F značí fourierovu transformaci.

Změny amplitudového a fázového kontrastu způsobené aberací objektivové čočky (CTF) je popisován funkcí

$$e^{i\gamma} \quad (5.5)$$

kde γ je dána výrazem

$$\gamma(\vec{R}) = -2\pi \left(\frac{1}{2} \Delta \lambda \vec{R}^2 - \frac{1}{4} C_S \lambda^3 \vec{R}^4 \right) \quad (5.6)$$

kde γ je fázový posun způsobený aberacemi, \vec{R} je prostorová frekvence ve fokální rovině objektivové čočky, C_S je koeficient sférické aberace, λ je vlnová délka elektronů primárního svazku a Δ je aplikovaný defokus (vzdálenost obrazové roviny od opravdové fokální roviny).

Vztah popisuje fázový posun způsobený sférickou aberací a aplikovaným defokusem. Výsledkem je, že některé z prostorových frekvencí jsou přenášeny se záporným kontrastem a nebo s nulovým přenosem. Proto se dataset snímá při více defokusech, čímž se řeší problém chybějících částí. [13]

Pro získání kvalitních dat je tedy nutné odstranit vliv mikroskopu a provést CTF korekci.

5.3.1 Manuální korekce

Manuální korekce spočívá ve spočítání rotačního průměru výkonového spektra obrazu, což je ale možné pouze při absenci astigmatismu, který musí být velmi dobře korigován. Amplitudový profil (druhá odmocnina intenzity) se ručně porovnává s modelem, který se uživatel snaží co nejlépe nafitovat na spočítanou křivku.

5.3.2 Automatická korekce

Nejjednodušší metoda CTF korekce je otočit fázi v částech spektra, kde $\sin \gamma$ mění svoje znaménko. V některých případech pouhá tato úprava postačí ke zpracování rekonstrukce za předpokladu, že není vyžadováno nejvyšší dosažitelné rozlišení.

5.3.3 Amplitudová korekce a Wienerova filtrace

Pokročilejší metoda je korekce jak fáze, tak amplitudy, u které jsou kompenzovány nejen samotné oscilace spektra, ale i jeho celkový úbytek na vyšších frekvencích pomocí úpravy obálky CTF. Korekce obrazu se provádí podle vztahu

$$Im_{COR} = F^{-1}\{F\{Im\}/F\{PSF\}\} = F^{-1}\{F\{Im\}/CTF\} \quad (5.7)$$

kde Im je zaznamenaný obraz a Im_{COR} obraz po korekci. Za předpokladu nulového šumu je možné provést korekce všude vyjma bodů s přenosem nula. V reálném případě malé hodnoty přenosu potlačují zájmový signál a relativně zvyšují zastoupení šumu v těchto oblastech. K potlačení tohoto jevu se provádí Wienerova filtrace zahrnující potlačení vlivu šumu.

$$Im_{COR} = F^{-1} \left\{ \frac{F\{im\}CTF}{CTF^2 + c} \right\} \quad (5.8)$$

kde $c = \frac{1}{SNR}$. [13, 25]

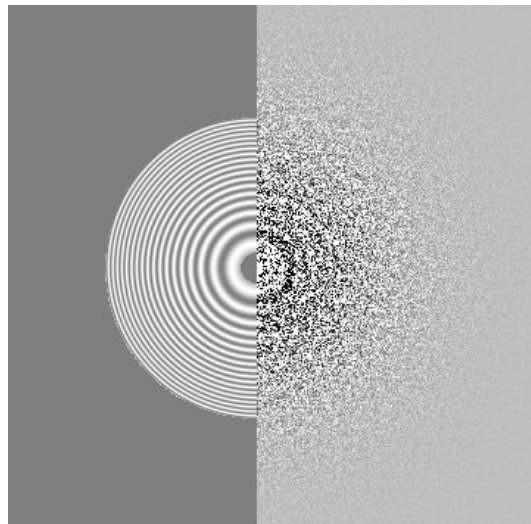
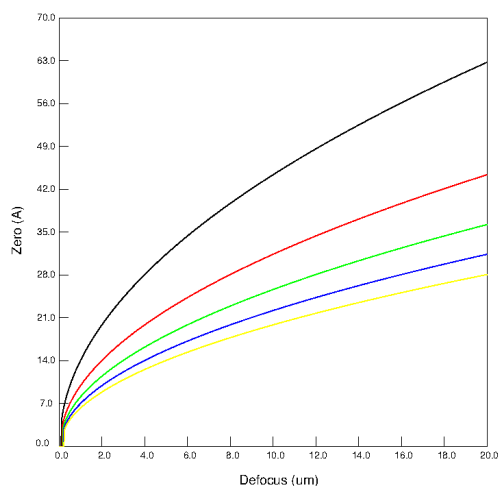
CTF korekce byla provedena programem ctffind3, který dokáže počítat i s daty zatíženými astigmatismem (odchyly aproximuje pomocí elipsy). Pro výpočet je nutné znát několik parametrů. Těmi jsou koeficient sférické aberace ($2,7mm$), urychlovací napětí (300 kV), zastoupení amplitudového kontrastu (0,1), fyzická velikost detekčního elementu v detektoru (Fallcon II: $14 \mu m$), zvětšení ($140\,000\times$), korigovaný rozsah frekvencí (od 5 do 100 \AA), korigovaný rozsah defokusů (od 0,5 do $5 \mu m$). Prakticky byl výpočet proveden skriptem:

```
#!/bin/tcsh
#PBS -N ctffind3           $ název výpočtu
#PBS -q normal             $ typ fronty, normal označuje možný výpočetní čas 24 hod
#PBS -l nodes=1:ppn=36:brno $ počet uzlů a procesorů na jeden uzel, jejich fyzické umístění
#PBS -l scratch=400mb      $
#PBS -l mem=36gb           $ velikost požadované paměti
#PBS -j oe                 $

# Environment
source ~/.tcshrc
# Submit: qsub -d 'pwd' relion_ctffit.pbs
# Command: mpiexec --bynode -n 20 XXXcommandXXX

mpiexec --bynode -n 36 'which relion_run_ctffind_mpi' --i "all_sel_micrographs.star"
--o "all_sel_micrographs_ctf.star" --ctfWin -1 --CS 2.7 --HT 300 --AmpCnst 0.1 --XMAG 140000
--DStep 14 --Box 512 --ResMin 100 --ResMax 5 --dFMin 5000 --dFMax 50000 --FStep 500 --dAst 0
--ctffind3_exe storage/brno2/home/radims/EMsoft/relion-1.3/ctffind3/ctffind3.exe

echo "done"
```



Obr. 5.6: Průběh prvních pěti CTF zero vzhledem k aplikovanému defokusu při 300 kV a koeficientu sférické aberace $C_S = 3mm$. Pro defokus $-4 \mu m$ je první CTF zero (černá křivka) 28 Å (vlevo). Výsledek CTF korekce na jednom z mikrogramů. V levé polovině je vypočítaný model. V pravé části pak fourierova transformace mikrogramu (vpravo).

