UNIVERZA V LJUBLJANI

FAKULTETA ZA FARMACIJO

SIMON TUREK

DIPLOMSKA NALOGA

VISOKOŠOLSKI STROKOVNI ŠTUDIJ LABORATORIJSKE BIOMEDICINE

Ljubljana, 2012

UNIVERZA V LJUBLJANI

FAKULTETA ZA FARMACIJO

SIMON TUREK

ISOLATION AND ATTEMPTED CRYSTALIZATION OF THE CATALYTIC SET DOMAINE OF THE MLL5 PROTEIN

IZOLACIJA IN POSKUS KRISTALIZACIJE KATALITIČNE DOMENE SET PROTEINA MLL5

DIPLOMSKA NALOGA

Ljubljana, 2012

All research work was carried out at the Centre for structural biochemistry (CBS - Centre de biochimie structurale) in Montpellier, France and at the University of Montpellier 1, Faculty of Pharmacy under the working supervision of Dr. Albane le Maire CR and the mentorship of Assist. Prof. Dr. Matjaž Jeras.

Acknowledgement

I would like to express my gratitude to Dr. Albane le Maire CR to introducing me to this project and helping me out through my work in Montpellier.

I would like to thank to entire Institute of CBS.

Velika zahvala, da je to delo dokončano, pa gre doc. dr. Matjažu Jerasu, ki me je med samim pisanjem diplomske naloge usmerjal, si v ta namen vzel veliko svojega časa in mi bil v veliko pomoč.

Zahvaljujem se družini in prijateljem za vse vzpodbudne besede tekom študija.

Statement

I declare that I have made this thesis/work independently and under the working supervision of Dr. Albane le Maire CR and the metorship of Assist. Prof. Dr. Matjaž Jeras.

Ljubljana, junij 2012 Simon Turek

Predsednik diplomske komisije: izr. prof. dr. Vojko Kmetec

Član diplomske kimisije: asist. Nejc Horvat

TABEL OF CONTECTS

ABSTRACT

Proteins are biochemical compounds usually consisting of one or more polypeptides typically folded into a globular or fibrous structure. Each polypeptide is a single linear polymer chain of amino acids interconnected via peptide bonds formed between terminal carboxyl and amino groups of adjacent amino acid residues. The amino acid sequence of each protein is defined by the coding sequence of a particular gene, present in a genome.

The functionality of a protein is usually linked to its folding, referred to as the structure-function relationship. In structural biology, several different methods are used for determination of protein structures. The goal of our work was to determine the 3D structure of the catalytic SET domain of the epigenetic enzyme MLL5**.**

Mixed-lineage leukemia protein 5 (MLL5) is a histone-lysine N-methyltransferase that specifically mono- and dimethylates 'Lys-4' of histone H3 (H3K4me1 and H3K4me2) thereby providing a specific tag for epigenetic transcriptional activation. The MLL5 is a key regulator of hematopoiesis, involved in terminal myeloid differentiation and in the regulation of hematopoietic stem cell (HSC) self-renewal. It also plays an essential role in retinoic-acid-induced granulopoiesis and acts as an important cell cycle regulator (1, 2, 3, 4). Namely, MLL5 is required for suppression of inappropriate expression of S-phase-promoting genes and to maintain the expression of determination genes in quiescent cells. Its overexpression inhibits cell cycle progression, while its knockdown induces cell cycle arrest at both, the G1 and G2/M phases (2). It can also be involved in methylation of DNA in the process of cancer formation. A new research field dealing with development of epigenetic drugs is quickly evolving. It seems that such drugs might become important constituents of combined anticancer therapies.

Our goal was to express the MLL5 protein in cultivated cells in the presence of Zn^{2+}. Namely the regulatory SET domain of the enzyme contains a region rich in cysteins which bind Zn^{2+} and this binding stabilizes its conformation. The SET domain alone is encoded by the aminoacid sequence from 328 - 451 bp (124 bp) within the coding region of the MLL5 gene.

We have successfully isolated and purified a small part of the MLL5 protein, corresponding to its regulatory SET domain (aminoacid sequence 323-473 bp) and tried to crystallize it in order to define its 3D structure by x-ray-crystallography. However in several attempts, we were only able to obtain amorphous crystals. Unfortunately this kind of material

is not good enough for further analyses. Although we were not successful in obtaining suitable crystals that would allow us to resolve the yet unknown 3D structure of the SET domain, our results represent a solid basis for further experiments.

RAZŠIRJENI POVZETEK

UVOD

Beljakovine so biokemijske spojine, zgrajene iz enega ali več polipeptidov. Običajno obstajajo v globularni ali nitasti obliki, ki jim zagotavlja biološko aktivnost oziroma funkcionalnost. Vsak polipeptid je polimer, sestavljen iz linearne verige aminokislin, ki so med seboj povezane s peptidnimi vezmi med amino in karboksilnimi skupinami sosednjih aminokislinskih ostankov. Zaporedje aminokislin v proteinu je opredeljeno z nukleotidnim zaporedjem kodirajočega področja ustreznega gena, ki je del celotnega genoma.

Protein MLL5 (protein mieloidne/limfoidne ali mešane celične linije levkemije 5) je histonska-lizin-N-metiltransferaza, ki specifično mono- ali dimetilira "Lys-4" histona H3 (H3K4me1 in H3K4me2), kar predstavlja posebno oznako za aktivacijo epigenetske transkripcije (1, 2, 3, 4). Protein MLL5 je homolog proteina trithorax (Trx) Drosophila, vsebuje en N-terminalni cinkov prstan vrste PHD (highly conserved putative plant homeodomain; močno ohranjena domnevna rastlinska homodomena, op.p.) in osrednjo domeno SET, kodira pa ga gen na kromosomskem odseku 7q22 (1) in je kandidat za supresorski gen tumorjev, saj je pogosto opuščen oziroma izbrisan pri akutnih mieloičnih levkemijah (5). Encim MLL5 je pomemben regulator celičnega cikla, saj zavira nepravilno izražanje genov, ki spodbujajo S-fazo ter vzdržuje ustrezno izražanje determinacijskih genov v mirujočih celicah. Čezmerna količina MLL5 zavira potek celičnega cikla, popolna odsotnost pa povzroči njegovo ustavitev, in sicer tako v fazi G1 kot G2/M (2, 3, 4). Dokazali so tudi, da lahko povzroči nastanek levkemij tako, da vpliva na preureditve in spreminjanje določenih genov v različnih kromosomih. Je ključni regulator hematopoeze, saj je vključen v terminalni del mieloične diferenciacije in v mehanizem samoobnavljanja krvotvornih matičnih celic (KMC). Igra tudi pomembno vlogo pri granulopoezi, ki jo izzove retinojska kislina, kjer deluje kot aktivator RAR alfa (RARA) na promotorje tarčnih genov (4). Encim MLL5 je ubikvitaren in se izraža tako v fetalnih tkivih kot tkivih odraslih oseb, v največjem obsegu pa se nahaja v fetalnih timusu in ledvicah ter v krvotvornem tkivu, jejunumu in malih možganih odraslih (5).

NAMEN

Namen našega dela je bil najprej izolirati in očistiti osrednjo katalitsko domeno SET proteina MLL-5. Nato pa smo načrtovali pripravo njenih čistih, dovolj velikih in urejenih kristalov, ki bi nam omogočili meritve difrakcije po obsevanju z žarki X, s čimer bi lahko na koncu določili do sedaj še neznano tridimenzionalno (3D) strukturo SET domene. Postopek kristalizacija beljakovine pa je sam po sebi zelo zahteven, med drugim tudi zaradi krhke narave tovrstnih kristalov. Določitev 3D domene SET proteina MLL5 bi doprinesla k boljšemu razumevanju delovanja tega proteina in potencialno imelo znaten pomen na nadaljnji razvoj zdravil. Kot so pokazale študije, ima MLL5, preko izvajanja metilacije DNA in histonov, tako v premalo kot tudi v preveč izraženem stanju, velik vpliv na nastanek tumorjev. Omenjeni encim predstavlja torej potencialno tarčo za nova epigenetska zdravila. Ta naj bi bila za razliko od genskih precej bolj učinkovita, vendar pa je do njihove uporabe še vedno precej daleč. Epigenetska zdravila ne vplivajo zgolj na posamezen gen oziroma na njegovo vklapljanje in izklapljanje, pač pa delujejo tako, da popravljajo napake v okvarjenem splošnem molekularnem regulacijskem mehanizmu, s tem pa seveda vplivajo na delovanje stotine genov, ki tak mehanizem uporabljajo. Današnja genska zdravila imajo vrsto slabosti. Med njimi je zagotovo ena od najpomembnejših vezana na premalo učinkovite in natančne dostavne sisteme, kar ima za posledico slabo ciljanje, toksičnost, kratkotrajno delovanje in pojav neželenih imunskih reakcij.

METODE

Začetek eksperimenta se je začnel s pripravo konstrukta, ki vsebuje 6-His »tag« (slika 6), nato sledi analiza izražanja konstruktov v B834 in Rosetta celicah. V samem procesu sta bila analizirana dva različna načina pospešitve izražanja proteina; IPTG in OnEx. Delo do te točke je bilo opravljeno s strani laboratorija OPPF iz Oxforda. Na podlagi njihovih rezultatov smo se odločili za delo s proteinom MLL5-3J (tabela I, stran 26). Sam protein MLL5-3J meri na začetku izolacijskega postopka 75 kDa. Z različnimi izolacijskimi postopki smo poskusili izolirati katalitično SET domeno, aminokislinsko zaporedje: 323 – 473, ki vsebuje veliko cisteina. Končna velikost izolirane SET domene naj bi bila 17 kDa.

Eksperiment je bil razdeljen v sledeče faze:

1. Priprava celic:
    - ✓ priprava petrijevk za celična kulture;
    - ✓ vnos konstrukta v seve E.coli;
    - ✓ priprava glicerolove raztopine za hrambo celic;
    - ✓ priprava pred-kulture;
    - ✓ priprava kulture;
    - ✓ »žetev« (op.p. harvesting) proteina iz gostiteljskih celic.

2. Čiščenje/izolacija proteina:
    - ✓ obdelava gostiteljskih celic z ultrazvokom;
    - ✓ priprava Ni kolone;
    - ✓ kalibracija aparata za kromatografijo;
    - ✓ afinitetna kromatografija (HisTrap ali Ni-kolona);
    - ✓ priprava kolone GST;
    - ✓ cepitev fuzijskega proteina z encimom P3C;
    - ✓ kalibracija kolone za gelsko filtracijo;
    - ✓ gelska filtracija (GF);
    - ✓ cepitev GSH Ni in/ali GST kolone:
    - ✓ koncentriranje izoliranega proteina s pomočjo »amicons« celic pod pritiskom in v epruvetah s pomočjo centrifuge;
    - ✓ karakterizacija proteina (ε - molarni ekstinkcijski koeficient in MW – molekularna masa);
    - ✓ ločitev proteina s SDS-page elektroforezo;
    - ✓ meritev koncentracije proteina (metoda po Bradford-u in na nano-drop spektrofotometer pri 280 nm);
    - ✓ cirkularni dihroizem (CD);
    - ✓ dinamično sipanje svetlobe (op.p. - dynamic light scattering – DLS).

3. Kristalizacija izoliranega/očiščenega proteina.
4. Optimizacija kristalizacijskih pufrov in samega kristalizacijskega procesa, z namenom pridobitve kristalov, ki bi jih lahko uporabili za nadaljnje analize.

REZULTATI IN RAZPRAVA

V okviru našega dela smo iz gojenih celic, ki so proizvajale rekombinantni protein MLL5, uspeli izolirati in očistiti njegovo domeno SET. V treh poskusih njene kristalizacije pa smo bili razmeroma neuspešni, saj smo pridobili le zelo majhne delce amorfnih oblik, ki pa žal niso bili ustrezni za nadaljnjo analizo. Ne glede na to, pa so bili rezultati našega zadnjega poskusa zelo obetavni in odpirajo možnosti za nadaljnje raziskave.

Prvi poizkus: MLL5-3J (ROSETTA, SEV E. coli), 2 x 700 mL LB MEDIJA *(Rast celične kulture v LB, indukcija IPTG)*

V prvem poskusu smo uporabili zamrznjena vzorca, ki sta vsebovala rekombinantni protein MLL5-3J, izražen v celicah »Rosetta« (sev E. coli) induciran z IPTG. Vsaka krioviala je vsebovala celice iz 700 mL kulture v mediju LB. Odtajane celice smo združili in po njihovi obdelavi z ultrazvokom ter filtraciji vzorca je sledila ločba proteina na Ni koloni s kromatografskim postopkom HisTrap.

Na grafu, predstavljenem na sliki 12 je prikazan tipičen kromatogram, na katerem so vidne vse tri faze kromatografskega procesa: nanos (L); izpiranje (W) in elucija (E). Drugi vrh v grafu predstavlja naš eluirani protein (MLL5-3J).

Na naslednjem grafu (slika 13) lahko določimo področje za MLL5-3J specifičnega vrha. Protein se je, kot je označeno na grafu, izpral s kolone med frakcijama 21 in 32. Zato smo se odločili, da vse omenjene eluirane frakcije uporabimo za njegovo nadaljnjo izolacijo oziroma čiščenje. Skupen volumen spranega proteina je bil 12 mL, v njem pa je bilo 3,4 mg/mL MLL5-3J. Koncentracijo proteina smo določili s postopkom nanodrop (Ɛ=55.405, MW=45.582). Po izračunih sodeč smo uspeli izolirati 41 mg proteina, kar je bila očitno posledica izjemnega izražanja rekombinantnega vektorja, razloga za to pa nismo mogli ugotoviti. Kot je razvidno iz kromatograma pa smo polega specifičnega dobili tudi nespecifičen vrh (frakcije 5 do 10). Zato smo predvidevali, da sam postopek ločbe ni bil najbolj učinkovit in se odločili, da opravimo še elektroforezo SDS-page. V ta namen smo pripravili naslednje vzorce oziroma eluirane frakcije (op.p. flow throughs): 5, 10, 23, 25, 28 in 31. S pomočjo slike gela po elektroforezni ločbi smo določili velkost in položaj našega proteina MLL5-3J (Slika 14). Protein je imel velikost okoli 70 kDa. Ugotovili smo, da so tudi frakcije od 5 do 10 vsebovale MLL5-3J, s čimer smo potrdili neučinkovitost ločbe. V naslednjem koraku smo uporabili postopek ločbe na koloni GST. Po nanosu vzorca in izpiranju nevezanih proteinov smo dodali encim proteazo P3C. Končna velikost eluiranega

proteina je bila 34 kDa. Med izpiranjem kolone s 24 mL elucijskega pufra smo izpirke, z volumni po 6 mL, zbrali v 4 različnih epruvetah. Po določitvi koncentracije proteina v vsaki epruveti, smo se odločili za nadaljnji postopek ločbe oziroma čiščenja, in sicer tokrat z gelsko filtracijo (GF), pri čemer smo uporabili le prvi dve epruveti, ki sta vsebovali največji količini MLL5-3J.

Rezultati, prikazani na slikah 15 in 16 so bili zelo vzpodbudni. Uspelo nam je dobiti simetričnen vrh, ki pa je bil nekako prevelik. Odločili smo se, da bomo združili izprane frakcije med 32 in 42. Ta vzorec smo nato analizirali s fotokorelacijsko spektroskopijo (DLS), in ugotovili, da ima tarčni protein velikost 34 kDa. To pa je bil rezultat, ki ga nismo pričakovali, saj bi morala biti velikost MLL5-3J po GF 17.8 kDa. Predpostavili smo, da smo dobili dimere. Sicer pa je bila koncentracija proteina v izpranem vzorcu 10 mg/mL, kar pomeni, da smo skupaj pridobili 6.18 mg MLL5-3J. Za potrditev rezultatov smo nato izvedli še elektroforezo SDS-page.

Rezultati elektroforeze (slika 17) prikazujejo učinkovitost različnih načinov ločbe proteina MLL5-3J, in sicer: FT; Ni-pool; eluatov El1, El2 in El3 po GF; eluatov El1 (GST1) in El2 (GST2) po GST in končne GF pred kristalizacijo. Ugotovili smo, da vsebujeta obe izprani frakciji El1 in El2, pridobljeni z GF, veliko proteina, in sicer 0.27 mg/mL, kar smo določili s postopkom nanodrop (Ɛ=12.295, MW=17.8 kDa). V vseh 24 mL vzorca smo torej imeli skupaj 6,48 mg proteina MLL5-3J. Njegova končna velikost pa je bila 17.8 kDa. Rezultati analize s krožnim dikroizmom CD so pokazali, da je večina proteina v beta konformaciji, kar je tudi pričakovano za domeno SET. Pred samim poskusom kristalizacije smo vzorec še koncentrirali s centrifugiranjem v filtracijski epruveti, in sicer do končnega volumna 600 µl. Koncentracija vzorca proteina je bila tako 10.3 mg MLL5-3J/600 µl. Žal pa je bil nadaljni postopek kristalizacije neuspešen, saj smo uspeli pridobiti zgolj drobne, amorfne delce, ki pa niso bili primerni za nadaljnje analize. Ugotovoli smo torej, da uporabljeni postopki izolacije niso bili dovolj učinkoviti.

Drugi poizkus: MLL5-3J (ROSETTA, SEV E. coli), 2 x 700 mL OnEx MEDIJA (*Rast celične kulture v OnEx*, *OnEx samoindukcija*)

V drugem poskusu smo uporabili zamrznjena vzorca, ki sta vsebovala rekombinantni protein MLL5-3J, izražen v celicah »Rosetta« (sev E. coli) in induciran s postopkom samoindukcije OnEx. Vsaka od odmrznjenih kriovial je vsebovala celice iz 700 mL kulture v mediju OnEx. Vsebini odtajanih kriovial smo združili ter po obdelavi celic z ultrazvokom in filtraciji vzorca, ki je vseboval naš tarčni protein, izvedli ločbo na Ni koloni s kromatografskim postopkom HisTrap.

Protein MLL5-3J smo izprali iz kolone s pomočjo gradienta imidazola. Kot je razvidno iz slike 18, vrh, ki ga je zaznal detektor ni bil simetričen. Odločili smo se, da združimo vse izprane frakcije med 7 in 30. Skupni volumen združenih eluatov je bil 48 mL, in sicer s koncentracijo 1.35 mg/mL MLL5-3J, ki je bila izmerjena s postopkom nanodrop (Ɛ=55.405, MW=45.582). Skupaj smo torej izolirali približno 65 mg tarčnega proteina. Tako kot v prvem poskusu smo izvedli še ločbo na koloni GST, s cepitvijo z encimom P3C. Za nadaljno GF smo uporabili prvi dve iz kolone GST izprani frakciji, pri čemer pa smo, v primerjavi s prvim poskusom, spremenili izpiralni pufer. Tokrat smo uporabili 100 mM Hepesa, 150 mM NaCl in 2 mM beta-mercaptoetanola, s pH=7.

Kromatogram na sliki 19 prikazuje vrh po gelski filtraciji, ki je simetričen, pravilne oblike in ni ne prevelik, ne premajhen. Združili smo izprane frakcije od 23 do 35 in tako dobili 12 mL vzorca, ki je vseboval 0.46 mg/mL MLL5-3J, kar smo izmerili s postopkom nanodrop (Ɛ=12.295, MW=17.8 kDa). Pred kristalizacijo smo preiskovani vzorec skoncentrirali do končnega volumna 1 mL, v katerem je bilo 10.8 mg MLL5-3J/mL. Žal pa tudi v tem primeru nismo bili uspešni pri izvedbi kristalizacije.

Tretji poskus: MLL5-3J (ROSETTA, SEV E. coli) 3 X 700 ML LB MEDIJA (*Rast celične kulture v LB, pred-kultura z dodatkom $ZnSO_4$, indukcija IPTG*)

V tretjem poskusu smo uporabili zamrznjene vzorce, ki so vsebovali rekombinantni protein MLL5-3J, izražen v celicah »Rosetta« (sev E. coli), induciran z IPTG. Celično pred-kulturo smo pripravlili z dodatkom $ZnSO_4$. Vsaka krioviala je vsebovala celice iz 700 mL kulture v mediju LB. Odtajane vsebine kriovial smo združili in po obdelavi celic z ultrazvokom vzorcu dodali DNAzo I, skupaj z njenim aktivatorjem $MgCl_2$. S tem smo razrezali prisotno DNA na manjše delce in tako preprečili potencialne motnje pri nadaljnjih analizah. Nanos vzorca na Ni-kolono je potekal s pomočjo peristaltične črpalke, s hitrostjo nanosa 0.75 mL/min. Spiranje kolone pa smo izvedli ročno, in sicer s 25 mL pufra PBS, s pretokom 5 mL/min. Za drugi nanos na kolono smo uporabili izpirke (flow through), ki smo jih zbrali po prvi ločbi vzorca na koloni, pri čemer je bila hitrost nanosa enaka kot v prvem primeru. Za spiranje kolone smo uporabili 50 mL pufra PBS. Spiranje smo nato ponovili še s 50 mL pufra, ki je vseboval 50 M Tris in 150 mM NaCl, s pH=7.5, in sicer zaradi priprave kolone za GST kromatografijo. Po dodatku encima P3C smo pričeli s spiranjem kolone. Meritve koncentracij proteina MLL5-3J v prvih dveh eluatih smo izvedli z metodo po Bradfordu. Skupni volumen obeh prvih izpirkov je bil 12 mL.

Nato smo izvedli še gelsko filtracijo (GF), ki pa ni bila tako učinkovita, kot je bilo pričakovali (slika 21). Specifični vrh je bil namreč previsok. Odločili smo se, da bomo združili vse frakcije med 8 in 43 ter tako dobili 70 mL, ki smo ga skoncentrirali s pomočjo celice amicon, pod zvišanim pritiskom. Še pred koncem pa smo postopek prekinili in vzorec ponovno razredčili, tokrat z disociacijskim pufrom, saj smo sumili, da vsebuje multimere tarčnega proteina. Nato smo ponovno izvedli GF, pri čemer pa smo za izpiranje uporabili disociacijski pufer. Kljub prizadevanjem rezultat GF tudi tokrat ni bil tak, kot smo ga pričakovali, a vendar je bila oblika vrha precej boljša kot prej (Slika 22). Po GF smo združili frakcije od 32 do 43 in izmerili koncentracijo MLL5-3J s postopkom na nanodrop (Ɛ=12.295, MW=17.8 kDa). Ta je bila 1,089 mg/mL, kar pomeni skupno 27,225 mg tarčnega proteina. Vzorec smo nato koncentrirali v koncentracijski epruveti s pomočjo centrifuge, in sicer na končni volumen 1,8 mL, s koncentracijo 10.4 mg MLL5-3J /mL (nanodrop; Ɛ=12.295, MW=17.8 kDa).

S fotokorelacijsko spektroskopijo (DLS) (Slika 23) smo potrdili, da je v našem vzorcu dejansko prisoten protein MLL5-3J s povprečno vrednostjo Z 7,08 d.nm in da ni prisotnih

multimerov. S postopkom cirkularnega dihroizma (CD) smo tudi potrdili, da je večina proteina v beta konformaciji, kar je pričakovano za domeno SET (slika 24). Tudi v zadnjem poskusu nam žal ni uspelo pridobiti kristalov, ki bi bili primerni za nadaljno analizo oziroma določanje tridimenzionalne strukture domene SET proteina MLL5.

OPTIMIZACIJA KRISTALIZACIJSKIH PUFROV

Rezultati preverjanja bistrosti kapljic med testiranjem različnih kristalizacijskih pufrov so prikazani v tabeli V. Tabela VI pa vsebuje podatke o indeksih proteinske disperzije (PDI), v primerih, kjer so bile kapljice bistre.

Legenda:

/ - nobene bistre kapljice;

Ø – bistra kapljica, ki vsebuje 10 mg/mL MLL5-3J;

• - bistra kapljica, ki vsebuje 10 mg/mL MLL5-3J z dodatkom SAM (S-adenozil metionina);

•• - bistra kapljica, ki vsebuje 25 mg/mL MLL5-3J;

All - vse kapljice so bistre.

V nadaljevanju smo analizirali le tiste kristalizacijske pufre, ki niso povzročili obarjanja proteina. Tiste med njimi, pri katerih so bile določene vrednosti PDI (indeks proteinske disperzije je pokazatelj namenjen primerjavi topnosti proteina v vodi) okoli 0,5 in ki so imeli pufrski vrh med 6 in 10, smo izbrali kot najboljše in najprimernejše za proces kristalizacije.

ZAKLJUČEK

Med izvajanjem celotnega procesa izražanja, indukcije, izolacije, čiščenja, koncentracije in poskusa kristalizacije proteina MLL5-3J, smo obravnavali številna vprašanja, ki so bila povod za spreminjanje in preizkušanje novih protokolov.

S tem namenom smo izvedli 3 različne poskuse, v okviru katerih smo testirali različne eksperimentalne pogoje in njihov vpliv na končni cilj, to je kristalizacijo tarčnega proteina. Tako smo na primer, pri pripravi optimalne produkcijske kulture Rosetta, seva E. coli za učunkovito »žetev« rekombinantnega proteina, preizkusili dva različna protokola, in sicer v drugem in tretjem poskusu. V tretjem poskusu smo kulturi dodali 100 μL $ZnSO_4$ (10 μM $ZnSO_4$), da bi povečali indukcijo proteina v celicah. Po obdelavi celic z ultrazvokom, smo v vzorec dodali DNA-zo I, skupaj z njenim aktivatorjem $MgCl_2$, ki razreže DNA v manjše delce, ki niso več moteči za nadalje postopke.

Poleg tega smo namesto pufra Tris uporabili Hepes in PBS. Vse do te eksperimentalne točke so bili postopki zelo učinkoviti, kot je razvidno iz rezultatov številnih analiz, ki smo jih naredili na posameznih stopnjah naših poskusov. Vendar pa smo bili žal v vseh treh eksperimentalnih poskusih neuspešni v zadnjem koraku, to je pri kristalizaciji tarčnega proteina MLL5-3J. Zato smo se na koncu odločili, da spremenimo pristop k problemu.

Poskušali smo najti kar najboljši pufer, ki bi nam omogočil učinkovito izolacijo in kristalizacijo proteina. Zato smo analizirali vrsto različnih pufrov in končali s seznamom tistih, ki bi naj po naših ocenah imeli dobre možnosti za dosego navedenih ciljev. Zavedati pa se moramo, da obstajajo tudi drugi pomembni dejavniki, ki bi jih morali obravnavati in preizkusiti, da bi optimizirali izolacijo in kristalizacijo naše ciljne beljakovine.

Dodatno pomembno vprašanje je izražanje rekombinantne beljakovine MLL5-3J v celični kulturi Rosette, sevu E. coli. Glede na naše rezultate lahko trdimo, da je bil ta del našega dela uspešen. Žal pa nismo analizirali stabilizacijskega učinka Zn^{2+} ionov na konformacijo naše beljakovine, kar bi seveda bilo zelo zanimivo.

Čeprav nismo uspeli pridobiti ustrezne oblike kristalov rekombinantne beljakovine MLL5-3J, s katerimi bi lahko izvedli nadaljnje strukturne analize katalitične domene SET tega epigenetskega encima, pa smo kljub temu prepričani, da smo naredili nekaj pomembnih prispevkov, ki bi lahko privedli do želenega rezultata v prihodnosti.

RÉSUMÉ

Les protéines sont des macromolécules biologiques composées d'une ou plusieurs polypeptides généralement pliées dans une structure globulaire ou fibreuse. Chaque polypeptide est une chaîne polymère linéaire d'acides aminés reliés entre eux par des liaisons peptidiques formés entre les groupes terminaux carboxyle et amino de résidus d'acides aminés adjacents. La séquence d'acides aminés de chaque protéine est définie par la séquence codante d'un gène particulier, qui est présent dans le génome.

La fonctionnalité d'une protéine est généralement liée à son pliage, dénommée la relation structure-fonction. En biologie structurale, plusieurs méthodes différentes sont utilisées pour la détermination des structures des protéines. Le but de notre travail était de déterminer la structure 3D du domaine catalytique SET d'enzyme MLL5 épigénétique.

Protéines leucémie mixte lignée 5 (MLL5) est une histone-lysine N-méthyltransférase spécifiquement mono-et dimethylates 'Lys-4' de l'histone H3 (H3K4me1 et H3K4me2), fournissant ainsi une balise spécifique pour l'activation transcriptionnelle épigénétique. Le MLL5 est un régulateur clé de l'hématopoïèse, impliqué dans la différenciation myéloïde terminale et dans la régulation de cellules souches hématopoïétiques (CSH) d'auto-renouvellement. Il joue également un rôle essentiel dans acide rétinoïque-induites granulopoïèse et agit comme un régulateur du cycle cellulaire importante (1, 2, 3, 4). Nommément MLL5 est nécessaire pour la suppression de l'expression inappropriée de phase S de promotion gènes et à maintenir l'expression de gènes dans des cellules de détermination de repos. Sa surexpression inhibe la progression du cycle cellulaire, tandis que son effet de choc induit un arrêt du cycle cellulaire à la fois, le G1 et G2 / M phases (2). Il peut également être impliqué dans la méthylation de l'ADN dans le processus de formation du cancer. Un nouveau domaine de la recherche portant sur le développement de médicaments épigénétiques évolue rapidement. Il semble que ces médicaments pourraient devenir importants constituants des combinés thérapies anticancéreuses.

Notre objectif était d'exprimer la protéine dans les cellules Rosetta cultivées en présence de Zn^{2+}. À savoir, le domaine SET réglementaire de l'enzyme contient une région riche en cystéines qui lient Zn^{2+} et cette liaison stabilise sa conformation. Le domaine SET seule est codé par la séquence aminoacide de 328 - 451 pb (124 pb) de la région codante du gène MLL5.

Nous avons réussi à isoler et purifier une petite partie de MLL5, avec la correspondance à sa domaine régulatrice SET (séquence nucleotide 323-473 bp) et essayé le

cristalliser pour définir sa structure 3D avec x-ray-cristallographie. Cependant, dans les plusieurs essaies, nous avons pu obtenir qu'amorphes cristaux. Malheureusement ce type de matériel ne suffit pas pour l'analyse additionnelle. Malgré que nous n'étions pas successives d'obtenir les cristaux approprie pour nous permettre de résoudre le structure 3D de domaine SET qui était encore inconnu, nos résultats représentent la base solide pour les recherches prochaines.

1 THE LIST OF ABBREVIATIONS

A

AA – aminoacid sequence

ACTB - beta-actin

AIRE – autoimune regulator

ASH2L - set1/Ash2 histone methyltransferase complex subunit ASH2 Ash2 (absent, small, or homeotic)

B

BP (bp) – base pair

C

CAM – chloramphenicol (2,2-dichloro-N-[1,3-dihydroxy-1-(4-nitrophenyl)propan-2-yl]acetamide)

CARB – carbenicillin(2S,5R,6R)-6-{[carboxy(phenyl)acetyl]amino}-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid)

CBS – centre de biochimie structurale

D

DLS - photo correlation spectroscopy

DTT – dithiothreitol ((2S,3S)-1,4-bis(sulfanyl)butane-2,3-diol)

E

EZH2 - histone-lysine N-methyltransferase

ε - molar extinction coefficient (M^{-1} cm^{-1})

F

FR - flow rate

G

GF - gel filtration

GSH – glutathione

GST - the glutathione S-transferase

H

HCFC1 - host cell factor

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

I

IPTG – isopropyl β-D-1-thiogalactopyraniside

K

kDa/kD - kilodaltons; measure of molecular weight or mass. One hydrogen atom has mass of 1 Da

KMC – krvotvorne matične celice

L

LB - lysogeny broth

M

mAU - a symbol for the milli-absorbance unit. An increase in absorbance of 1 mAU corresponds to a reduction in transmittance of about 0.2305%.

MLL5 - myeloid/lymphoid or mixed-lineage leukemia protein 5

MW - molecular weight in kilo Daltons

MWCO - molecular weight cut-off

N

NR - nuclear receptor

O

OGT - UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit

ONEX – overnight extension

P

PDI - the protein dispersibility index

P3C - 3C Protease,Cysteine Protease

PBS - phosphate Buffered Saline

PcG - polycomb-group

PEV - position-effect-variegation

PPP1CA - serine/threonine-protein phosphatase PP1-alpha catalytic subunit

PPP1CB - serine/threonine-protein phosphatase PP1-beta catalytic subunit

PPP1CC - serine/threonine-protein phosphatase PP1-gamma catalytic subunit

R

RARA - restriction enzyme

RPM – rounds per minute

S

SAM - S-Adenosyl methionin

STK38 - serine/threonine-protein kinase 38

T

TGS - tris-glycine-SDS

Tris - 2-Amino-2-hydroxymethyl-propane-1,3-diol

TRX/MLL - histone-lysine N-methyltransferase HRX

2 INTRODUCTION

All of the laboratory work has been performed as a part of the EU project Trainomics, at the CBS (Centre de Biochimie Structurale), University 1 of Montpellier, France, dedicated to the study of SET domains and methyltransferase activity of MLL5. The project includes several European laboratories.

2.1 THE MISSION OF TRAINOMICS

2.1.1 THE AIMS OF THE PROJECT

The basic goal of Trainomics is training of young investigators for the development of novel epigenetic drugs and nuclear receptor (NR) modulators by performing iterative cycles of drug design based on the known 3D protein structures, characterisation of molecular features of potential drug action and analysis of its global gene regulatory activities by applying "omics", including the novel LChIP-seq technologies. In parallel, the anti-cancer activities of novel modulators will also be studied in vitro, ex vivo (leukemic blasts and patient-matched primary tumour/normal cell cultures) and in vivo (animal models, for example zebrafish and mice). Big, small and medium pharmaceutical enterprises will be integrated into this drug development pipeline by providing their scientific contribution, pharmaceutical and intellectual property know-how, as well as training of the young investigators involved.

The scientific objectives of the project are:

- ✓ synthesis of modulators of epigenetic enzymes (epi-enzymes);
- ✓ synthesis of modulators of protein/protein interactions (NRs, epi-enzymes);
- ✓ definition of crystal structures of NR complexes and epigenetic enzymes bound to modulators;
- ✓ characterisation and validation of epi-drugs and their crosstalk with the NR signalling;
- ✓ global analysis of NR and chromatin-modifying complexes:
- ✓ development of the laser-based chromatin immunoprecipitation (LChIP-seq).

The consortium combines knowledge from a multiplicity of fields to teach young investigators a variety of modern techniques applied to cancer biology and to develop novel structure-based modulators for the treatment of a particular disease. This corresponds

conceptually to a “drug discovery pipeline” at the academic level with high innovation potential and should be ideal to expose young investigators to the necessity of multidisciplinary collaboration in the reality of today’s pharmaceutical industry.