5.4 Sběr částic

Sběr velkého množství částic je závažným problémem v kryoEM. Pro studie těsně nad atomovým rozlišením je nutné nasbírat statisíce až miliony částic, což však není možné (vzhledem k časové náročnosti) provádět ručně. Poloautomatické způsoby generují set nalezených částic, který pak člověk prochází a odstraňuje falešně pozitivní výsledky. To je velká pomoc při požadavku deseti tisíc částic, pro více je ale i tento způsob velmi časově náročný. Automatické algoritmy je možné rozdělit do dvou skupin.

5.4.1 Algoritmy využívající templát

Algoritmy založené na kroskorelaci využívají referenci generovanou z referenční 3D struktury nebo setu zprůměrovaných ručně nasbíraných částic. Metody vycházejí z lineárního modelu formování obrazu, který říká, že výsledný zašuměný obraz je lineární kombinací šumu (s nulovou střední hodnotou, známým výkonovým spektrem) a původního obrazu. Za předpokladu bílého šumu se technika omezuje na korelační techniku signálové detekce.

Problémem je velké množství všech možných 2D projekcí známé 3D struktury, které prodlužuje výpočet. Další nevýhodou algoritmů využívajících templáty je velký

počet falešně pozitivních výsledků. Každý objekt se zhruba stejnou velikostí a průměrnou intenzitou jakou má templát bude mít vysoký korelační koeficient. Proto je nutné provést po automatickém vyhledávání částic další zpracování, které je odstraní.[21]

5.4.2 Algoritmy hledající charakteristické vlastnosti

Na rozdíl od algoritmů využívajících templáty, které zpracovávají mnoho obrazových pixelů, používají metody, které jsou založené na rozpoznávání charakteristických vlastností objektu, jen pár jeho charakteristických vlastností (geometrické vlastnosti – pozice rohů, rovných úseků, kontur, statistické vlastnosti – momenty). Výhodou je jejich invariance ke změně velikosti, rotaci nebo změně osvětlení. Nevýhodou je pak obtížnost stanovení vlastností hledané struktury v málo kontrastních obrazech.[21]

5.4.3 Aplikovaný postup

Vzhledem k velkému počtu zpracovávaných mikrogramů byla hned ze začátku zavrhnuta varianta ručního sběru částic. Z důvodu velmi nízkého kontrastu zkoumaného komplexu EcoR124I byly zavrhnuty i metody využívající charakteristické vlastnosti hledaných objektů.

Vyhledávání částic v mikrogramech bylo provedeno pomocí prahování kroskorelační mapy spočítané z prohledávaného mikrogramu a jedné ze známých projekcí zjištěných elektronovou tomografií (rovnice 5.9).

$$c(x', y') = \sigma \sum_x \sum_y f(x, y) g(x + x', y + y') \quad (5.9)$$

Kde $f(x, y)$ je obraz, $g(x, y)$ reference. [22]. Podrobný popis algoritmu implementovaného do programu RELION je v článku Sjors H. W. Scheres: Semi-automated selection of cryo-EM particles in RELION-1.3 [23].

Odkazy na jednotlivé referenční projekce se zaznamenávají do souboru refs_autopick.star, který má tvar:

```
data_

loop_
_rlnImageName #1
000001@/storage/brno2/home/radims/relion13_first/EcoR124I/reference/Sel40_avg_sc_project.spi
000002@/storage/brno2/home/radims/relion13_first/EcoR124I/reference/Sel40_avg_sc_project.spi
000003@/storage/brno2/home/radims/relion13_first/EcoR124I/reference/Sel40_avg_sc_project.spi
000004@/storage/brno2/home/radims/relion13_first/EcoR124I/reference/Sel40_avg_sc_project.spi
. . .
```