2.1.2 THE MISSION OF THE CBS

Centre de Biochimie Structurale (CBS) at the University 1 of Montpellier is involved in determination of 3D structures of biomolecules and biomolecular assemblies, as well as in elucidation of their structure-function relationships. The strength of the CBS lies in the multidisciplinary character of the research that combines atomic level structures, thermodynamics, kinetics, microscopy and bioinformatics. The CBS is fully equipped for studies in crystallography (TECAN Genesis and Cartesian nano X8 crystallization robots, Rigaku X-ray diffractometer with Osmic mirrors and MAR research image plate), fluorescence spectroscopy, NMR and atomic force microscopy. The CBS lab will provide for the project the structure-based knowledge on NRs, epigenetic enzymes and protein-protein interactions. They s will express and purify the proteins using various expression systems and chromatographic methods. Before crystallization trials (1.800 crystallization conditions will be screened with the crystallization robots) the integrity of the samples will be analyzed by a plethora of techniques such as electrophoresis, dynamic light scattering, mass spectrometry or NMR spectroscopy. Diffraction data will be collected at the European Synchrotron Radiation Facility (ESRF; Grenoble). Resolution of the structures will be performed by the Molecular Replacement technique using the Protein Data Bank (PDB) or by the Multi-wavelength Anomalous Dispersion (MAD) methodology if produced in selenomethionine-labeled forms. Experimental models will be constructed, refined and analyzed.

2.2 PROTEINS

The expression protein comes from Greek word “protos”, meaning the first element. Proteins are essential for life, as they needed for growth, repair, good functioning and structure of living cells. For example hormones like insulin, which is essential for controlling the level of blood sugar; enzymes, such as amylases, lipases and proteases being crucial for digestion of food; antibodies that represent indispensible component of our immune system

and muscle proteins that allow muscular contraction, are only few examples of their extremely important functions within the human organism.

2.2.1 PROTEIN BASICS (6)

Protein 3D structure is very complex and closely packed. This is a key feature allowing their various functions. Proteins are macromolecules or heteropolymers as they are made of various combinations of 20 different L-amino acids, also referred to as aa residues.
The term peptide is frequently used for about 40 aa residues. It appears that 40-50 aa residues represent the lower limit for a functional domain size. Obviously a certain number of aa residues is necessary for a particular biochemical function of the protein. Protein sizes can range from the lower limit (40-50) to several hundred aa residues. Aminoacids can be classified as either essential or non-essential. The essential aminoacids cannot be produced by the metabolism of the body, while the non-essential ones can be made endogenously.
Very large aggregates can also be formed from protein subunits, for example many thousand actin molecules are assembled into an actin filament. Large protein complexes with RNA, the ribozymes, are also found in ribosomal particles.
Each protein has its own specific number and sequence of aminoacids, which makes it unique in the body. Depending on its particular aa sequence, the resulting protein carries out specific function in the body. Its shape or folding is important for its function.

2.2.2 PROTEIN STRUCTURES (6, 7)

There are four different types of protein structures, primary, secondary, tertiary and quaternary.

2.2.2.1 The primary protein structure

In proteins, aminoacids are linked together in a form of chains by peptide (amide) bonds that form the molecular backbone. A peptide bond is formed between a basic amino group ($-NH_2$) on one aminoacid and an acidic carboxyl group (-COOH) on another. The general formula of an aminoacid is H_2N –CHR –COOH. The R group can be just anything from an atom to a very complex molecule. The term “polypeptide” refers to a long chain of aminoacids.

Once the protein chain has been made, it must fold up properly before it can do its job. The structure and folding of each protein is specific. How the aminoacid sequence causes the protein to fold properly is not yet completely understood. A protein can be made up of one or more separate polypeptides. It can also be made of sheets with aminoacid chains lining up together, forming spiral structures. Clearly, the properties of polypeptides and proteins depend on their aminoacid composition.
The primary structure of a protein is its linear sequence of amino acids together with defined locations of possible disulfide (-S-S-) bridges.

2.2.2.2 The secondary protein structure

Most proteins contain one or more stretches of amino acids that take on a characteristic structure in the 3-D space. The most common of these are the alpha helix and the beta conformation. The secondary structure is basically a result of local hydrogen bonds being created along the polypeptide backbone, giving the protein its strength and flexibility.

• The **alpha-helix** is a consequence of hydrogen-bonding within the polypeptide chain, for example in muscle proteins. The R groups of amino acids all extend to the outside and the helix which is right-handed makes a complete turn every 3.6 amino acids, therefore being twisted in a clockwise direction. The carbonyl group (-C=O) of each peptide bond extends parallel to the axis of the helix and points directly at the -N-H group of the peptide bond, 4 amino acids below it. In that way a hydrogen bond [-N-H·····O=C-] can easily be formed between them.

• The **beta conformation** is created by hydrogen-bonding between adjacent polypeptide chains, for example in silk fibroin. It consists of pairs of chains lying side-by-side and being stabilized by hydrogen bonds between the oxygen atom of the carbonyl moiety on one chain and the -NH group on the adjacent one. The chains are often anti-parallel, meaning that the N-terminal to C-terminal direction of the first chain is reverse to the other.

2.2.2.3 The tertiary protein structure

The tertiary protein structure results from interactions between aminoacid R-groups in a polypeptide, such as non-covalent bonds (hydrogen bonds, ionic bonds, hydrophobic interactions) and weak, covalent bonds (disulphide bonds between cysteine residues). The physiological function of a protein greatly depends on its tertiary structure. If this is disrupted, the protein is said to be denatured, and it loses its activity. For example, denatured enzymes

lose their catalytic power or denatured antibodies can no longer bind antigen. A mutation in the gene encoding a particular protein is a frequent cause of altered tertiary structure.

The tertiary structure of many proteins is built from several domains. Often each domain has a separate function within the protein, such as: binding a small ligand (e.g., a peptide within the molecule), spanning the plasma membrane (in transmembrane proteins), containing a catalytic site (in enzymes), DNA-binding (in transcription factors) or providing a surface needed for specific binding to another protein.
In some cases, each domain of a protein is encoded by a separate exon within the protein coding gene.

2.2.2.4 The quaternary protein structure

Often, two or more polypeptides need to interact in order to allow a protein to perform its particular function. If this is the case, at least dimeric structures are formed. Such complexes of two or more polypeptide chains are usually held together by noncovalent forces (hydrogen bonds, ionic bonds) and less commonly by hydrophobic interfaces and inter-chain disulphide bonds, but in precise ratios and with a precise 3-D configuration. A typical exaple of such interaction is the noncovalent association of beta-2 microglobulin with the heavy chain of a class I histocompatibility molecule.

2.2.3 PROTEIN FUNCTIONS (6, 7, 8, 9)

The three-dimensional structure of a given protein is a result of interactions within its internal environment. The knowledge about the protein structure can tell us a lot about how they perform their physiological or pathological tasks in the body or within the cell.
There are several factors that can greatly affect protein structure and thereby also their function. For example in aqueous environment the hydrophobic aminoacid R-groups are positioned towards the protein's interior. The temperature and pH values have a great impact on protein activity, as just minimal changes in temperature or pH can interfere with the non-covalent bonding within the protein, causing disruption of its three-dimensional structure and the loss of activity (denaturation). Denatured proteins can also clump together and become insoluble in a process called coagulation.

There is a huge diversity of protein functions and here are some examples:

- ✓ structural – collagen in muscle fibres;
- ✓ storage – wheat gliadins or barley hordeins;
- ✓ enzymatic activity - hydrolases, transferases, isomerases, polymerases, ligases, etc.;
- ✓ transport - oxygen transfer with haemoglobin;
- ✓ messenger - insulin, certain other hormones and neurotransmitters;
- ✓ antibodies - proteins that bind to specific foreign antigens/particles;
- ✓ regulation - proteins involved in regulating DNA synthesis.

2.3 EPIGENETICS (10)

Epigenetics ("epi" means outside) is area part of molecular biology, which examines changes in the expression of genes within the organism that are not associated with mutations in DNA sequence. Namely, gene expression may also be affected by the environmental factors and not only by changes in the coding sequence of DNA. Epigenetics is mainly concerned with molecules which have the ability to bind to a particular section of DNA causing activation or inactivation of individual gene transcription. This is a consequence of biochemical reactions that add certain moieties, for example simple methyl groups to the DNA chain. Such methylation can prevent the translation of the genetically encoded information into a particular protein. The most obvious example of epigenetic activity and significance is cell differentiation. All cells in a multicellular organism contain the same genetic code and chemical modifiers such as hormones, growth factors, etc., but can nevertheless during development from stem cells differentiate into specialized tissue elements. Therefore epigenetics represents one of the basic principles of eukaryotic biology.

Contemporarily there is a great expansion in researching epigenetic enzymes, being potential targets for cancer treatment. Epigenetic drugs have a much wider impact than genetic ones because they can affect large number of genes and not just one. Recent research has shown that hypo- and hyper methylation can both lead to over differention of cells. By using molecules which could increase or decrease the methylation processes, we would hypothetically be able to interfere with cancer formation and apply them for its treatment.

2.4 PROTEIN DOMAINS

All proteins consist of different domains. In chromatin regulators, functionally the most important ones are the PHD and the SET domains.

2.4.1 THE PHD FINGER

The PHD finger (Plant Homeo Domain) was discovered in 1993 as a Cys4-His-Cys3 motif within the homeodomain protein HAT3 from Arabidopsis thaliana. The PHD finger motif resembles the metal binding RING domain (Cys3-His-Cys4) and the FYVE (four cysteine-rich proteins) domain. It occurs as a single finger, but is often found in clusters of two or three, and it also occurs together with other domains, such as the chromodomain and the bromodomain (11).

The lenght of the PHD finger is approximately 50-80 aminoacids. It is found in more than 100 human proteins. Several of these proteins are located in the nucleus and are involved in chromatin-mediated gene regulation. The PHD finger occurs in proteins such as: transcriptional co-activators p300, Polycomb-like protein, Trithorax-group proteins, like ASH1L (a probable histone-lysine N-methyltransferase ASH1L), ASH2L (Set1/Ash2 histone methyltransferase complex subunit ASH2) and MLL (Myeloid/lymphoid or Mixed-Lineage Leukemia protein), the autoimmune regulator AIRE (transcription factor expressed in the medulla of the thymus), the Mi-2 complex (part of histone deacetylase complex), the co-repressor TIF1 and many others (12).

The NMR structure of the PHD finger isolated from the human WSTF (Williams-Beuren Syndrome Transcription Factor) shows that the conserved cysteines and histidine coordinate two Zn^{2+} ions. In general, the PHD finger adopts a globular fold, consisting of a two-stranded beta-sheets and an alpha-helix. The region consisting of these secondary structures and the residues involved in coordinating the zinc-ions are highly conserved among species (13).

2.4.2 THE CENTRAL SET DOMAIN

The SET domain is a 130-amino acid long, evolutionarily conserved sequence motif, present in chromosomal proteins that function as modulators of gene activation from yeast to mammals. It appears to be a protein-protein interaction domain, as it has been proven to

mediate interactions with a family of proteins that display similarity with dual-specificity phosphatases. The SET domain consists of two regions known as SET-N and SET-C. The SET-C domain forms an unusual and conserved knot-like structure with a probable functional importance. An insert region SET-I and flanking regions of high structural variability are additional parts of the SET overall structure (13).

2.5 THE MLL5 PROTEIN

Full name: Histone-lysine N-methyltransferase, Myeloid/lymphoid or mixed-lineage leukemia protein 5 (trithorax homolog, Drosophila) (Figure 1)

Size: 1858 aminoacids, 204 965 Da

Location: chromosome 7q22 (Figure 2)

Taxonomic lineage: Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

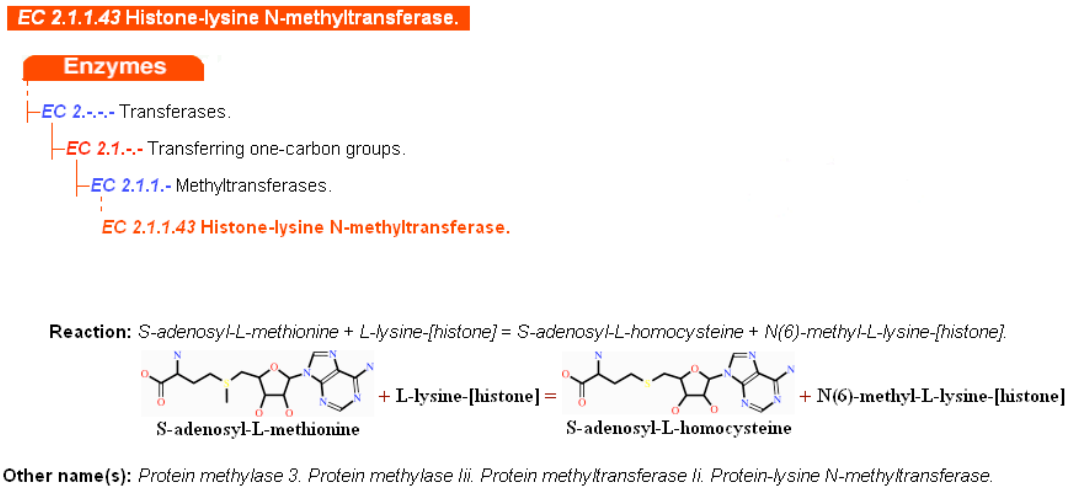

Figure 1: The MLL5 classification (15).

Figure 2: The localization of MLL5 gene on the chromosome 7 (16).

The enzyme MLL5 can structurally be divided into four components (Figure 3, 4 and 5):

- the PHD domain: bp from **118 to 166** - total lenght of 49 bp,
- the SET domain: bp from **328 to 451** - total length of 124 bp,
- the potencial coiled coil: bp from **559 to 615** - total lenght of 57 bp,
- the compositional bias (Pro-rich): bp from **1433 – 1846** - total lenght of 414 bp.

Figure 3: Relative localizations of four structural components coding areas within the MLL5 gene: the PHD domain (upper line), the SET domain (second line), the potential coiled coil (third line) and the compositional bias (lower line) (4).

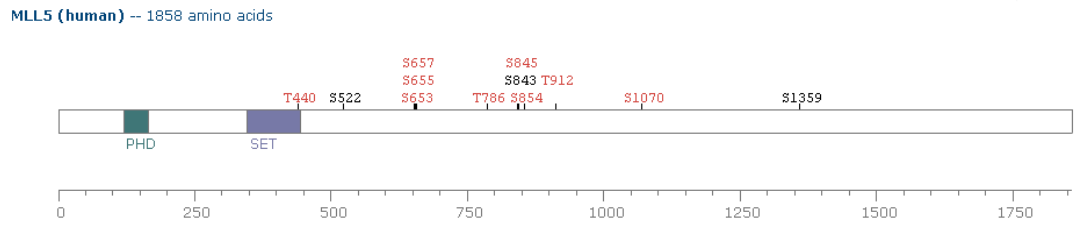

Figure 4: Positions of the PHD and SET domain coding regions within the MLL5 gene (17).

The whole aminoacid sequence of the protein MLL5 (1858 aa residues in total) is presented in Figure 5.

```
        10         20         30         40         50         60
MSIVIPLGVD TAETSYLEMA AGSEPESVEA SPVVVEKSNS YPHQLYTSSS HHSHSYIGLP

        70         80         90        100        110        120
YADHNYGARP PPTPPASPPP SVLISKNEVG IFTTPNFDET SSATTISTSE DGSYGTDVTR

       130        140        150        160        170        180
CICGFTHDDG YMICCDKCSV WQHIDCMGID RQHIPDTYLC ERCQPRNLDK ERAVLLQRRK

       190        200        210        220        230        240
RENMSDGDTS ATESGDEVPV ELYTAFQHTP TSITLTASRV SKVNDKRRKK SGEKEQHISK

       250        260        270        280        290        300
CKKAFREGSR KSSRVKGSAP EIDPSSDGSN FGWETKIKAW MDRYEEANNN QYSEGVQREA

       310        320        330        340        350        360
QRIALRLGNG NDKKEMNKSD LNTNNLLFKP PVESHIQKNK KILKSAKDLP PDALIIEYRG

       370        380        390        400        410        420
KFMLREQFEA NGYFFKRPYP FVLFYSKFHG LEMCVDARTF GNEARFIRRS CTPNAEVRHE

       430        440        450        460        470        480
IQDGTIHLYI YSIHSIPKGT EITIAFDFDY GNCKYKVDCA CLKENPECPV LKRSSESMEN

       490        500        510        520        530        540
INSGYETRRK KGKKDKDISK EKDTQNQNIT LDCEGTTNKM KSPETKQRKL SPLRLSVSNN

       550        560        570        580        590        600
QEPDFIDDIE EKTPISNEVE MESEEQIAER KRKMTREERK MEAILQAFAR LEKREKRREQ

       610        620        630        640        650        660
ALERISTAKT EVKTECKDTQ IVSDAEVIQE QAKEENASKP TPAKVNRTKQ RKSFSRSRTH

       670        680        690        700        710        720
IGQQRRRHRT VSMCSDIQPS SPDIEVTSQQ NDIENTVLTI EPETETALAE IITETEVPAL

       730        740        750        760        770        780
NKCPTKYPKT KKHLVNEWLS EKNEKTGKPS DGLSERPLRI TTDPEVLATQ LNSLPGLTYS
```

```
       790        800        810        820        830        840
PHVYSTPKHY IRFTSPFLSE KRRRKEPTEN ISGSCKKRWL KQALEEENSA ILHRFNSPCQ

       850        860        870        880        890        900
ERSRSPAVNG ENKSPLLLND SCSLPDLTTP LKKRRFYQLL DSVYSETSTP TPSPYATPTH

       910        920        930        940        950        960
TDITPMDPSF ATPPRIKSDD ETCRNGYKPI YSPVTPVTPG TPGNTMHFEN ISSPESSPEI

       970        980        990       1000       1010       1020
KRRTYSQEGY DRSSTMLTLG PFRNSNLTEL GLQEIKTIGY TSPRSRTEVN RQCPGEKEPV

      1030       1040       1050       1060       1070       1080
SDLQLGLDAV EPTALHKTLE TPAHDRAEPN SQLDSTHSGR GTMYSSWVKS PDRTGVNFSV

      1090       1100       1110       1120       1130       1140
NSNLRDLTPS HQLEVGGGFR ISESKCLMQD DTRGMFMETT VFCTSEDGLV SGFGRTVNDN

      1150       1160       1170       1180       1190       1200
LIDGNCTPQN PPQKKKVSLL EYRKRQREAR KSGSKTENFP LISVSPHASG SLSNNGDGCA

      1210       1220       1230       1240       1250       1260
SSNDNGEQVD HTASLPLPTP ATVYNATSEE TSNNCPVKDA TASEKNEPEV QWTASTSVEQ

      1270       1280       1290       1300       1310       1320
VRERSYQRAL LLSDHRKDKD SGGESPCVSC SPSHVQSSPS SHSNHIPQLQ AKGPVPSFSE

      1330       1340       1350       1360       1370       1380
LMEDPDPENP EPTTTNECPS PDTSQNTCKS PPKMSKPGSP GSVIPAQAHG KIFTKPDPQW

      1390       1400       1410       1420       1430       1440
DSTVSASEAE NGVHLKTELQ QKQLSNNNQA LSKNHPPQTH VRNSSEQLSQ KLPSVPTKLH

      1450       1460       1470       1480       1490       1500
CPPSPHLENP PKSSTPHTPV QHGYLSPKPP SQQLGSPYRP HHSQSPQVGT PQREPQRNFY

      1510       1520       1530       1540       1550       1560
PAAQNLPANT QQATSGTLFT QTPSGQSSAT YSQFNQQSLN STAPPPPPPP PPSSSYYQNQ

      1570       1580       1590       1600       1610       1620
QPSANFQNYN QLKGSLSQQT VFTSGPNQAL PGTTSQQTVP GHHVTPGHFL PSQNPTIHHQ

      1630       1640       1650       1660       1670       1680
TAAAVVPPPP PPPPAPGPHL VQQPNSHQQH SVAHVVGPVH AVTPGSHIHS QTAGHHLPPP

      1690       1700       1710       1720       1730       1740
PPPPGPAPHH HPPPHPSTGL QGLQAQHQHV VNSAPPPPPP PPPSSVLASG HHTTSAQALH

      1750       1760       1770       1780       1790       1800
HPPHQGPPLF PSSAHPTVPP YPSQATHHTT LGPGPQHQPS GTGPHCPLPV TGPHLQPQGP

      1810       1820       1830       1840       1850       1860
NSIPTPTASG FCPHPGSVAL PHGVQGPQQA SPVPGQIPIH RAQVPPTFQN NYHGSGWH
```

Figure 5: Complete aminoacid sequence of the protein MLL5 (4). The sequences of the four structural components are shown in different colours: the PHD domain in green, the SET domain in red, the potential coiled coil in light blue and the compositional bias in violet.

The MLL5 protein is located in the nucleus specle of the cell. It is widely expressed in both adult and fetal tissues. Its highest levels are found in fetal thymus and kidneys, as well as in adult hematopoietic tissues, jejunum and cerebellum (5).

2.5.1 THE FUNCTIONS OF MLL5

The MLL5 protein is a divergent member of the drosophila trithorax-related (SET) domain and plant homeodomain (PHD) domain-containing chromatin regulators that are involved in the regulation of transcriptional "memory" during cell differentiation. It forms intranuclear protein complexes that may play an important role in chromatin remodeling and cellular growth suppression. Basically, it is a histone methyltransferase which specifically mono- and dimethylates 'Lys-4' of histone H3 (H3K4me1 and H3K4me2) (1, 2, 3, 4). The H3 'Lys-4' methylation represents a specific tag for subsequent epigenetic transcriptional activation. This enzyme is a key regulator of hematopoiesis, as it has been shown that it is involved in terminal myeloid differentiation and in regulation of hematopoietic stem cell (HSC) self-renewal, by a mechanism that involves DNA methylation (4).

It plays an essential role in retinoic-acid-induced granulopoiesis by acting as a co-activator of RAR-alpha (RARA) in target gene promoters and acts as an important cell cycle regulator, participating in cell cycle regulatory network machinery at multiple cell cycle stages (3). Namely, the MLL5 is required to suppress inappropriate expression of S-phase-promoting genes and maintain expression of determination genes in quiescent cells. Its overexpression inhibits cell cycle progression, while its knockdown induces cell cycle arrest at both, the G1 and G2/M phases (1).

The components of the MLL5-L complex (the name of the complex, where each of the following proteins is found), composed at least of MLL5 are (18):

- ✓ the STK38 (serine/threonine-protein kinase 38);
- ✓ the PPP1CA (serine/threonine-protein phosphatase);
- ✓ the PP1-alpha catalytic subunit;
- ✓ the PPP1CB (serine/threonine-protein phosphatase PP1-beta catalytic subunit);
- ✓ the PPP1CC (serine/threonine-protein phosphatase PP1-gamma catalytic subunit);
- ✓ the HCFC1 (Host cell factor);
- ✓ the ACTB (Beta-actin);

- ✓ the OGT (UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110 kDa subunit).

It has been shown that by alternative splicing, 7 isoforms of the MLL5 protein can be produced in humans (4).

MLL5 is also a candidate tumor suppressor gene. It is a frequently deleted element found in cytogenetic aberrations of acute myeloid malignancies (19). Recent studies have shown that it can cause leukemia by rearranging and fusing with other genes in different chromosomes, which is a process under epigenetic control (20). Therefore the MLL5 protein represents a promising target for development of epigenetic drugs that could be potentially used for treating certain forms of cancer.

3 THE WORK PLAN

The goal of our work will be to define the structure of the catalytic, i.e. the SET domain of the MLL5 protein via determining its 3D crystal structure, which is still unavailable. For that purpose we will use x-ray crystallography and NMR analysis. In order to produce soluble proteins for crystallization, different constructs for each region studied will be designed by using appropriate bioinformatics tools. As already mentioned in the previous chapter, the MLL5 gene codes for the following structural components of the protein:

- ✓ the SET domain from bp 328 to bp 451,
- ✓ the PHD domain from bp 118 to bp 166,
- ✓ the potencial coiled coil from bp 559 to bp 615,
- ✓ the compositional bias (Pro-rich) from bp 1433 to bp 1846.

All molecular biology procedures needed for the preparation of different protein constructs will be performed in collaboration with OPPF (Oxford Protein Production Facility, Oxford). Namely, at the OPPF, a versatile suite of pOPIN vectors suitable for expression of proteins or their targeted domains have been developed. The majority of these utilize multiple promoter systems so that single constructs can be screened for their expression, both in E. coli or insect host cells, thereby avoiding the need to prepare and use multiple vectors for each target. All the constructs that will be used have been expressed and screened in E. coli in order to select the best ones that will allow the production of sufficient amounts of soluble proteins.

All fusion constructs will contain a 6-His purification”tag” to simplify the downstream expression screening and purification processes. Expression screens in E.coli of all the constructs are being carried out in order to select the best construct that will alow producing sufficient amount of soluble proteins, in collaboration with OPPF. After that, proteins will be expressed in E.coli, and then purified using adequate affinity chromatography, gel filtration… All fusion constructs will be prepared by the OPPF lab and inserted into two different plasmids, i.e. the OPIN J and the OPIN F in order to obtain the two corresponding fusion proteins with the His taq (OPIN F) or His-GST taq (OPIN J).