Samotné vyhledávání částic v celém datasetu pak bylo realizováno skriptem uvedeným níže. Krok 1 představuje samotné prohledávání jednotlivých mikrogramů. Jako vhodný práh byla stanovena hodnota 0,05 , která poskytuje dostatečné množství nalezených částic a přitom negeneruje příliš mnoho falešně pozitivních výsledků. Vzhledem k obrovskému počtu možností, které je nutné propočítat (1413 mikrogramů, 9 projekcí otáčených po 5 stupních) byl výpočet prováděn na čtyřech uzlech po deseti procesorech.

```
#!/bin/tcshrelicion_autopick.pbs
#PBS -N autopick_all
#PBS -q normal
#PBS -l nodes=4:ppn=10:brno
#PBS -l mem=10gb
#PBS -j oe

# Environment
source ~/.tcshrc

# qsub -v DATADIR='pwd' XXXscriptXXX
#cd $DATADIR

# Command: mpiexec --bynode -n 20 XXXcommandXXX

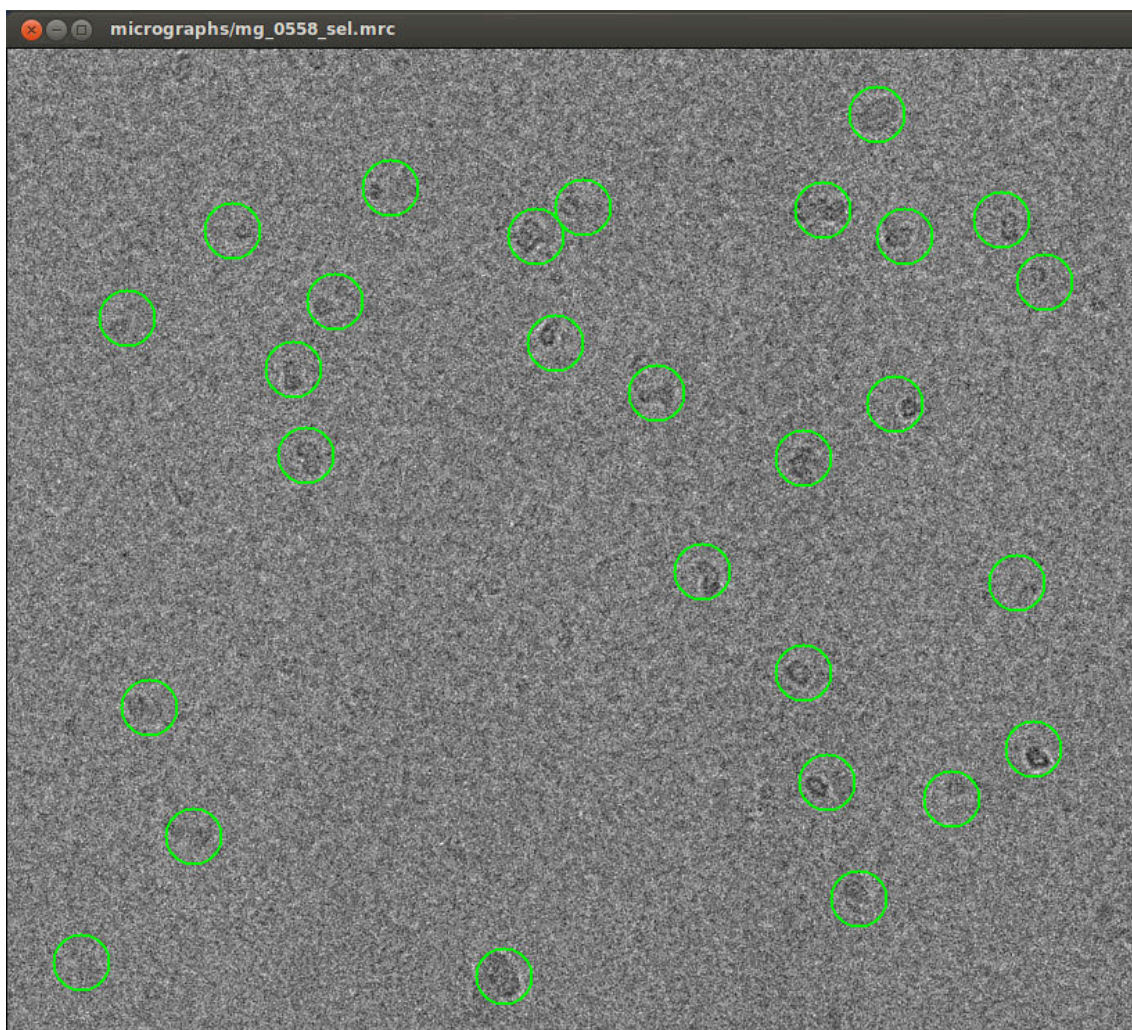
# Step1: find particles in all micrographs
mpiexec --bynode -n 40 'which relion_autopick_mpi' --i selected_micrographs.star --o autopick
--particle_diameter 200 --angpix 1 --ref refs_autopick.star --ang 5 --lowpass 10 --threshold 0.05
--min_distance 120

# Step2: extract automatically picked particles
#mpiexec -n 15 'which relion_preprocess_mpi' --o particles_autopick
--mic_star all_micrographs_ctf.star --coord_suffix _autopick.star --extract --extract_size 100
--norm --bg_radius 28 --white_dust -1 --black_dust -1 --invert_contrast

echo "done"
```

5.4.4 Extrakce částic

Nalezené částice je pro další zpracování nutné vyextrahovat z mikrogramů na základě souřadnic nalezených v předchozím kroku. Důležitým parametrem je stanovení vhodné velikosti boxu tak, aby zahrnoval celé částice (optimální mezera při okrajích je okolo 20%) i při jejich ne zcela přesném zaměření. Příliš velká box size však stěžuje zatřídění částic do jednotlivých tříd. Další hledisko při jejím výběru hraje čas potřebný ke spočítání 2D i 3D fourierovy transformace oblasti dané velikosti. Nevhodně zvolená box size tak prodlužuje výpočet až několikanásobně. Náročnost je možné spočítat pomocí příkazu `bfft -ver 7 -test 2,280,320` (funkce pro korekci ctf z balíku `bsoft` s parametry nejvyšší verbosity, testování 2D fourierovy transformace, interval). Při aplikaci na enzym EcoR124I s velikostí asi 200 Å je prohledávaný interval od 280 do 320 Å. Porovnání výpočetní náročnosti je v tabulce 5.1.



Obr. 5.7: Jeden z filtrovaných mikrografů s vyznačenými nalezenými částicemi (zele-
lené koužky).

Velikost boxu [pix]	2D fft [s]	Velikost boxu [pix]	3D fft [s]
288	0.00200	280	0.77296
300	0.00219	288	0.87606
280	0.00221	294	0.87606
...
311	0.02034	311	9.30659
317	0.02199	317	10.69155
283	0.03702	283	15.47310

Tab. 5.1: Časová náročnost výpočtu fourierovy transformace v závislosti na zvolené
velikosti boxu. Ve 3D případě je čas nejhoršího 20 krát delší než nejlepšího.