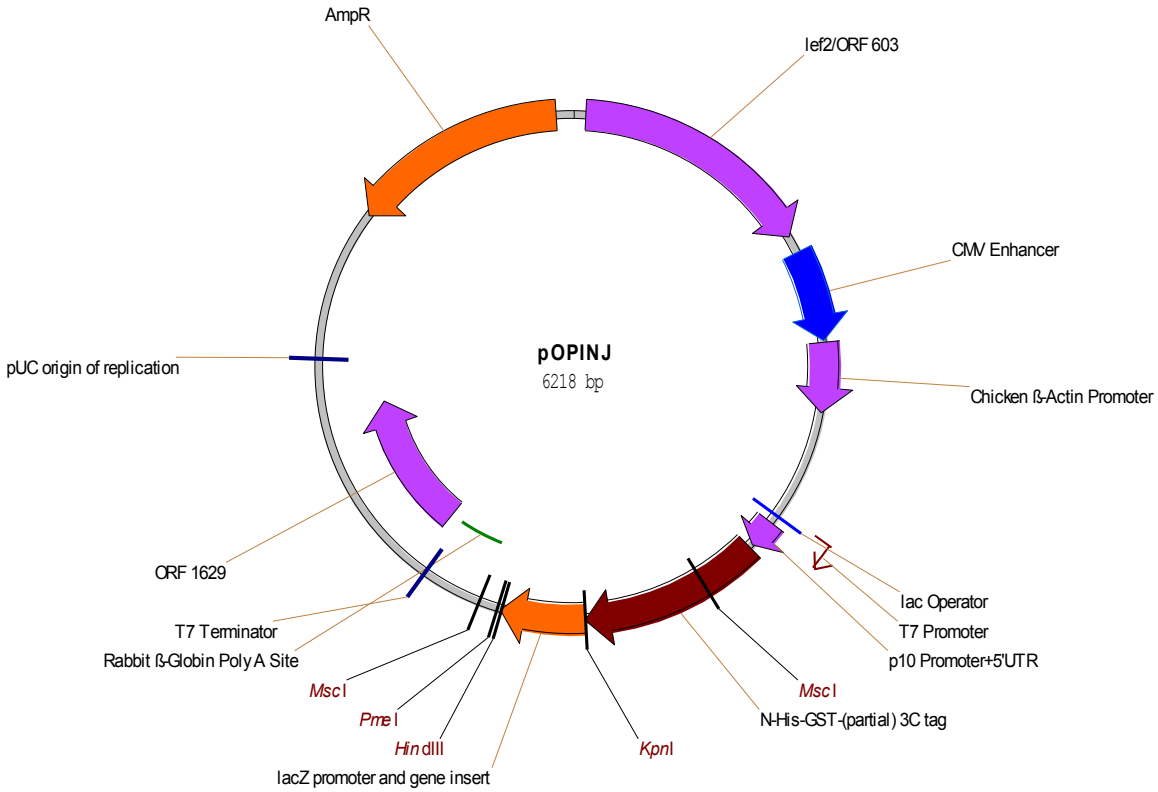

Figure 6: A schematic representation of the pOPINJ plasmid, as constructed by the OPPF.

```
                        N-His-GST-(partial) 3C tag
                                      ~~~~~~~~~~~~~~~~~~~
                                       M  A  H    H  H  H
2301                                  ATGGCACA TCACCATCAC
                                      TACCGTGT AGTGGTAGTG
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       H  H  M  S    P  I  L    G  Y  W    K  I  K  G    L  V  Q  ·
2351  CATCACATGT CCCCTATACT AGGTTATTGG AAAATTAAGG GCCTTGTGCA
      GTAGTGTACA GGGGATATGA TCCAATAACC TTTTAATTCC CGGAACACGT
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      · P  T  R    L  L  L  E    Y  L  E    E  K  Y    E  E  H  L  ·
2401  ACCCACTCGA CTTCTTTTGG AATATCTTGA AGAAAAATAT GAAGAGCATT
      TGGGTGAGCT GAAGAAAACC TTATAGAACT TCTTTTTATA CTTCTCGTAA
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      ·  Y  E  R    D  E  G    D  K  W  R    N  K  K    F  E  L
2451  TGTATGAGCG CGATGAAGGT GATAAATGGC GAAACAAAAA GTTTGAATTG
      ACATACTCGC GCTACTTCCA CTATTTACCG CTTTGTTTTT CAAACTTAAC
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       G  L  E  F    P  N  L    P  Y  Y    I  D  G  D    V  K  L  ·
2501  GGTTTGGAGT TTCCCAATCT TCCTTATTAT ATTGATGGTG ATGTTAAATT
      CCAAACCTCA AAGGGTTAGA AGGAATAATA TAACTACCAC TACAATTTAA
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                    MscI
                   ~~~~~~
      · T  Q  S    M  A  I  I    R  Y  I    A  D  K    H  N  M  L  ·
2551  AACACAGTCT ATGGCCATCA TACGTTATAT AGCTGACAAG CACAACATGT
      TTGTGTCAGA TACCGGTAGT ATGCAATATA TCGACTGTTC GTGTTGTACA
                   N-His-GST-(partial) 3C tag
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

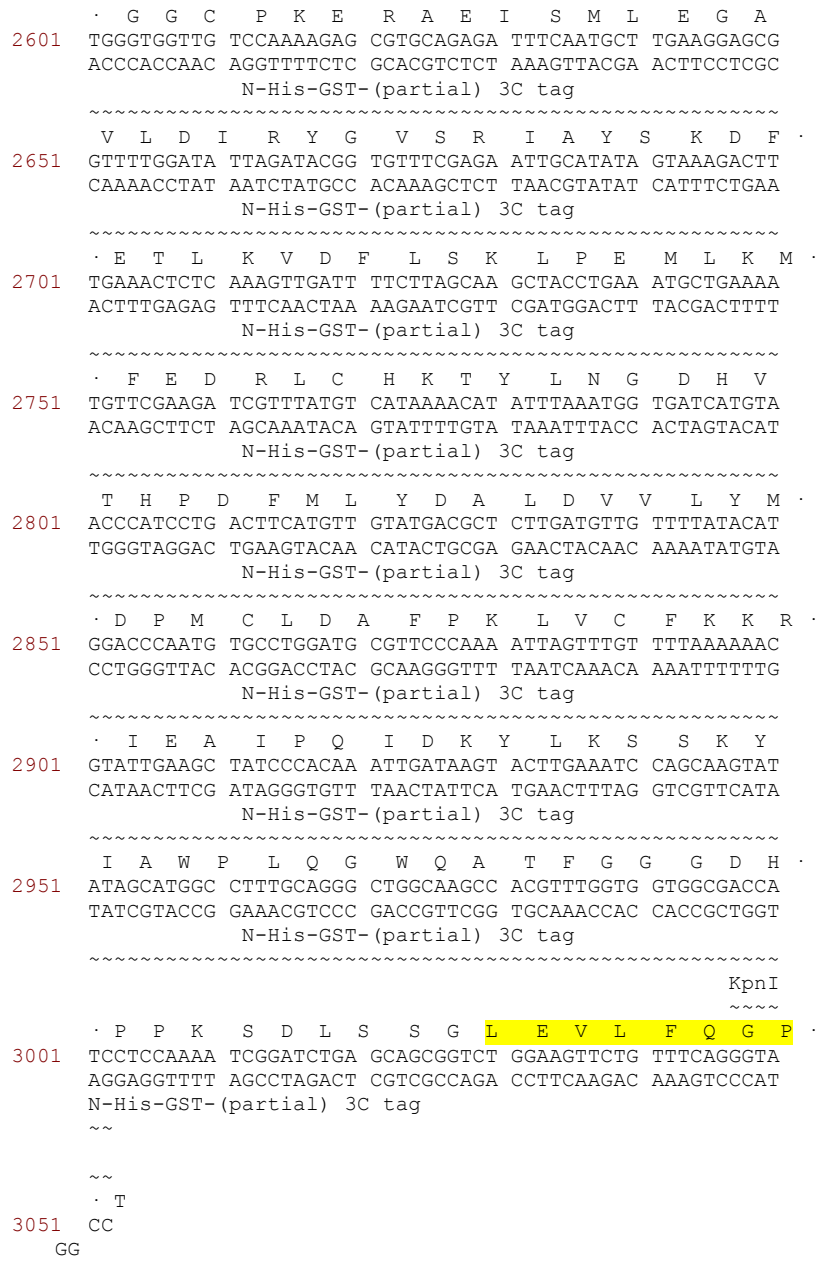

Figure 7: Aminoacid sequence of the inserted plasmid.

Expression screenings will be carried out at the OPPF for 3 constructs transferred either into B834 or Rosetta E. coli strains. Based on the preliminary results of these screenings we will decide with which constructs and E. coli strains to continue (See Table I, page 26).

The full size of the MLL5-3J protein at the beginning of its isolation process is 75 kDa (kilodaltons). Our goal will be to isolate the aminoacid sequence from 323 to 473 that contains the catalytic SET domain, being rich in cysteins. For that purpose we will cut out all

parts of the protein that are not of our interest. In order to do that, we will use different approaches, for example filtration through the Ni column, restriction with the protease P3C (a 3C cysteine protease), filtration through the GST column and gel filtration (GF).
The end size of the isolated SET domain should be 17 kDA.

To summarize, our work will be divided into following steps:

1.) Preparation of cells:
   - ✓ preparation of Petri dishes for cell cultures;
   - ✓ constructs insertion into E.Coli strains (Rosetta);
   - ✓ preparation of the glycerol stock solution;
   - ✓ preparation of pre-cultures;
   - ✓ preparation of cell cultures;
   - ✓ harvesting of proteins from host cells.

2.) Purification of the protein:
   - ✓ sonication of host cells;
   - ✓ preparation of the Ni column;
   - ✓ equilibration of the chromatography apparatus;
   - ✓ affinity chromatography (His Trap or Ni-column);
   - ✓ preparation of the GST column;
   - ✓ P3C cleavage of the fusion protein;
   - ✓ calibration of the column for gel filtration;
   - ✓ gel filtration (GF);
   - ✓ GSH Cleavage of the GST or/and Ni-column;
   - ✓ concentration of the isolated protein by amicons cells under pressure and tubes via centrifuge;
   - ✓ characterization of the protein (ε - molar extinction coefficient and MW – molecular weight);
   - ✓ separation of the protein with SDS-page electrophoresis;
   - ✓ measurements of protein concentration (Bradford method and nanodrop spectrophotometry at 280 nm);
   - ✓ circular dichroism (CD);
   - ✓ dynamic light scattering (DLS).

3.) Crystallization of the purified protein.

4.) Optimization of crystallization buffers and the crystallization process itself, in order to obtain suitable crystals for further analysis.

4 MATERIALS AND METHODS

4.1 MATERIALS

4.1.1 SMALL LABORATORY MATERIALS

- ✓ epruvettes (250 µl, 500 µl, 1 ml, 10 ml, 25 ml, 100 ml);
- ✓ single use plastic pipettes and pipetting devices (Pipetman ultra produced by Gilson¸ Medibase);
- ✓ laboratory glassware (graduated cylinders, jars, erlenmeyer flasks, beakers);
- ✓ syringe needles (BD);
- ✓ kit for acryl gel preparation: plastic holder, glass plate, comb, silicon plates (Pierce);
- ✓ magnetic stir/stirrer (Gilson);
- ✓ filters – brown 50 µl and yellow 0.45 µm (Ministar);
- ✓ crystallization plates type EasyXtal 15-Well;
- ✓ corning microtiter plates; 1:4, 96 well.

4.1.2 EQUIPMENT

- ✓ pH meter: Inolab WTW SERIES pH720;
- ✓ peristaltic pump: type Miniplus3 (Gilson);
- ✓ automatic purifier for Chromatography (AKTA);
- ✓ gel filtration column: Hiload tm 26/60 superdex 75 pergrade (Average particle size 13 µm 13 µm);
- ✓ spectrophotometer: Nanodrop 2000 Theremo Scientific;

- ✓ centrifuge: Megafuge 10 Hereaus Theremo Scientific;
- ✓ centrifuge: Sorvall elution RC, Thermo Scientific;
- ✓ centrifuge: KR 25i, Jouan;
- ✓ crystallization robots: Tecan genesis RSP 100 with cartesian dispensing system;
- ✓ electrophoresis kit (Pierce);
- ✓ a heat-shock device (Gilson);
- ✓ Ni-column;
- ✓ GST-column.

4.1.3 CHEMICALS AND BUFFERS

- ✓ buffer for resuspending cell pellets: 50 mM Tris-HCl, 150 mM NaCl, pH 8.0;
- ✓ TGS (Tris-glycine-SDS) running buffer: 25 m M Tris, 192 mM, 0.1% SDS, pH 6.8;
- ✓ imidazole gradient buffer: 50 mM Tris, 300mM NaCl, 0.5 M imidazole, pH 8.0;
- ✓ GSH (Glutatione) buffer: 20 mM TRIS (2-Amino-2-hydroxymethyl-propane-1,3-diol), 100 mM GSH, pH8;
- ✓ washing buffer - PBS (Phosphate Buffered Saline) – Ni-column - 2nd and 3rd trial: 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4;
- ✓ hepes buffer: 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 100 mM NaCl, 5 mM DTT, pH 7;
- ✓ washing buffer - GF, the 1st trial: 50 mM TRIS, 150 mM NaCl, 1 mM DTT, pH 7.5;
- ✓ washing buffer - GF, the 2nd and the 3rd trial: 0.1 M HEPES, 150 mM NaCl, 2 mM beta-mercaptoethanol, pH 7;
- ✓ second washing buffer - GF, the 3rd trial: 50 mM TRIS, 500 mM NaCl, 5 mM DTT, pH 7.5;
- ✓ buffer for reconcentration: 50 mM TRIS, 500 mM NaCl, 5 mM DTT, pH 7.7;
- ✓ column equilibration buffer: 100 mM HEPES, 150 mM NaCl, 2 mM beta-mercaptoethanol;
- ✓ SAM - S-adenosyl-L-methionine, powder, Gentaur Laboratories;
- ✓ **lysogeny broth (LB) for agar plates:** to a flask of volume at least 2 L, add:
    - 10 g tryptone, 5 g yeast extract, 5 g NaCl,
    - 800 mL of distilled water (stir the solution until everything is completely dissolved),

- add 400 µL of 5 M NaOH with stirring to adjust the pH,
- bring the liquid level up to 1.000 ml with distilled water,
- add 15 g of granulated agar to the liquid and stir until the agar is dissolved,
- remove the stir bar, cover the flask with aluminum foil and autoclave for 20 min using the liquid cycle.

✓ lysogeny broth (LB) - a nutritionally rich mediumfor growing bacteria (culturing and maintaining of recombinant strains of E.coli). LB generally includes the following ingredients: peptides and casein peptones, vitamins (including B vitamins), trace elements (e.g. nitrogen, sulphur, magnesium) and minerals. Preparation of 1 liter of LB medium:

- 10 g tryptone;
- 5 g yeast extract;
- 10 g NaCl;
- suspend the solids in ~800 ml of distilled or deionized water;
- add further distilled or deionized water, in a measuring cylinder to ensure accuracy, to make a total of 1 litre;
- autoclave at 121 °C, pH adjusted to 7.5 with NaOH.

✓ **ZYM 50-52 SOLUTION –** used for the onex medium (stock solution 50xM)

1.25 M Na_2HPO_4

+ 1.25 M KH_2PO_4

+ 2.5 M NH_4Cl

+ 0.25 M Na_2SO_4

H_2O was poured up to the volume of 500 ml and the resulting solution autoclaved for 15 min.

✓ **stock solution 50x 5052** - for the OnEx medium

125 g of 25 % glycerol

+ 12.5 g of 2.5 % glucose

+ 50 g of 10 % lactose

H_2O was poured up to the volume of 500 ml and the resulting solution autoclaved for 15 min.

✓ **ZY -** for the OnEx medium (for 1 L of medium)

10 g of 1% NZ amine

+ 5 g of 0.5% yeast extract

H_2O was poured up to the volume of 1000 ml and the resulting solution autoclaved for 15 min. It was prepared just before the usage.

- ✓ **ONEX MEDIUM** (for 500 ml) – a rich medium for the auto-induction
  463 ml of ZY
  + 1 ml of 1M $MgSO_4$
  + 10 ml of 50x5052
  + 10 ml of 50xM
- ✓ **carbenicillin (CARB)** (Gentaur Laboratories) - to prepare 0.1 M CARB:
  dissolve 0.5 g CARB in 10 ml ddH_2O (50 mg CARB/ml);
- ✓ **chloramphenicol (CAM)** (Mehta Pharmaceutical) - to prepare 0.1 M CAM:
  dissolve 0.34 g CAM in 10 ml EtOH (34 mg CAM/ml);
- ✓ **P3C - 3C Protease (Cysteine Protease)** (Sino Biological) - this enzyme recognizes the cleavage site: Leu-Glu-Val-Leu-Phe-Gln-↓-Gly-Pro. 2 ml of P3C (protease) with the concentration of 0.5 mg/ml is mixed with 3 ml of washing buffer. Incubation time: 16 hours (overnight), working temperature: 4 °C.

4.2 METHODS

4.2.1 PRODUCTION OF THE PROTEIN

4.2.1.1 Expression screening of construct

Different expression screenings have been carried out for all 3 plasmids containing constructs after being transferred either into B834 and Rosetta cells. The screening was performed by the OPPF.

The **B834** is a parental strain for BL21, which is widely known as the strain of choice for expression of target proteins in bacterial systems. These protease-deficient hosts are methionine auxotrophs and allow high specific-activity labeling of target proteins with 35S-methionine and selenomethionine for crystallography (18).

The **Rosetta** host strains are also BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in E. coli. These strains supply tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant

plasmid. Thus the Rosetta strains provide for "universal" translation which is otherwise limited by the codon usage of E. coli (19). (See Table 1, page 26)

4.2.1.1.1 IPTG AND ONEX™ INDUCED EXPRESSION OF THE PROTEIN

The protein was induced in two different ways – by IPTG and OnEx. The screening was performed for both of them in both types of cells and with all three constructs.

IPTG - Isopropyl-Thio-2-D-Galactopyranoside is used for inducing the protein expression from the lac promoter and its various derivatives. This molecule mimics allolactose, a lactose metabolite that triggers transcription of the lac operon. Unlike allolactose, the sulphur (S) atom of IPTG creates a chemical bond which is non-hydrolyzable by the cell, therefore preventing the cell from "eating up" or degrading the inductant. As a consequence the IPTG concentration remains constant. IPTG binds to the lac repressor and releases the tetrameric repressor from the lac operator in an allosteric manner, thereby allowing the transcription of genes in the lac operon, such as the gene coding for beta-galactosidase, a hydrolase enzyme that catalyzes the hydrolysis of β-galactosides into monosacharides.

In cloning experiments, colonies that have been transformed with the recombinant plasmid rather than a non-recombinant, need to be identified. X-gal is a substance that can be metabolised by beta-galactosidase to produce a blue product. Thus cells expressing beta-galactosidase grown in the presence of X-gal and IPTG (inducer of expression) will turn blue. Where a DNA fragment has been inserted into the LacZ (one of the genes for beta-galactosidase) there will be no action upon X-gal and the cells will not turn blue, thus identifying the cells that carry recombinant plasmid rather than the non-recombinant one.
The lac operon is an operon required for the transport and metabolism of lactose in E. coli and some other enteric bacteria. It consists of three adjacent structural genes, lacZ, lacY and lacA. The lac operon is regulated by several factors including the availability of glucose and of lactose (20).

With the **OnEx™** (Overnight Express Autoinduction System), a period of cell growth is followed by spontaneous induction of protein expression without monitoring cell density and without conventional induction with IPTG. The method is based on media components

that are metabolized differentially to promote cell growth to high density and automatically induce protein expression from lac promoters. Addition of these components to traditional glucose-free E. coli culture media, such as LB broth, results in high cell densities, autoinduction of expression, and maximum soluble protein yields (21). Cell mass and the target protein yield are often increased several-fold as compared with conventional protocols using IPTG induction.