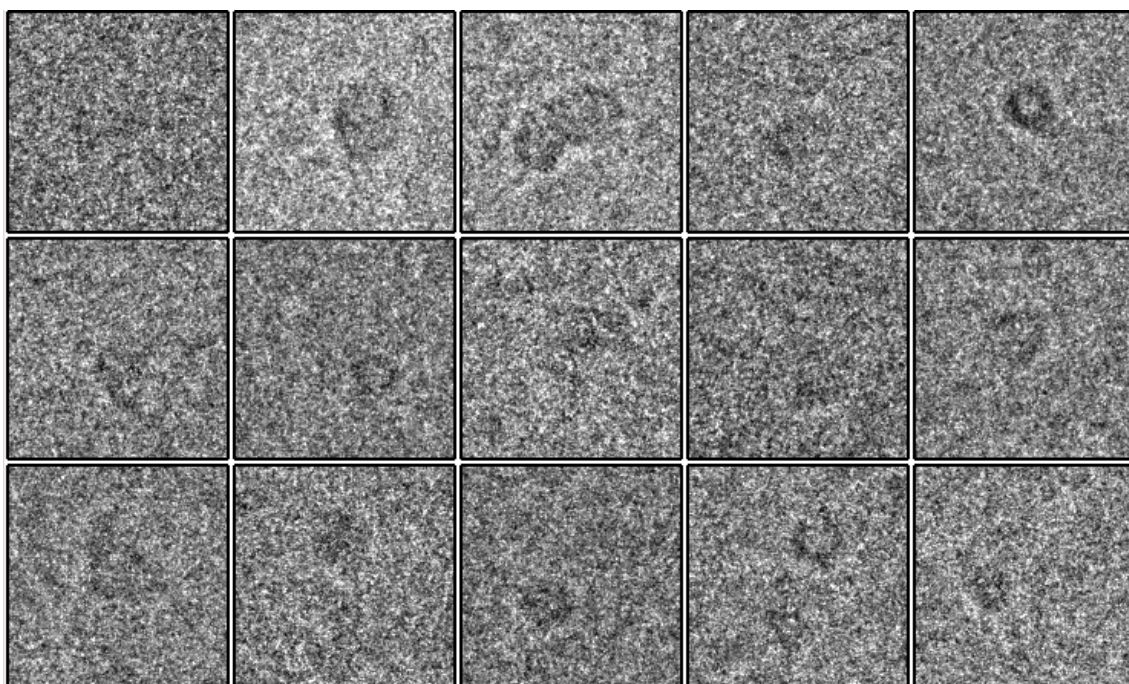
Samotná extrakce je pak prováděna příkazem označeným jako druhý krok v předchozím skriptu.

5.5 Zatřídění nalezených částic – závisle na referenci

RELION má implementovaný nástroj pro vytřídění falešně pozitivních výsledků. Pomocí příkazu

```
'which relion_particle_sort' --i particles_autopick.star --ref refs_autopick_resized.star  
--angpix 1 --particle_diameter 200
```

seřadí všechny nalezené částice podle jejich podobnosti s předkládanými referencemi (particle sorting). To umožňuje vybrat jen kvalitní částice a jejich identifikátory uložit do nového STAR souboru. To bohužel u zpracovávaného datasetu nestačilo a bylo nutné přikročit k ručnímu odstranění falešně pozitivních částic. Při selekci po seřazení bylo odstraněno 3200 částic, při ruční probírce dalších 7026. Dále je tedy zpracováváno 8974 částic, k jejichž dalšímu čištění dojde během roztřídění do jednotlivých dvojrozměrných tříd (2D classification).



Obr. 5.8: Nalezené a extrahované částice seřazené vzhledem k podobnosti s referencí.

5.6 Zatřídění nalezených částic - nezávisle na referenci

Je iterativní proces rozřazující předložené částice do předem stanoveného počtu tříd. Ve zpracovávaném datasetu bylo použito 80 tříd. Po prvním iniciálním rozřazení se uloží odkazy na částice spadající do kvalitních tříd do nového star souboru, který se tak stává novým vstupem pro další krok iterace. Proces je prováděn skriptem

```
#!/bin/tcsh
#PBS -N 2Dclass_all
#PBS -q normal
#PBS -l nodes=2:ppn=8:brno
#PBS -l scratch=500gb
#PBS -l mem=32gb
#PBS -j oe

# Submit: qsub -d 'pwd' relion_AllParts_2Dclass.pbs

# Environment
source ~/.tcshrc

# Command: mpiexec --bynode -n 20 XXXcommandXXX

mpiexec --bynode -n 8 'which relion_refine_mpi' --o Class2D/autopick_sort
--i particles_autopick_sort.star --particle_diameter 200 --angpix 3 --ctf
--iter 25 --tau2_fudge 2 --K 80 --flatten_solvent --zero_mask --strict_highres_exp 15
--oversampling 1 --psi_step 10 --offset_range 5 --offset_step 2 --norm --scale --j 8
--memory_per_thread 4

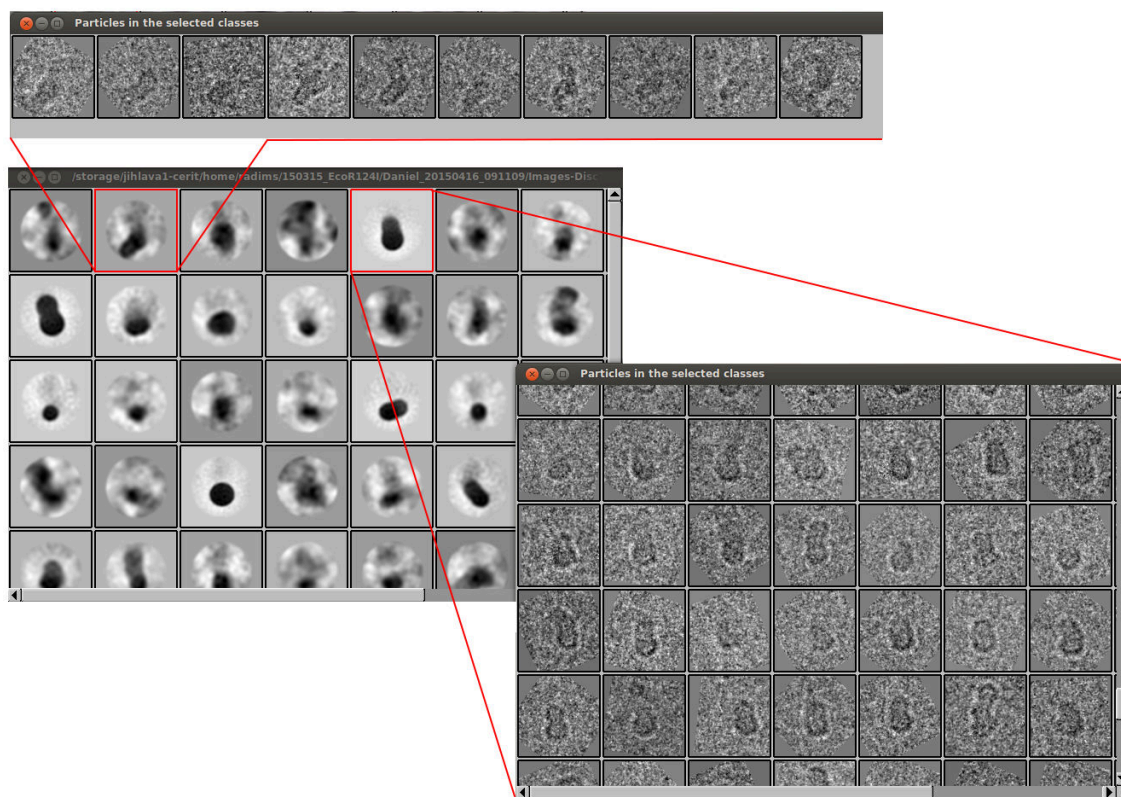
echo "done"
```