Table I: Specification of MLL5 protein isolates.

name	Protein	Doma	Boundaries	Size (insert)	Vector	Size (construct)	B834_ IPTG _sol	B834_ IPTG _insol	B834_ Auto _sol	B834_ Auto _insol	Rosetta_ IPTG_sol	Rosetta_ IPTG_insol	Rosetta_ ONEX_sol	Rosetta_ ONEX_insol
ML-3J	hMLL5	3	323-473	17,743	OPINJ	45,683	*(*)	***	**	***	***	***	*(*)	***
ML-3F	hMLL5	3	323-473	17,645	OPINF	19,786	*	**	(*)					
ML-1F	hMLL5	1	1-570	67,569	OPINF	66,71		*		*		*		(*)

*Poor induction **Good induction ***Very good induction

Based on the results presented in Table (screening results provided by the OPPF) we decided to continue with the MLL5-3J (3 – represents domain and J represents the used vector) and with the construct size of 45.683 Da. We also decided to use only Rosetta cells in both protein induction conditions, i.e. IPTG and OnEx.

4.2.1.2 Preparation of LB agar gel in Petri dishes

A flask of pre-prepared LB agar gel composition was heated in a microwave oven until all the material was completely liquefied, after which it was poured into sterile Petri dishes in aseptic conditions. The agar plates were solidified in approximately 20 minutes. Restrictive, i.e. selective CAM (0.1 M) and CARB (0.1 M) solutions were placed onto whole gel surfaces in order to select the E. coli cells containing the plasmids. After that the plates were stored in a refrigerator at 4 °C.

4.2.1.3 Transfer of plasmids into E. coli

Two aliquots of Rosetta cells were taken from refrigerator and marked accordingly. We always used one as a control and one as a working sample. They were left on ice for 10 min. After adding 1 µl of the concentrated plasmid MLL5– 3J, which was prepared by the OPPF and stored in refrigerator at -70 °C, into the working sample we continued the incubation on ice for 30 min. Subsequently both samples were transferred to 42 °C for 90s to heat-shock the cells, thereby allowing the entrance of the plasmid into bacteria. After that, they were put on ice for additional 5 min. Into each sample 1 ml of the LB was added and

after gentle mixing at 240 rpm, the cells were left at 37 °C for one hour. Both, the control and the working sample were then transferred near the flame onto two separate sterile LB agar plates with a pipette and both samples were dispersed on the whole surface. Petri dishes were marked accordingly and put in an incubator at 37 °C over night. First 15 min they were incubated with the agar layer on top, so that the plates could dry, and then with the agar layer on bottom. Next day after the non-growth of non-transfected cells on the control plate was checked and confirmed, the plates were sealed and put in the refrigerator at 4 °C.

4.2.1.4 Stock glycerol

In order to store the cells in refrigerator at -70 °C for further analyses we prepared a stock solution by mixing 700 µl of LB pre-culture medium with antibiotics and 700 µl of 50% glycerol in each cryovial. We transferred the cells from agar plates into appropriate cryovials in the vicinity of the flame. The tubes were marked accordingly and stored at -70 °C.

4.2.1.5 Preparation of pre-cultures

For the preparation of a pre-culture, 1% of glucose solution, previously sterilized by filtration through a 0.20 µl Ministar filter, was added to the LB, in order to increase the cell growth. Then antibiotics, i.e. 0.1 mM CARB and 0.1 mM CAM, as well as cells, taken either from agar plates or from the stock solution kept at -70 °C, were added. The whole procedure was carried out in the presence of an open flame, so that cell cultures were kept sterile. Subsequently the pre-cultures were put in the incubator at 37 °C, overnight, under constant shaking at 220 rpm.

An illustrative example

To 100 ml LB 100 µl from the 0.1 M CARB stock solution, 100 µl from the 0.1 M CAM stock solution and 0.1 g glucose were added. The resulting solution was splitin two erlenmeyer flasks and then either the thawed cells from the frozen stock (-70 °C) or those taken from Petri dishes, stored in a refrigerator at 4° C were added to start the growth of pre-cultures.

100 ml LB
+ 0.1 g glucose
+ 100 µL of the 0.1 M CAM stock solution

+ 100 µL of the 0.1 M CARBstock solution
+ ROS(A) Cells (frozen or from Petri dish)
+ **100 µL $ZnSO_4$** (10 µM $ZnSO_4$, but only for the 3rd trial)

4.2.1.6 Preparation of cell cultures

Cell cultures were prepared in two different ways according to the method used for the growth induction, i.e. IPTG or OnEx.

IPTG

For the preparation of IPTG culture, the preculture, 0.1 mM Carbenicillin (CARB) and 0.1 mM Chloramphenicol (CAM) and 1% glucose were used. The cultures were put in the incubator at 37 °C with continuous shaking at 225 rpm, for 4 hours, and then the growth of the Rosetta cells was checked by measuring the absorbance at 600 nm. If the absorbance was between 0.6 – 1.0, 350 µl of 1 M IPTG was added to the culture (finial concentration of 0.5 mM IPTG) to induce the expression of the protein. The cell culture was then incubated at 20 °C overnight for approximately 20 hours.

LB medium (700 ml)
+ 700 µl of the 0.1 M CARB stock solution
+ 700 µl of the 0.1 M CAM stock solution
+ 0.7 g glucose (the solution has to be microbiologically filtered)
+ 10 ml preculture

OnEx

For the preparation of OnEx culture, the preculture, 0.1 mM Carbenicillin (CARB) and 0.1 mM Chloramphenicol (CAM) and 1% glucose were used. Then the cultures were transferred in the incubator at 37 °C with continuous shaking at 225 rpm for 4 hours and then for additional 20 hours at 25 °C.

OnEx medium (500 ml)
+ 500 µl of the 0.1 M Carb stock solution
+ 500 µl of the 0.1 M CAM stock solution
+ 10 ml of the ROS (A) cell suspension – the LB pre-culture

+ 0.5 g glucose (the solution has to be microbiologically filtered)

4.2.1.7 Harvesting of Rosetta host cells

When the cell growth in cultures was successful, the cells were transferred into special centrifuge tubes and centrifuged for 20 min at 18 000 rpm and the temperature of 4 °C. In the mean time the new tubes were prepared, each filled with 25 ml of 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 and put on the ice. After centrifugation the supernatants were poured off and the cell pellets were re-suspended and transfered into the new tubes containing buffers. The tubes were labelled and stored at -70 °C. This freezing step already fragilizes the cells to facilitate the isolation of the protein.

4.2.2 PURIFICATION OF THE PROTEIN

4.2.2.1 Sonication of host cells

Cryovial containing frozen cells, which were kept at -70 °C, were taken from the freezer and put into not too warm water for a couple of minutes in order to be thawed. In each thawed sample, an antiprotease - ½ EDTA free tablet was added. In order to additionally limit the protease activity all the tubes were kept on ice during the whole procedure. The next step in protein isolation was the sonication of cells. Each sample was sonicated for 2 minutes at the amplitude of 70%, for 3 consecutive rounds. After sonication each sample was separated into two tubes, which were kept on ice. They were balanced and centrifuged at 18.000 x g for 40 min at 4 °C. The supernatants, containing proteins were carefully removed and filtered through the production filters (ministar yellow - 50 µm and brown – 0.45 µm) into a measurement cylinder.

4.2.2.2 HisTrap (Ni column) purification

HisTrap is a pre-packed, ready-to-use column for the preparative purification of His-tagged recombinant proteins by immobilized metal affinity chromatography, using pre-charged Ni Sepharos High Performance column. The special design of the column, together with the high performance matrix, provides fast, simple and easy separations in a convenient format. HisTrap HP columns can be operated with a syringe, peristaltic pump or liquid chromatography system such as ÄKTA™ design or HPLC (22).

The polyhistidine binds strongly to divalent metal ions such as nickel and cobalt. The protein can be passed through a column containing immobilized nickel ions, which bind the polyhistidine tag. All untagged proteins pass through the column. The protein can then be eluted with imidazole (in our case), which competes with polyhistidine tag for binding to the column, or by a decrease in pH, which decreases the affinity of the tag for the resin.

4.2.2.3 Preparation of the Ni column

The HisTrap pre-packed column of 1 or 5 ml was used. First the column was washed with water by applying a volume which was 10 times the column volume. The water was pumped through the column with the peristaltic pump. For the stripping and reloading of the column it was manually washed, first with 10 ml EDTA solution and after that again with water (10 times the volume of the column). The last step was manually packing the column with a Ni suspension. This was put into injector and injected slowly into the column.

4.2.2.4 Equilibration of the chromatography apparatus

Before the chromatography apparatus (AKTA) was used, all its tubes and loops were first washed with water. Then, after the buffers were prepared, the equilibration was started. The pump A1 was put into loading buffer and B1 into elution buffer. On display we chose run → show details →column which will be used (5ml) → flow rate (in relation to the column used)→start concentration of 4% → direct sample 0→step10
The equilibration of the apparatus lasted about 30 minutes.

4.2.2.5 Histrap chromatography apparatus

We used the Akta apparatus that allows automatic protein purification. When the apparatus was equilibrated vand washed, the protein detection and separation procedure could be started with the use of a Ni column. As described, the sample was filtered and collected, usually in a measure cylinder, which was attached to the holder and the sample loop was linked to it. Loop A1 was then put into buffer A (loading buffer) and loop B1 into buffer B (elution buffer). The red loop, named also a flow through or F3, was placed in a special bottle to collect the flow through. On the display the following run orders were chosen: → show details → choose a column → no pump wash (washing was manual) → start concentration 4% → volume of the sample 2.000 ml → wash out unbound sample 20.00 ml. After the settings were chosen the analysis was named, for example Ni (column) ML3J (protein) OnEx

or LB 14022011 (date). The protein was loaded in a column. After the sample was fully loaded, the apparatus was started withfilling the column with the buffer in order to wash out unbound specific proteins. The elution of the protein was then performed with the imidazole gradient.
Alternatively peristaltic pump can be used, but in this case all the steps have to be done manually.

4.2.2.6 The GST column and cleavage of fusion protein with the protease P3C

This purification method is based on the high affinity of the GST for glutathione. When applied to the affinity medium, the GST-tagged proteins bind to the glutathione ligand, and then the impurities are removed by washing with binding buffer. We used the 5 ml GST-column which was first equilibrated with the binding buffer (flow rate 5 ml/min). Then the sample and binding buffer (flow rate 5 ml/min) were applied.
Into the column filed containing the sample we injected 5 ml of the P3C enzyme and left the column overnight (approximately 16 hours) in the cooling room. After that we started the elution of our protein from column. Since the protease is fused to GST it is easily removed from the cleavage reaction site by using elution buffer based on glutathione.
All steps can also be done fully manually by washing the columm using peristaltic pump with a flow rate of 5 ml/min.

4.2.2.7 The GSH cleavage

After the protein waseluted from the GST column, a GSH cleavage was done in that column, just to make sure that there is no more protein left in it. For that purpose we prepared the 20 ml of the GSH buffer: 20 mM Tris-HCl, 100 mM GSH, adjusted to pH 8 which was then injected into column with a peristaltic pump.

4.2.2.8 Gel filtration (GF)

Gel filtration chromatography separates proteins, peptides, and oligonucleotids on the basis of their size. Molecules move through a bed of porous beads, diffusing into the beads to greater or lesser degrees, depending on their size. Smaller molecules diffuse further into the pores of the beads and therefore move through the bed more slowly, while the larger ones enter the pores less or not at all and thus move through the bed more quickly. Both molecular weight and three-dimensional shape contribute to the degree of retention. Gel filtration