Po první iniciální iteraci s indexem 0 bylo odstraněno 906 částic tvořících 28 tříd z 80. To znamená, že asi 10 % dat utvořilo 35 % tříd. To je dáno zařazením jen malého množství odlišných částic do své vlastní třídy. Díky jejich vyřazení dochází k postupnému čištění dat a vylučování falešně pozitivních výsledků sběru částic. V dalších iteracích se počet částic snižoval jak je uvedeno v tab. 5.2.

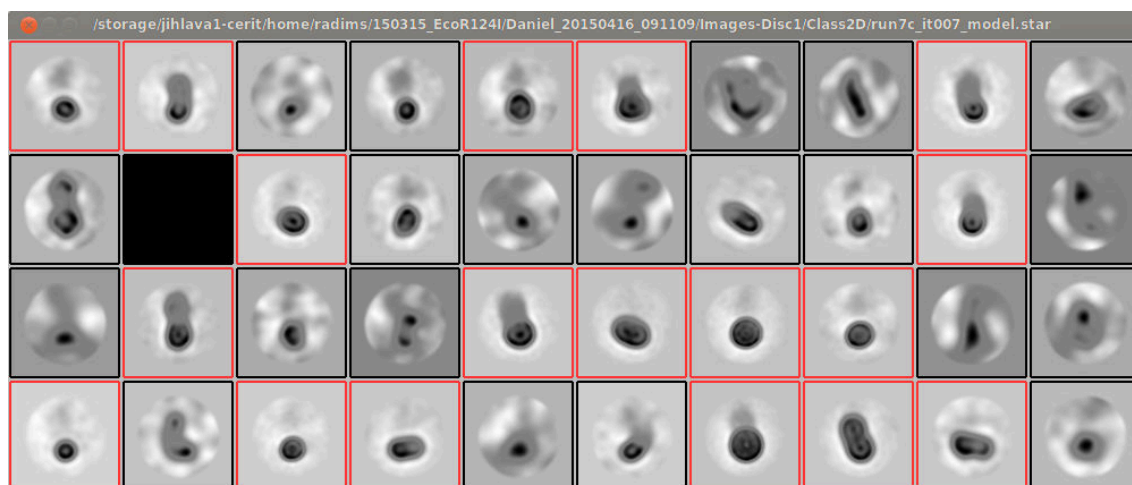
Běh	Počet zbývajících částic
1	8974
2	8068
3	7575
4	7410
5	7241
6	6940

Tab. 5.2: Počet zůstávajících částic po jednotlivých iteracích třídění do tříd.

Po sedmi bĕzích iterací (25 iterací v jednom) jsou již patrné jemnější struktury vypovídající o struktuře studovaného enzymu EcoR124I.5.10



Obr. 5.9: Čištění datasetu pomocí odstraňování malých a špatně zarovnaných skupin. Příklad nekvalitních částic tvořících si v malém počtu svoji vlastní skupinu (nahore) a kvalitní skupiny vyznačující se dobrým zarovnáním – nízkým pozadím (vpravo dole).



Obr. 5.10: Výsledek sedmého běhu iterací již naznačuje vnitřní strukturu částic (červeně ohraničené třídy). Upraveno transformací LUT.

6 ZÁVĚR

Diplomová práce prověřila možnost zpracování dat z kryoEM relativně malých (<500 kDa) heterogenních komplexů s nízkou symetrií. V úvodních dvou kapitolách se práce věnuje obecným vlastnostem kryoelektronové mikroskopie, tvorbě obrazu a kontrastu v EM. Dále jsou popsány vlastnosti studovaných restričních endonukleáz typu I (EcoR124I), způsob přípravy a biochemická charakterizace vzorku EcoR124I použitého v této studii.

Práce dále popisuje teoretické základy a praktické aspekty EM tomografie, které bylo provedeno na zkoumaném komplexu pro získání prvotního 3D modelu. Ačkoliv 3D subtomogram komplexu EcoR124I byl získán při nízkém rozlišení (45 Å), výsledný subtomogram se skládal ze tří podjednotek – M2S a dvou R.

Z tohoto subtomogramu byl připraven model využitý pro automatické hledání částic v datech pro jednočásticovou analýzu. Před samotným kros-korelačním vyhledáváním částic byly mikrografy vyfiltrovány pomocí DWT s prahem NeightShrink-Sure což významně zvýšilo kontrast částic vůči pozadí a usnadnilo jejich hledání.

Těžiště práce je v popisu a provedení SPA v kapitole 4. Zde jsou postupně probírány jednotlivé kroky včetně dílčích výsledků zpracování.

Závěrem jsou 2D částice komplexu EcoR124I po několika iteračních krocích, které mohou být využity pro následné 3D rekonstrukce komplexu.