chromatography may be used for the analysis of molecular size and separation of different components in a mixture, or for salt removal or buffer exchange from a preparation of macromolecules (23). For our experiment we used Superdex 75 (Figure 8).

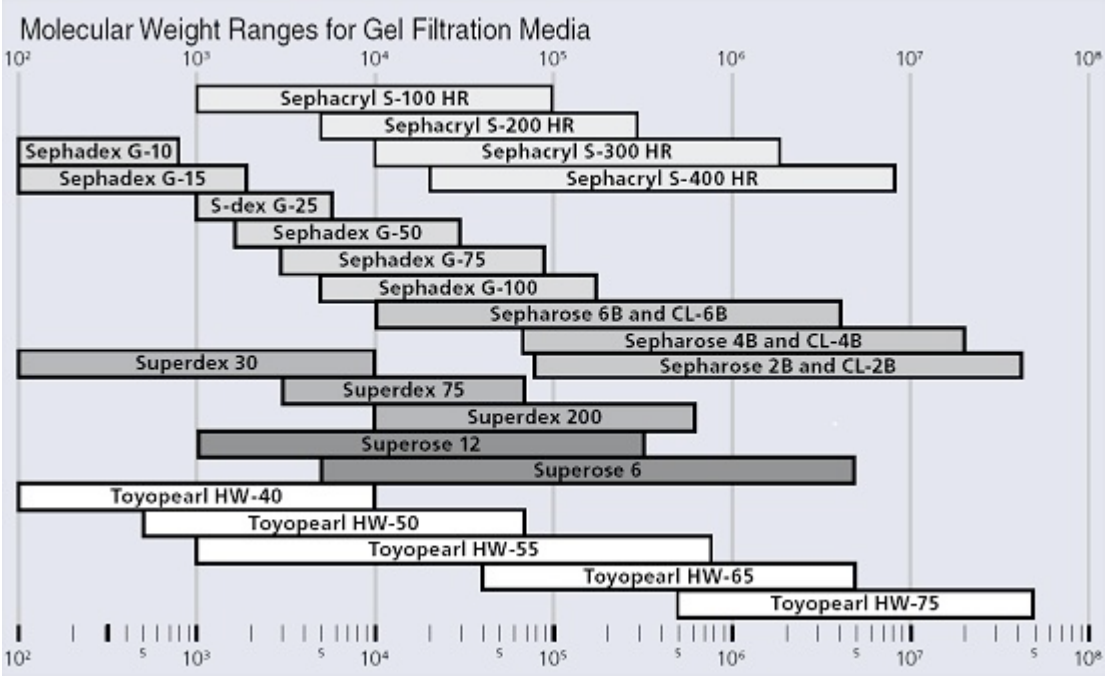

Figure 8: Molecular weight ranges of different gel filtration media.

4.2.2.9 Calibration of the column for gel filtration

To make sure that the column for gel filtration is appropriate, the calibration of the gel filtration process was performed. In the first step the loop was washed with the calibration buffer (100 mM Hepes, 150 mM NaCl, 2 mM beta-mercaptoethanol). Usually, the column was washed with the 3 times the volume of the loop. The loop and the column (S75 26/60 superdex) were then attached to the chromatography apparatus. To check for possible dysfunctions, the volume of the loop was injected into the apparatus. The sample that was used for the calibration of the column is a mixture of 4 components:

Ovalbumin 3 mg, Mw 43.000 Da
Conalbumin 3 mg, Mw 75.000 Da
Ribonuclease A 3 mg, Mw 13.700 Da

Chymotrypsinogen A 3 mg, Mw 25.000 Da

All components were dissolved in buffer (100 mM Hepes, 150 mM NaCl, 2 mM beta-mercaptoethanol). On the display we chose our column, adjusted the parameters and started the run. After the first step of equilibration was finished the sample was transferred to injector and injected into the loop of the apparatus manually after which it was injected in the column automatically. The display commands were set up as follows: Manual→Other→Next breakpoint→Execute.

4.2.2.10 Protein purification by gel filtration (GF)

After the column S75 26/60 was equilibrated we started the protein purification by gel filtration. We used the sample obtained from the GST-column. The GF column was first calibrated with the working buffer. Then the sample was injected manually into the chromatography apparatus.

After purification we decided which eluates to pool for next analyses. If the purification was performed with the Akta apparatus, our decision regarding pooling was based on the graphically displayed results. If it was done by simply using peristaltic pump, it was decided based on the measurments of the MLL5 protein concentration in eluate samples (nano drop spectrophotometer or Bradford method).

4.2.2.11 Concentration of the target protein

The concentration process was the last step in preparing the target protein sample before crystallization. It can be done in a special centrifuge tube by centrifugation or in an amicon cell under pressure.

When the first method was applied, special tubes with 5.000 MWCO (molecular weight cut-off) were usually used. The sample was concentrated by centrifugation at 3.000 x g, for 3 minutes. Between each centrifugation period the protein was resuspended with a pipette. As the proteins are very sensitive and can coagulate easily if there are some air-bubbles present, this step was done with extreme care.

The second method for efficient concentration of the target protein is based on the use of amicon cells and is carried out under pressure. Amicon stirred cells provide high flow rates with solutions containing up to 10% macrosolute concentration. Gentle magnetic stirring minimizes concentration polarization and shear denaturation. After the concentration of the

protein is finished the membrane is washed with the same buffer that was used for gel filtration in order to wash out the protein.

4.2.2.12 Characterization of the target MLL5-3J protein (Ɛ, MW)

The target protein that was purified and isolated was MLL5-3J. The ratios of the molecular weight (MW) and the $\frac{1}{12}$ of the mass of the carbon-12 (^{12}C) were calcualted and represent dimensional numbers [kDa]. These data were then inserted into a nano drop system in order to measure the correct absorbance of concentrated protein solutions. User-selected values for molar extinction coefficient Ɛ (M^{-1} cm^{-1}) and molecular weight (MW) in kilo Daltons for their respective protein reference are shown in Table II.

Table II: Characterization of the target protein

MLL5-3J OPIN J	Fused protein	Cleaved protein
Ɛ ($M^{-1}cm^{-1}$)	55.405	12.295
MW (kDa)	45.582	17.8

4.2.2.13 Separation of the protein using SDS –page

The SDS-page gels were made in order to check the presence and the purity of the target protein. 15% polyacrylamide gels were prepared and used for electrophoresis.

The preparation protocol for 4 gels:

Separating phase

Acrylamid	10.4 ml
Tris 1,5 M, pH 8.8	4.8 ml
H_2O	4.8 ml
SDS 15%	428 µl
Temed	56 µl
APS 25%	40 µl

Temed and APS 25% were added at the end. Once the solution was prepared, it was poured between glass plates and filled with isopropanol to offset the line.

Stacking phase

Acrylamid	1.68 ml
Tris, pH 6.8	2.48 ml
H_2O	5.6 ml
SDS 10%	72 µl
Temed	16 µl
APS 25%	20 µl

Before inserting the stacking phase, isopropanol was completely removed. When stacking phase was inserted, the combs were put on top. After 30 minutes the gels were ready and then stored in a cooling room until used.They were wrapped up in wet towels in order to keep them wet.

To run a gel we put it on a holder and sunk it into TGS running buffer (25 m M Tris, 192 mM, 0.1% SDS, pH 6.8). The samples to be analysed were heated for 5 min at 95 °C. Each loading "pocket" received 20 µl of the sample, while the last one was loaded with 4 µl of the protein size marker (ladder). Electrophoresis was performed at 160 V for first 10 min and then at 200 V. After approximately 45 min the gel was removed, washed in water and then put into the staining colour instant blue for approximately 1 hour. After one hour the gel was transferred onto the light table to determine the presence and the position of our target protein.

4.2.2.14 *Measurement of the target protein concentration (Bradford method and nano drop spectrophotometry at 280nm)*

There are two ways of determining protein concentrations. Normally we have used the nano drop spectrophotometry, but in few cases, where this technique failed, the Bradford method was used.

4.2.2.14.1 THE NANO DROP SPECTROPHOTOMETRY

To measure the protein concentration at 280 nm we used Nanodrop 2000 Theremo Scientific. There is a possibility to enter the molar and mass extinction coefficient of protein by choosing "Other protein (Ɛ & MW)" respectively in order to calculate the concentration automatically from the absorption (see table II). We wiped the sample pedestal with a paper tissue and placed 2 µl blank (the corresponding buffer to the sample). Pressed "Blank" .Drop

was wiped away when the blank was measured and we put 2 µl of the protein sample on the pedestal. Pressed "Measure". There is no need to dilute the protein sample as in normal spectrophotometers. Usually concentrations up to 50 mg/ml can be handled without dilution (24).

4.2.2.14.2 THE BRADFORD METHOD

The colorant (Bio-Rad protein assay) was diluted with water in a proportion of 1/5. For one measurement 1 ml of colorant solution was transferred in each measurement cell (cuvette). Different volumes of the starting solution of 2 mg/ml of BSA (bovine serum albumin) diluted with water were transferred into four different cuvettes (0.5 µl, 1 µl, 2.5 µl, 5 µl) while the fifth was kept free of BSA used as a negative control. Then in another two cuvettes, each containig 1 ml of the colorant and the samples from two different elutions, containing our target protein were transferred subsequently all cuvettes were left in a dark room for 10 min and then the absorption at 595 nm was measured. Based on the results obtained a calibration curve was plotted from which the concentrations of our samples were extrapolated (see Table 5, Figure 25).

4.2.2.15 Circular dichroism (CD)

Circular dichroism is a great tool for rapid determination of the secondary structure and folding properties of proteins that have been obtained using recombinant technologies or being purified from tissues. CD is most widely used to determine whether an expressed, purified protein is folded, or if a mutation affects its conformation or stability. In addition, it can also be applied to study protein interactions. The following protocol details the basic steps for obtaining and interpreting CD data, as well as explains the methods for analyzing spectra in order to estimate the secondary structural composition of proteins. CD has the advantage that the measurements of multiple samples containing 20 µg or less protein in physiological buffers can be performed already infew hours. However, this technique does not give the residue-specific information that can be obtained by X-ray crystallography or NMR (25).

4.2.2.16 Dynamic Light Scattering (DLS)

Dynamic light scattering, known also as photo correlation spectroscopy, is a technique which can be used for determination of small particle sizes. When small particles are hit by light, the light beam scatters in all directions provided that the particles are small compared to

the wavelength applied, i.e. below 250 nm. The distances between the scatters in the solution are constantly changing with time due to the Brownian motion of small particles. The scattered light undergoes constructive or deconstructive interference, which is a surrounding process. The information about the time scale of the scatter movements is contained within the measured intensity of fluctuation (26).

4.2.3 CRYSTALLIZATION OF THE TARGET PROTEIN

The phase diagram of protein crystallization is a schematic representation of how the protein and its precipitate concentration are related. Protein crystals are formed only in supersaturated solutions. As shown below, in the Figure 9, low protein and/or precipitate concentrations will cause undersaturation and consequently no crystal formation will be induced.

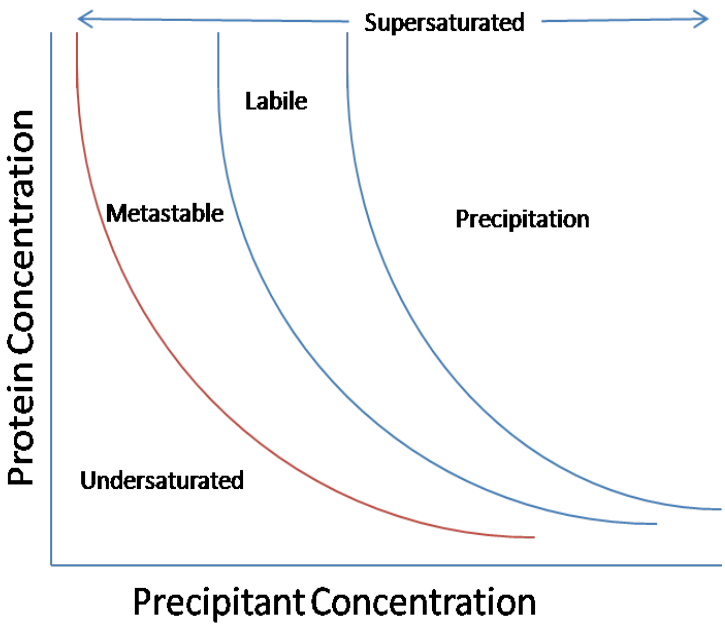

Figure 9: A schematic graphical display of correlations between protein and precipitant concentrations (31).

The red curve in Figure 9 that separates undersaturated conditions from the supersaturated ones is known as the solubility curve. A clear benefit of determining this curve is that it can help to guide us when analyzing particular protein crystal growth conditions. Crystallization setup that is undersaturated or in the metastable phase will appear clear, however, the latter

phase has the potential of crystal growth if seeded (spiked) with crystals. Precipitation occurs when the protein comes out of solution as an aggregate, but as such it is not suitable for crystallographic studies. The labile zone (or nucleation zone) is important since this is the point where the crystal nucleation and initial growth occurs. As the crystals form, the protein concentration will be depleted causing the move from the labile to metastable zone (27).

4.2.3.1 Protein crystallization

The goal of crystallization is to produce crystals, which should be well-ordered, lacking contaminants and large enough to provide a diffraction pattern when hit by x-rays. Usually this diffraction pattern can be used for determining the protein's three-dimensional structure. The procedure of protein crystallization is inherently difficult because of the fragile nature of protein crystals. Namely, the surfaces of proteins are irregularly shaped, which results in the formation of large channels within any protein crystal. Therefore, the non-covalent bonds that hold together the lattice must often be formed through several layers of solvent molecules. In addition to overcoming the inherent fragility of protein crystals, the successful production of x-ray analysis worthy crystals is dependent upon a number of environmental factors. Because so much variation exists among proteins, each individual one requires unique conditions for successful crystallization. Therefore, attempting to crystallize a protein without a realiable and highly optimized protocol can be very tedious. Some of the factors that require consideration are protein purity, pH, protein concentration, temperature and addition of precipitants. For sufficient homogeneity, usually the protein should be at least 97% pure. The pH conditions are also very important, as different pH values can result in different packing orientations. Therefore high capacity buffers, such as Tris-HCl, are often necessary to maintain a particular pH value. Precipitants, such as ammonium sulphate or polyethylene glycol, are compounds that cause the protein to precipitate out of solution (28, 29).