LITERATURA

- [1] WANG, Liguang a Fred J SIGWORTH. Cryo-EM and single particles. Physiology (Bethesda, Md.) [online]. 2006, roč. 21, s. 13–18. ISSN 1548-9213. Dostupné z: doi:10.1152/physiol.00045.2005
- [2] MILNE, Jacqueline L S, Mario J. BORGNIA, Alberto BARTESAGHI, Erin E H TRAN, Lesley a. EARL, David M. SCHAUDER, Jeffrey LENGYEL, Jason PIERSON, Ardan PATWARDHAN a Sriram SUBRAMANIAM. Cryo-electron microscopy - A primer for the non-microscopist. FEBS Journal [online]. 2013, roč. 280, č. 1, s. 28–45. ISSN 1742464X. Dostupné z: doi:10.1111/febs.12078
- [3] KARLÍK, Miroslav. Úvod do transmisní elektronové mikroskopie. Vyd. 1. Praha: České vysoké učení technické v Praze, 2011, 321 s. ISBN 978-80-01-04729-3
- [4] Titan Krios: Visualizing Life at the Molecular Level. Product Data - FEI Company. 2010, 4 s.
- [5] Tecnai Spirit TEM: Easy-to-use solutions for life sciences imaging. Product Data - FEI Company. 2009, 8 s.
- [6] An Introduction to electron microscopy. FEI. 2010, 36 s., ISBN 978-0-578-06276-1 Elektronická verze dostupná na: <http://www.fei.com/introduction-to-electron-microscopy/>
- [7] KOLAROVA. 53. Restrikční nukleázy - význam v genovém inženýrství [online]. 2011, (cit. 5.12.2014) Dostupné na: <http://biologie-v-kostce.blogspot.cz/2011/05/58-restrikcni-nukleazy-vyznam-v-genovem.html>
- [8] KENNAWAY, Christopher K., James E. TAYLOR, Chun Feng SONG, Wojciech POTRZEBOWSKI, William NICHOLSON, John H. WHITE, Anna SWIDERSKA, Agnieszka OBARSKA-KOSINSKA, Philip CALLOW, Laurie P. COOPER, Gareth a. ROBERTS, Jean Baptiste ARTERO, Janusz M. BUJNICKI, John TRINICK, G. Geoff KNEALE a David T F DRYDEN. Structure and operation of the DNA-translocating type I DNA restriction enzymes. Genes and Development [online]. 2012, roč. 26, č. 1, s. 92–104. ISSN 08909369. Dostupné z: doi:10.1101/gad.179085.111
- [9] LAPKOUSKI, Mikalai, Santosh PANJIKAR, Pavel JANSČAK, Ivana Kuta SMATANOVA, Jannette CAREY, Rüdiger ETTRICH a Eva CSEFALVAY. Structure of the motor subunit of type I restriction-modification complex

- EcoR124I. Nature structural & molecular biology [online]. 2009, roč. 16, č. 1, s. 94–95. ISSN 1545-9993. Dostupné z: doi:10.1038/nsmb.1523
- [10] LOENEN, Wil a M, David T F DRYDEN, Elisabeth a. RALEIGH a Geoffrey G. WILSON. Type I restriction enzymes and their relatives. Nucleic Acids Research [online]. 2014, roč. 42, č. 1, s. 20–44. ISSN 03051048. Dostupné z: doi:10.1093/nar/gkt847
- [11] JANSČAK, Pavel, D. T F DRYDEN a Keith FIRMAN. Analysis of the subunit assembly of the type IC restriction-modification enzyme EcoR124I. Nucleic Acids Research [online]. 1998, roč. 26, č. 19, s. 4439–4445. ISSN 03051048. Dostupné z: doi:10.1093/nar/26.19.4439
- [12] CSEFALVAY, Ladislav, Daniel NEMEČEK, Eva CSEFALVAY. Purification and Initial Imaging of the type I DNA Restriction Modification Enzyme EcoR124I by Electron Microscopy, [interní materiály CEITEC]
- [13] ORLOVA, E. V. a H. R. SAIBIL. Structural analysis of macromolecular assemblies by electron microscopy. Chemical Reviews [online]. 2011, roč. 111, č. 12, s. 7710–7748. ISSN 00092665. Dostupné z: doi:10.1021/cr100353t
- [14] University of Bern, [online], 2015, [cit. 5.12.2014] Dostupné na: <http://www.ana.unibe.ch/exmo/research.html>
- [15] HEZMANN, Bernard. Bsoft [online]. 2015 [cit. 2015-05-19], Dostupné na: <http://lsbr.niams.nih.gov/bsoft/#> (poslední aktualizace 23.2.2015)
- [16] Electron tomography: methods for three-dimensional visualization of structures in the cell. 2nd ed. Editor Joachim Frank. New York: Springer, c2006, xiv, 455 s. ISBN 9781441921727
- [17] JIN, Liang, Anna Clare MILAZZO, Stuart KLEINFELDER, Shengdong LI, Philippe LEBLANC, Fred DUTTWEILER, James C. BOUWER, Steven T. PELTIER, Mark H. ELLISMAN a Nguyen Huu XUONG. Applications of direct detection device in transmission electron microscopy. Journal of Structural Biology [online]. 2008, roč. 161, č. 3, s. 352–358. ISSN 10478477. Dostupné z: doi:10.1016/j.jsb.2007.10.007
- [18] Falcon II Direct Electron Detector, FEI [online], 2015, [cit. 15.5.2015] Dostupné na: <http://http://www.fei.com/accessories/falcon-II/>
- [19] ANISIMOVA, Elena, Jan BEDNÁŘ, Petr PÁTA, Elena ANISIMOVA, Jan BEDNÁŘ a Petr PÁTA. Zpracování obrazu pomocí vlnkové transformace Image processing using the wavelet transform. 2013

- [20] NEELIMA, M a Mahaboob PASHA. Wavelet Transform Based On Image Denoising Using Thresholding Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*. 2014, roč. 3, č. 9, s. 7906–7908. ISSN 2278-1021
- [21] ZHU, Y. Automatic particle selection: results of a comparative study. *Journal of Structural Biology* [online]. 2004, roč. 145, č. 1-2, s. 3–14 [vid. 14. duben 2015]. ISSN 10478477. Dostupné z: doi:10.1016/j.jsb.2003.09.033
- [22] NICHOLSON, W. V. a R. MALLADI. Correlation-based methods of automatic particle detection in electron microscopy images with smoothing by anisotropic diffusion. *Journal of Microscopy* [online]. 2004, roč. 213, č. 2, s. 119–128. ISSN 00222720. Dostupné z: doi:10.1111/j.1365-2818.2004.01286.x
- [23] SCHERES, Sjors H.W. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology* [online]. 2015, roč. 189, č. 2, s. 114–122. ISSN 10478477. Dostupné z: doi:10.1016/j.jsb.2014.11.010
- [24] FITZPATRICK, Anthony W P, Ulrich J. LORENZ, Giovanni M. VANACORE a Ahmed H. ZEWAIL. 4D cryo-electron microscopy of proteins. *Journal of the American Chemical Society* [online]. 2013, roč. 135, č. 51, s. 19123–19126. ISSN 00027863. Dostupné z: doi:10.1021/ja4115055
- [25] Cryo-EM. 1st ed. Editor Grant J Jensen. Amsterdam: Elsevier, c2010, xlx, 410 s., [10] s. obr. příl. *Methods in enzymology*, vol. 481. ISBN 978-0-12-374906-2
- [26] SCHERES, Sjors H.W. Single-particle processing in RELION-1.3. MRC Laboratory of Molecular Biology [online]. 2014, 38 s. Dostupné z: http://www2.mrc-lmb.cam.ac.uk/groups/scheres/relion13_tutorial.pdf

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

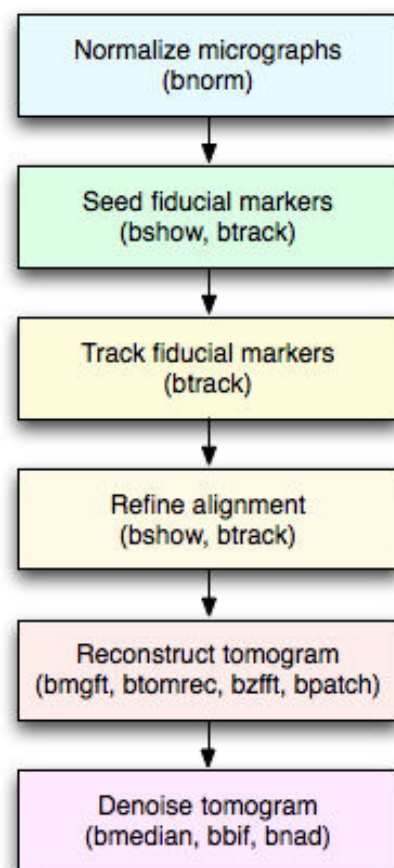
CCD	Charge-coupled device – zařízení s vázanými náboji
CMOS	Complementary Metal Oxide Semiconductor – doplňující se kov-oxid-polovodič
cryoEM	Cryo-Electron Microscopy – kryoelektronová mikroskopie
CTF	Contrast Transfer Function – transformační funkce kontrastu
DDD	Direct Detection Device – zařízení pro přímou detekci
DQE	Detector Quantum Efficiency – kvantová účinnost detektoru
DWT	Dyadic Wavelet Transform – dyadická vlnková transformace
FWHM	Full Width at Half Maximum – plná šířka v půlce maxima
HGT	Horizontal Gen Transfer – horizontální genový transfer
MSE	Mean Square Error – střední kvadratická chyba
NAD	Non linear Anisotropic Diffusion – nelineární anisotropická difuze
PSF	Point Transfer Function – bodová transformační funkce
RE	Restriction Endonuclease – restriční endonukleáza
RELION	REgularised LIkelihood Optimisation
RM	Restriction / Modification – restriční / modifikační
SNR	Signal to Noise Ratio – poměr signál/šum
SPA	Single Particle Analysis – jednočásticová analýza
STEM	Scanning Transmission Electron Microscope – skenovací transmisní elektronový mikroskop
SWT	Stationary Wavelet Transform – stacionární vlnková transformace
TEM	Transmission Electron Microscope – transmisní elektronový mikroskop
TRDs	Target Recognition Domains – cílové rozpoznávací domény

SEZNAM PŘÍLOH

A Pracovní postupy	57
A.1 Tomografie	57
A.2 Jednočásticová analýza	58
B Software	59
B.1 Bsoft	59
B.2 RELION	60
C Konferenční abstrakt	61
C.1 CryoEM reconstruction of small nucleoprotein complexes: Application to the EcoR124I restriction-modification enzyme	61

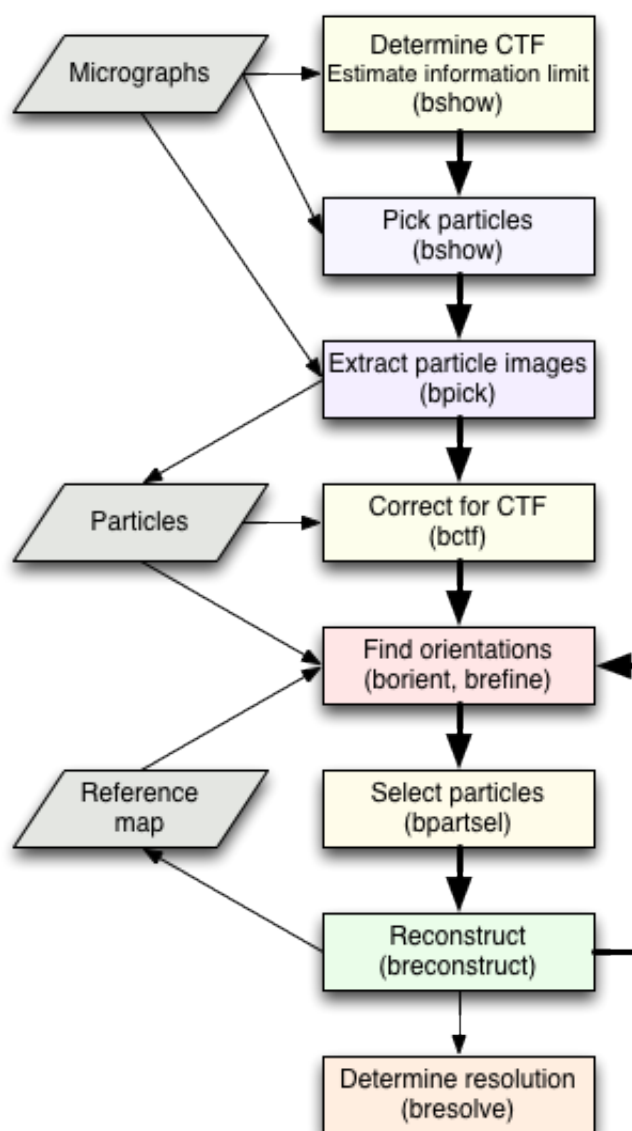
A PRACOVNÍ POSTUPY

A.1 Tomografie



Obr. A.1: Blokové schéma při zpracování tomografických dat v programovém balíku Bsoft včetně používaných příkazů.[15]

A.2 Jednočásticová analýza



Obr. A.2: Blokové schéma jednočásticová analýzy v programovém balíku Bsoft včetně využívaných příkazů.[15]

B SOFTWARE

V dnešní době je dostupná obrovská plejáda softwarových nástrojů pro zpracování dat z kryoEM. Jejich aktualizovaný seznam je dostupný na webových stránkách: http://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy

Základní charakteristiky využití softwarových nástrojů:

B.1 Bsoft

Webové stránky: <http://bsoft.ws>

Současná verze: 1.8.8

Kontakt: Bernard_Heymann@nih.gov

Bsoft je programový balík a platforma pro vývoj softwaru pro zpracování obrazu na molekulární úrovni ve strukturní biologii. K problémům ve strukturní biologii je přistupováno ve velice modulárním designu dovolujícím rychlý vývoj nových algoritmů bez zátěže problémů spojených s I / O soubory. To poskytuje jednoduché a přístupné prostředí, ze kterého lze čerpat pro jiné softwarové balíčky. Balík obsahuje několik kompletních pracovních postupů např. pro jednočásticovou analýzu a tomografii, které umožňují výměnu parametrů a tak možnost distribuovaného zpracování na rozdílných počítačových klastrech. Dostupné jsou též nástroje pro modelování struktur jako jsou atomové koordináty nebo modely ve velkém měřítku pro interpretaci velkých molekulárních komplexů jako jsou viry nebo sub-celulární organely. Podporuje:

Operační systémy: Unix (Mac OS X, IRIX, Linux, AIX, Solaris, Tru64), VMS

Formáty obrazu: BioRad, Brix, CCP4, Digital Instruments, Digital Micrograph, DSN6, EM, Goodford, GRD, HKL, Imagic, IP, JPEG, MFF, Image Magick, MRC, PIC, PIF, PNG, PNM, SPE, Spider, Suprim, TIFF, XPLOR, RAW

Cena: Free/Open Source (Public domain)

Napsáno v: C++, Tcl/Tk

Primární publikace pro citování: Heymann, J.B. (2001). "Bsoft: image and molecular processing in electron microscopy". *Journal of Structural Biology* 133 (2-3): 156-69. Další reference: Heymann, J.B.; Belnap, D.M. (2007). "Bsoft: Image processing and molecular modeling for electron microscopy". *Journal of Structural Biology* 157: 3-18. Heymann, J.B.; Cardone, G.; Winkler, D.C.; Steven, A.C. (2008). "Computational resources for cryo-electron tomography in Bsoft". *Journal of Structural Biology* 161: 232-242.

B.2 RELION

Webové stránky: <http://www2.mrc-lmb.cam.ac.uk/relion> Současná verze: 1.3

Kontakt: scheres@mrc-lmb.cam.ac.uk

Hlavní součástí programu RELION (REgularised Likelihood Optimisation) je jeho část sloužící k jemnému zarovnání mnoha 3D rekonstrukcí nebo 2D "class average" v kryo elektronové mikroskopii na základě empirického Bayesovského přístupu. Nástroj je vyvíjen skupinou Sjors Scherese z MRC Laboratory of Molecular Biology. RELION obsahuje nástroje díky nimž je schopný sám zjistit některé údaje přímo z dat a tak ulehčuje jejich zpracovávání. Například omezení plynoucí z hladkosti makromolekulárních komplexů - to omezuje rozsah spektra ve Fourierově prostoru. Během vývoje softwarového balíku se program změnil v obecný balík, ve kterém je v současné době možné provést většinu jednočásticových analýz.

Podporuje: Unix (Linux, Mac OS X, etc.)

Formáty obrazu: MRC, Spider

Cena: Free/Open Source

Primární publikace pro citování: Scheres SHW (2012). "A Bayesian View on Cryo-EM Structure Determination". J. Mol. Biol. 415 (2): 406-418.

doi:10.1016/j.jmb.2011.11.010. PMID 22100448.

C KONFERNČNÍ ABSTRAKT

C.1 CryoEM reconstruction of small nucleoprotein complexes: Application to the EcoR124I restriction-modification enzyme

R. Skoupý

*Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00
Brno, Czech Republic*

skoupy.radim@gmail.com

Cryo-electron microscopy has undergone significant advances in hardware development that push limits of 3-D structure determination of large macromolecular complexes of proteins and nucleic acids. New generation of electron microscopes (such as FEI Titan Krios) with parallel sample illumination, stable electro-magnetic lenses and new specimen compustage provided a platform for obtaining data extending to near-atomic resolution. Additionally, advent of direct detectors with 4-times better sensitivity and capability to record several images per second revolutionized the field of cryo-electron microscopy of biological specimen.

We used this new instrumentation to image the type I restriction-modification enzyme EcoR124I that is at the edge of standard cryoEM procedures and reconstructions: it is only 500 kDa large, has low symmetry and most likely is flexible in its open conformation [1]. The data obtained from Titan Krios with a direct detector provided data that can be processed and lead to high-resolution 3D reconstruction. However, we had to develop a new approach for processing of the aquired data to extract the individual EcoR124I particles from imaged micrographs. We used wavelet filter (i.e. dyadic and stationary wavelet transform with different treasholds) to enhance the contrast of the particles in the micrograph that facilitated identification of their positions and automated picking from thousands of collected micrographs. Next, we obtained the initial low-resolution model of the EcoR124I structure from cryo-electron tomography, a technique that allows 3-D reconstruction without any initial assumption about the 3-D structure. The average from 300 subtomogram particles was used to automatically locate particles in the filtered micrographs, using the algorithm implemented in the image processing package Relion [2]. Our next efforts are focused on refining the 3D structure to subnanometer resolution using the original dataset obtained from Titan Krios.

1. Ch. K. Kennaway, J. E. Taylor, Chun Feng Song et al: Structure and operation of the DNA-translocating type I DNA restriction enzymes, *Genes and Development.*, 26, (2012), 92-104. 2. Sjors H. W. Scheres: A Bayesian View on Cryo-EM Structure Determination, *Journal of Molecular Biology*, 415 (2011), 406-418