The optimal concentration range for our purified protein MLL5-3J is around 10 mg/ml. When the protein was isolated, the preparation of the sample for the crystallography was started. The first step was the centrifugation at 4 °C at 14.000 rpm for 10 minutes. The sample was then divided in 3 tubes. In the first tube about 100 µl of the sample was transferred and it was stored in a freezer at -70 °C. The second and the third tube were each filled with a half of the remaining volume of the protein solution and in the third one, SAM (S-adenosyl-L-methionine) was added in order to promote the crystal growth.

The crystallization was performed by using two different robots, i.e. the Tecan genesis RSP 100 and the Cartesian dispensing system. First the water container was washed with water in order to clean all the loops in the apparatus. Then several media were used (a list of Amex media: phclear, anion, clear strategy 1, clear strategy 2, clear strategy 3, clear strategy 4, compass, PEG2, JSCG+, JSCG 1, JSCG 2, JSCG 3 and JSCG 4).

Those have been kept in a cooling room. They were obligatory centrifuged before being used. The selected medium was placed into the holder 1 of the Tecan robot and a new empty plate (Corning 1:4, 96 well) was inserted into holder 2. The apparatus divided the applied media automatically by transfering 70 µl aliquots in each well. The microtiter plate was labeled and the medium that was not needed any more was placed back into the cooling room. In-between the the Cartesian dispensing apparatus was prepared. The bottle containing water was changed for a new one which was first sonicated for 5 minutes. The bottle was attached to the apparatus carefully so that there were no air bubbles present and the apparatus was washed. A pre-prepared plate containig appropriate medium was put in the apparatus (Cartresian dispensing) on the position 2 and the plate containing two MLL5-3J protein containing samples, namely 10 µl of sole protein in positions 12B and 12F, and 10 µl of protein supplemented with with SAM (S-adenosyl-L-methionine) in positions 11B and 11F. The apparatus then dispersed 100 nl drops containing our target protein and 100 nl drops of medium, resulting in different concentrations of the MLL5-3J. When the final preparation of the microplate plate was finished, a plastic cover was put on it, because drops are very small and can dry easily if not protected. The observation was done under the light microscope and the results were noted.

4.2.4 OPTIMISING OF CRYSTALLIZATION BUFFER CONDITIONS

In order to improve the whole crystallization porcedure we decided to optimise experimental conditions by applying several different buffers to our protein sample. Before we continued with analysis of the protein we decided to screen for the optimal conditions created by different buffers by DLS. Therefore we have prepared 22 different 100 mM buffers, in volumes of 15 ml, with pH values ranging from 3 to 10 (Jancarik et al. 2004). By testing them we tried to to choose the most appropriate one that would maximize the chance for appropriate homogeneity and crystallization of our target protein sample.

Table III: 22 buffers for screening write by Jancarik et al.

#	**Buffer (100 mM)**	**pH**		#	**Buffer (100 mM)**	**pH**
1	Glycine	3		12	Cacodylate	6.5
2	Citric acid	3.2		13	Ammonium acetate	7
3	Citric acid	4		14	MOPS	7
4	Sodium acetate	4.5		15	Sodium phosphate	7
5	Sodium phosphate	5		16	HEPES	7.5
6	Sodium citrate	5.5		17	Tris HCl	7.5
7	Sodium phosphate	6		18	Imidazole	8
8	Bis-tris	6		19	Tris Hcl	8
9	MES	6.2		20	CHES	9
10	ADA	6.5		21	CHES	9.5
11	Bis-tris-propane	6.5		22	CAPS	10

We prepared crystallization plates, 15-Well type EasyXtal and the EasyXtal 15-Well tools were used for the manual setup of hanging-drop protein crystallization trials. The plates have 15 wells, each with a capacity of 500 µl. Their optical properties and SBS footprint format (rack robotics friendly) make them suitable for automatic visualization systems. The greaseless screw-in crystallization supports are easy to open and close, therefore facilitating crystallization screening and variations of crystallization conditions (30).

In each well 500 µl of a tested buffer was transfered. On the bottom side of the greaseless screw-in crystallization support 3 drops of the same buffer corresponding to the one being present in the actual well were transferred. Each of these drops has a volume of 1 µl. On the other side of the glass, the deposited drops were marked, as follows:

Ø - no spot, meaning MLL5-3J 10 mg/ml,

· - one spot, meaning MLL5-3J 10 mg/ml with SAM and

·· - two spots, meaning MLL5-3J 25 mg/ml.

Susequently 1µl of appropriate protein-containig sample was injected into each pre-signed buffer drop. The wells containing each buffer tested were marked accordingly as well. The plates were left at room temperature for 24 hours. After that the drops were checked under the light microscope and the notes, as presented below, were taken while observing whether the

drops were clear or not. The sign / meant that there were no clear drops and the other signs were applied as follows:

Ø – a clear drop containig MLL5-3J 10 mg/ml;

• - a clear drop containig MLL5-3J 10 mg/ml with SAM;

•• - a clear drop containg MLL5-3J 25 mg/ml.

All the buffers where the drops were clear were further screened with DLS. Fot this purpose a mixture of 135 µl of each particular buffer and 15 µl of the MLL5 protein sample. We began the screening with a sample containing 10 mg/ml MLL5-3J, where we observed the lowest peak and PDI. After all the samples were screened, the screening was repeated with addition of 15 µl of 5 M NaCl, to see whether there was any effect of a high salt concentration.

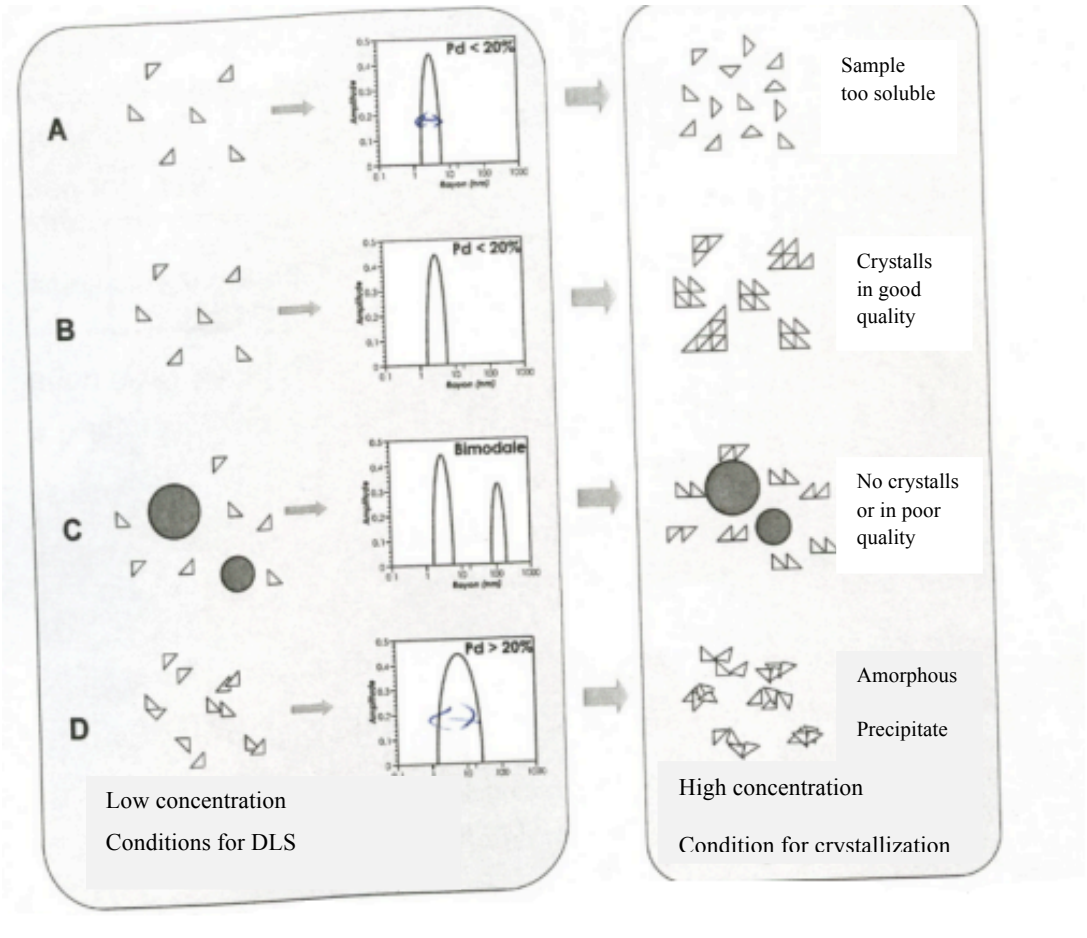

Figure 10: Different conditions that can be used for protein crystallization.

5 RESULTS AND DISCUSSION

5.1 RESULTS

5.1.1 EQUILIABRATION OF THE SUPERDEX 75 COLUMN

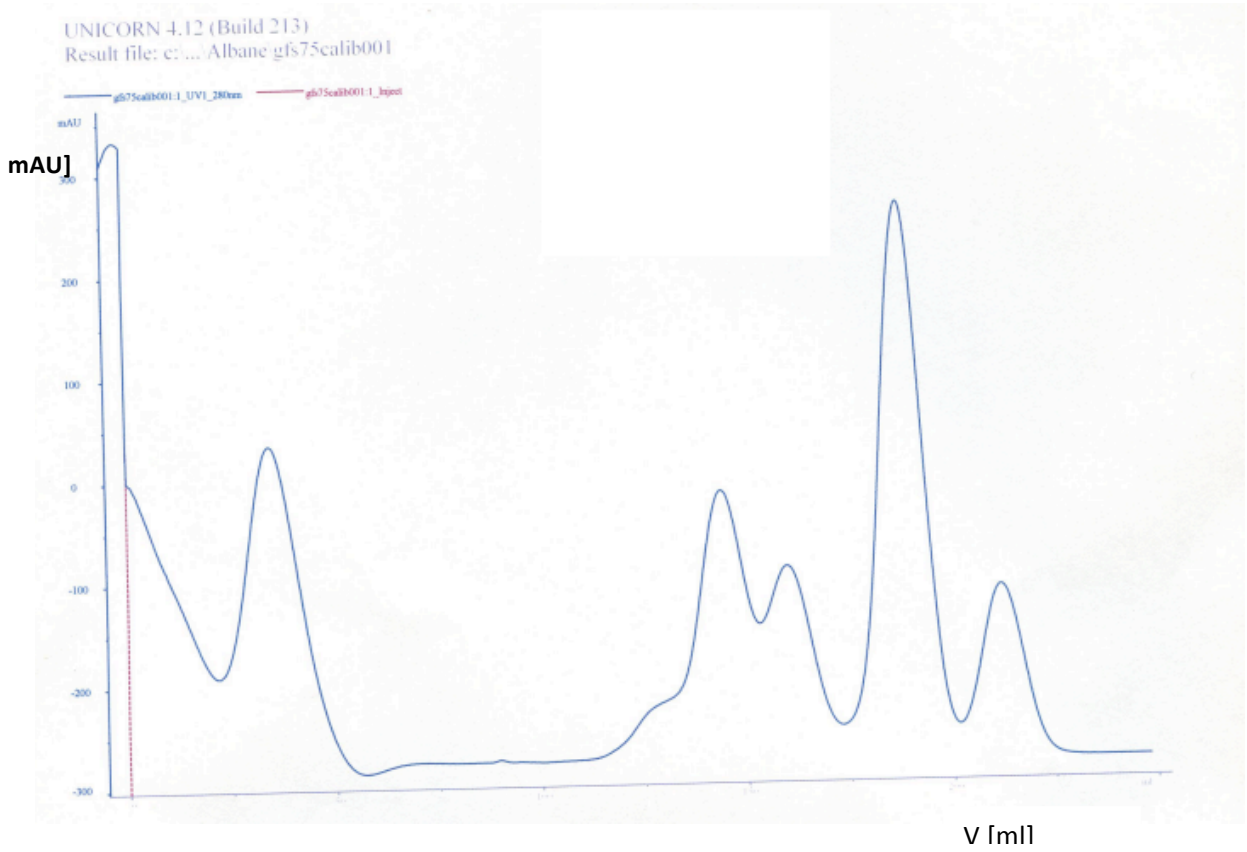

Figure 11: The equilibration curve of the Superdex 75; conditions: FR (Flow rate) 2.5 ml/min, detector: UV detector.

The equilibration of the Superdex 75 column was necessary to achieve a better separation of our targeted MLL5-3J protein by GF. The equilibration result is presented graphically in Figure 11, showing expected size-based distribution of peaks representing eluted control proteins.

5.1.2 THE 1ST TRIAL MLL5-3J (ROSETTA, E. coli STRAIN) WITH 2 x 700 ml OF LB MEDIUM (*Cell culture growth in LB, the IPTG induction*)

For the first trial in a row of our experiments we used frozen sample, containing MLL5-3J recombinant protein, expressed in Rrosetta cells (a strain of E. coli) and induced by IPTG. Two cryotubes were thawed, each containing harvested cell pallets from the 700 ml LB culture. The contents of both were combined and after sonication and filtration the sample was purified by the HisTrap chromatography using Ni-column.

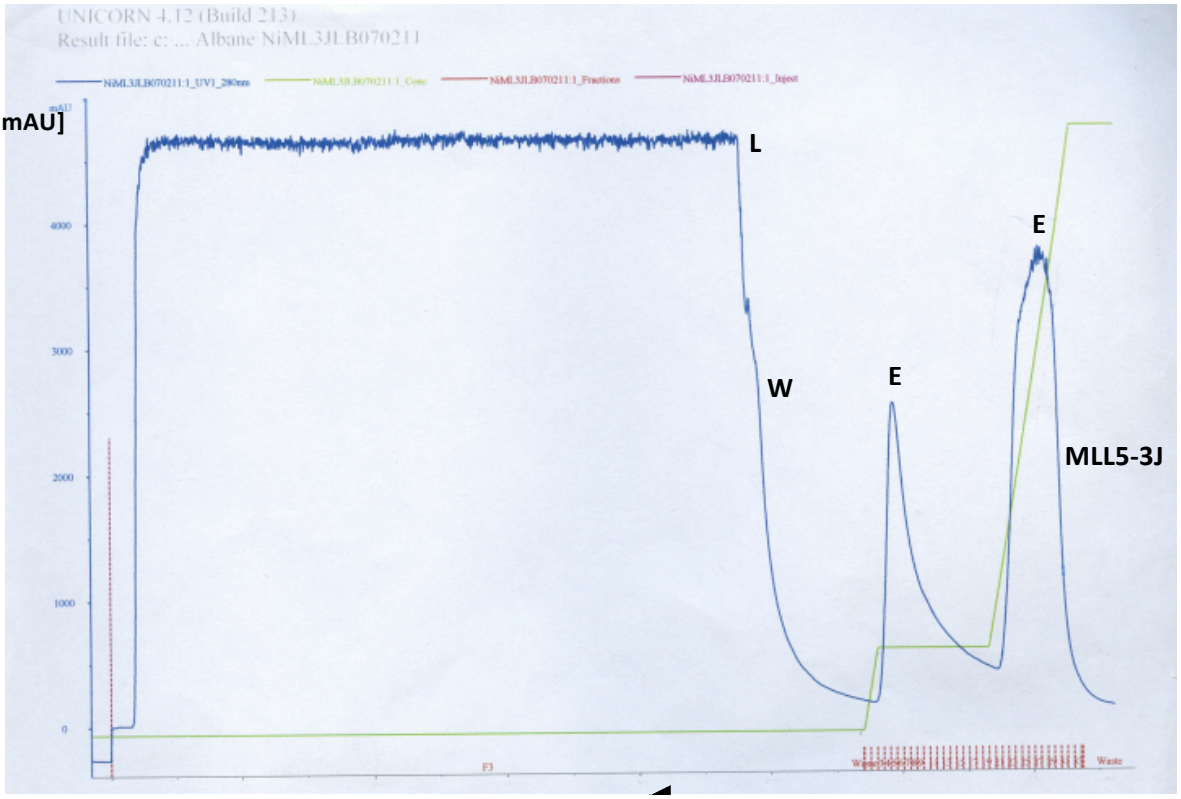

Figure 12: The 1st trial, HisTrap MLL5-3J purification; Ni column, LB medium; conditions:FR 0.5 ml/min, start conc. 4%, volume of the sample 2000 ml, wash-out 20 ml, detector: UV detector.

In Figure 12 a typical HisTag chromatograph is presented, showing all three procedural phases, i.e. loading (L), washing (W) and elution (E). The 2nd peak represents our target protein (MLL5-3J).

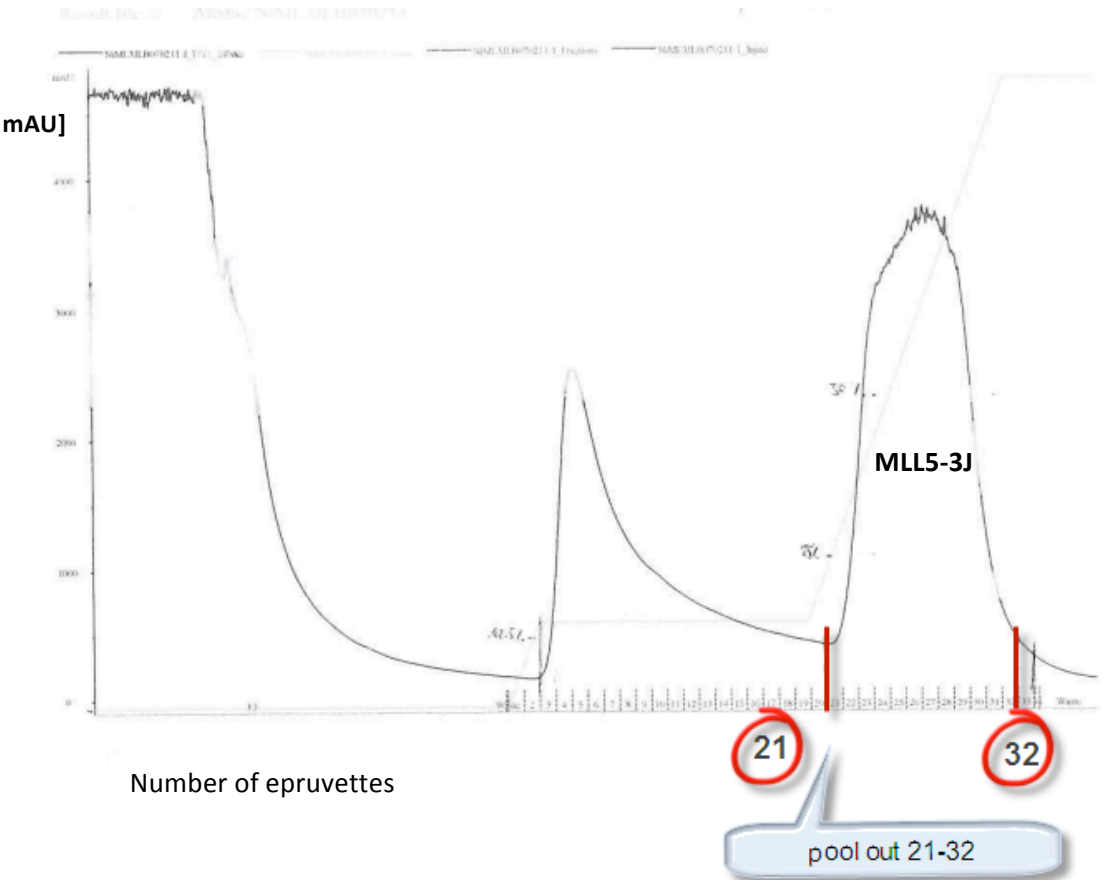

Figure 13: Same as Figure 12, but with the eluted MLL5-3J peak zoomed in.

On the graph we can specify the area of the peak. We observed that protein has been eluted between fractions 21 and 32. Therefore we decided to pool them, as marked on the cromatograph. The total volume of pooled eluted samples was 12 ml and contained 3.4 mg/ml of MLL5, as measured by nono drop (Ɛ =55.405, MW=45.582). So, alltogether we finished the HisTrap procedure with 41 mg of the MLL5-3J, which means that we had a massive overexpresion of the recombinant protein in our production Rosetta cultures. Anyhow, we do not know what has been the reason for this. As seen from the graph, there was also a nonspecific peak (1st peak) present, eluted between the fractions 5 and 10. We assumed that this purification step may have not been very efficient. After the HisTrap separation on the Ni-column was finished, we decided to run an SDS-page gel. For that purpose we have sampled the following eluted fractions or flow throughs (FTs): 5, 10, 23, 25, 28 and 31.

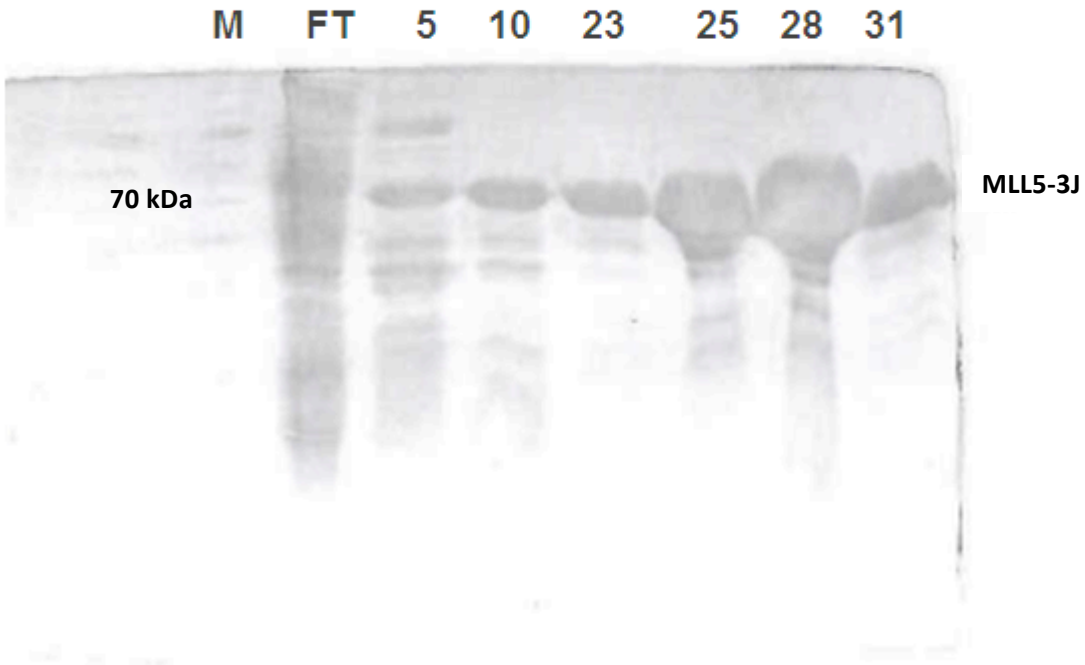

Figure 14: The 1st trial, the 1st SDS-page electrophoresis of selected HisTrap eluate fractions or flow throughs (FTs).

By running SDS-page gels we defined the size of the MLL5-3J protein present in selected eluates. The protein was spotted as dark lines with a size near 70 kDa. We saw that the fractions 5 and 10 still contain some of the target protein, showing that the purification procedure was not fully efficient. Therefore we decided to proceed with additional purification step, by using the GST-column. We loaded the sample on the column, washed the unbound proteins and added the P3C protease. The size of the target protein, following this step was 34 kDa. The GST column was eluted with 24 ml of elution buffer. The MLL5-3J protein was chased in four different tubes, each containing 6 ml of eluate. After measuring the concentration of MLL5-3J in each eluted fraction we decided to conutinue the purification with gel filtration, but only taking the first two eluates (tubes), that contained the gros quantity of protein.

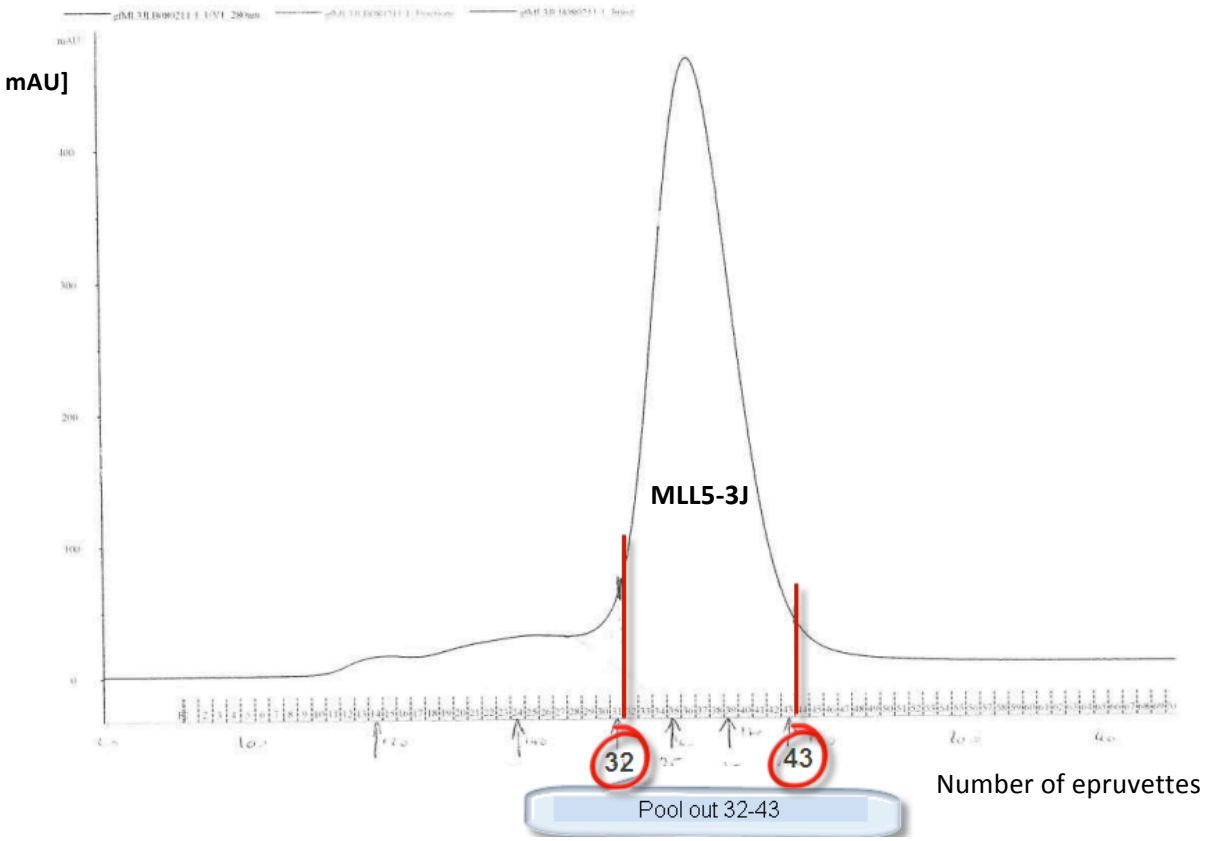

Figure 15: The 1st trial, gel filtration (GF) of the MLL5-3J protein containing sample obtained after GST purification step; LB medium; conditions: FR 0.5ml/min, detector: UV detector.

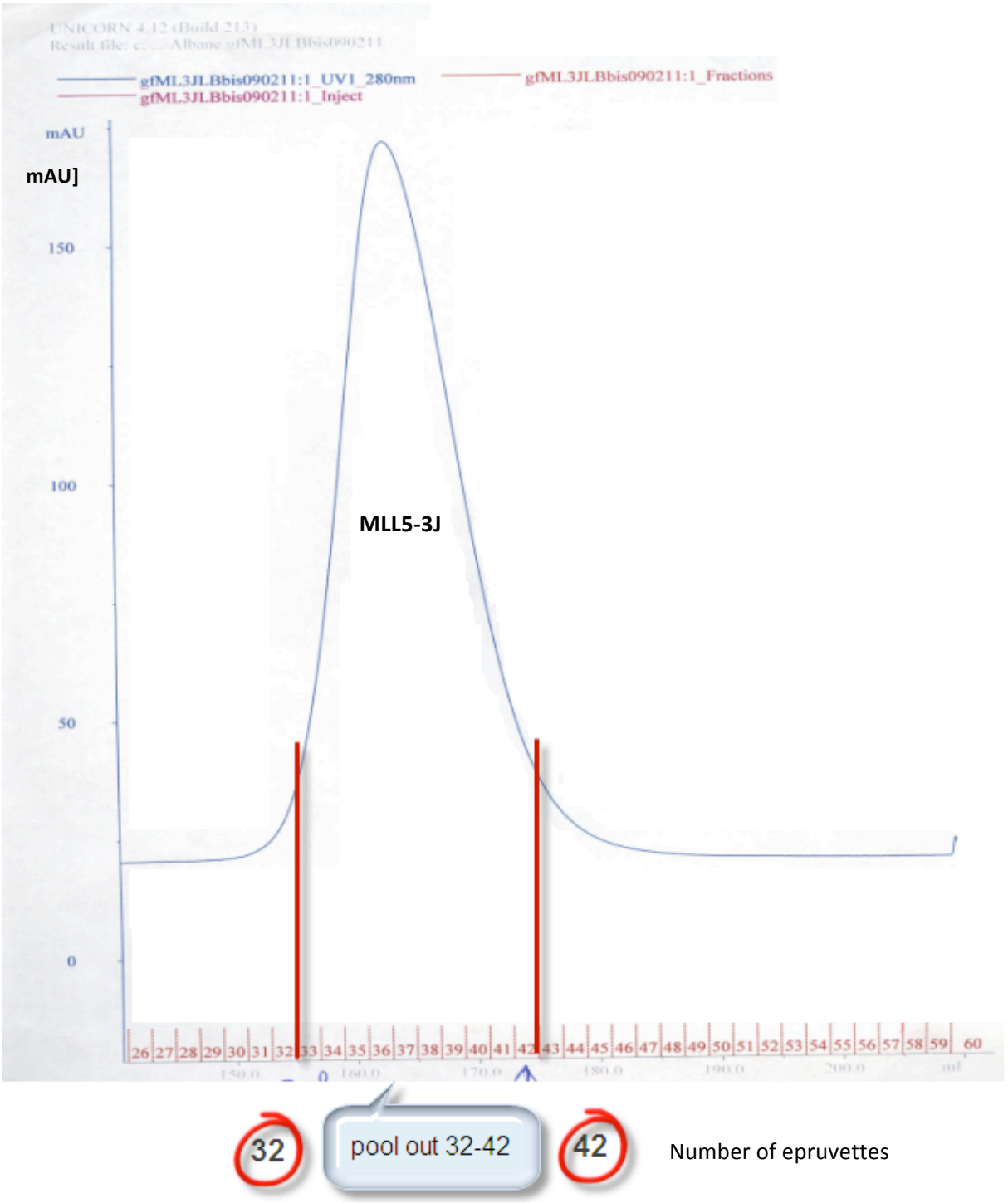

Figure 16: Same as Figure 15 but with the eluted MLL5-3J peak zoomed in.

The results of the GF were quite promising. We succeeded to obtain a symmetric peak, but it was too high. Based on these results we decided to pool the eluted fractions from 32 to 43. The total pooled volume was evaluated by DLS analysis, showing that the size of the

eluted protein was 34 kDa. However this was not we have expected as after GF, the size of the MLL5-3J protein should be 17.8 kDa. So we assumed that we got its dimer. The concentration of the eluted sample was 10 mg/ml and a total of 6.18 mg were obtained. In order to confirm the protein size we decided to run the SDS-page gel electrophoresis.

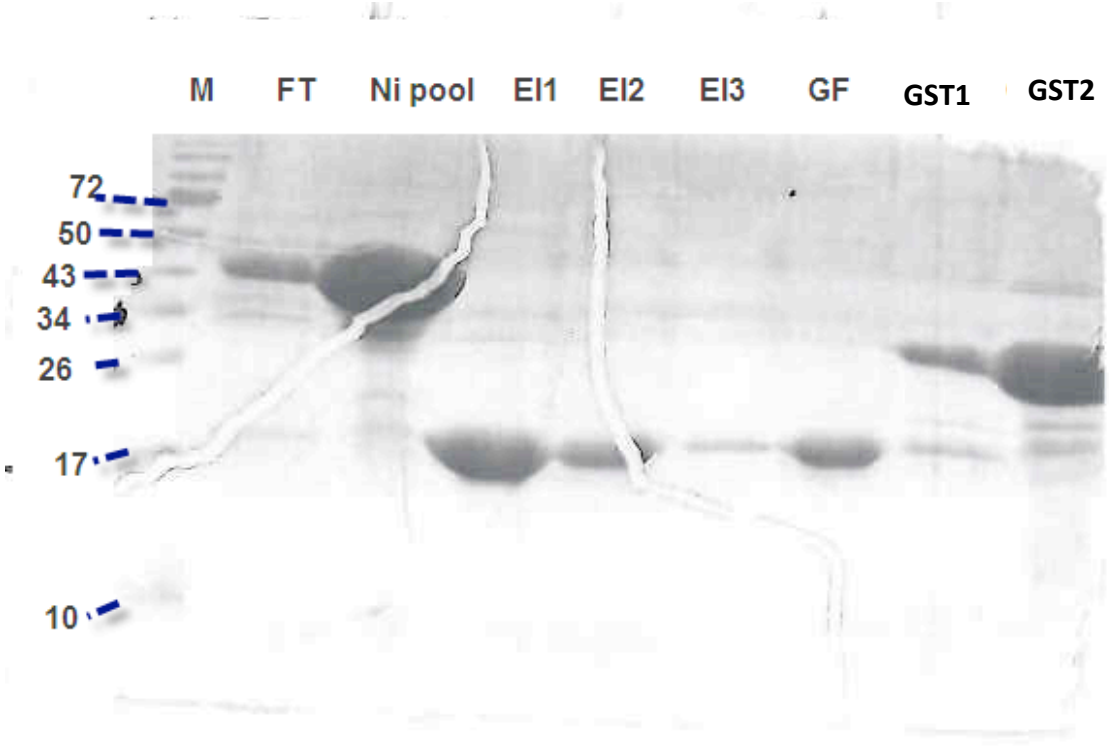

Figure 17: The 1st trial, the 2nd SDS-page gel of selected FT, Ni-pool, GST and GF eluate fractions.

The SDS-page gel electrophoresis shown in Figure 17 was used for evaluation of the MLL5-3J protein in different samples: FT; Ni-pool; El1, El2 and El3 from GF; El1 (GST1) and El2 (GST2) from GSTand GF (before crystalization). As we can see the eluates El1 and El2 obtained by GF contain a lot of protein. We measured its concentrations in both these eluates by using nano drop (Ɛ=12.295, MW=17.8 kDa) and found that the MLL5-3J protein concentration was 0,27 mg/ml and that the total amount of the protein in a 24 ml of the sample was 6,48 mg. The protein's final size was 17 kDa. So according to these results we considered that the gel filtration was successful. We also made the CD (circular dichroism analysis). The protein was mainly in a beta sheet configuration, what is expected for the SET domain. Before crystallization we concentrated the sample by centrifugation, using a filtration

tube. The sample was concentrated to the volume of 600 µl. The final concentration was 10.3 mg of MLL5-3J/ml. Subsequently we proceeded with crystallization but without success, as we were not able to obtainany kind of crystal form, not even the amorphic one. We couldn't find the reason for that, so we can only asume that the protein isolation and purification part was not efficient enough.

5.1.3 THE 2ND TRIAL – MLL-3J (ROSETTA E. COLI STRAIN) 2 X 700 ML OF ONEX MEDIUM (*Cell culture growth in LB; onex autoinduction*)

In the second trial of our experiment we used frozen samples containing the recombinant MLL-3J protein, expressed in Rosetta cells. This time the cell culture was not induced with the IPTG, but with the OnEx autoinduction system. Two cryotubes were thawed, each of them containing harvested cell pellets from a 700 ml OnEx culture. We combined the contents of both tubes and after the sonication and filtration we started the target protein isolation with HisTrap chromatography, using the Ni-column.

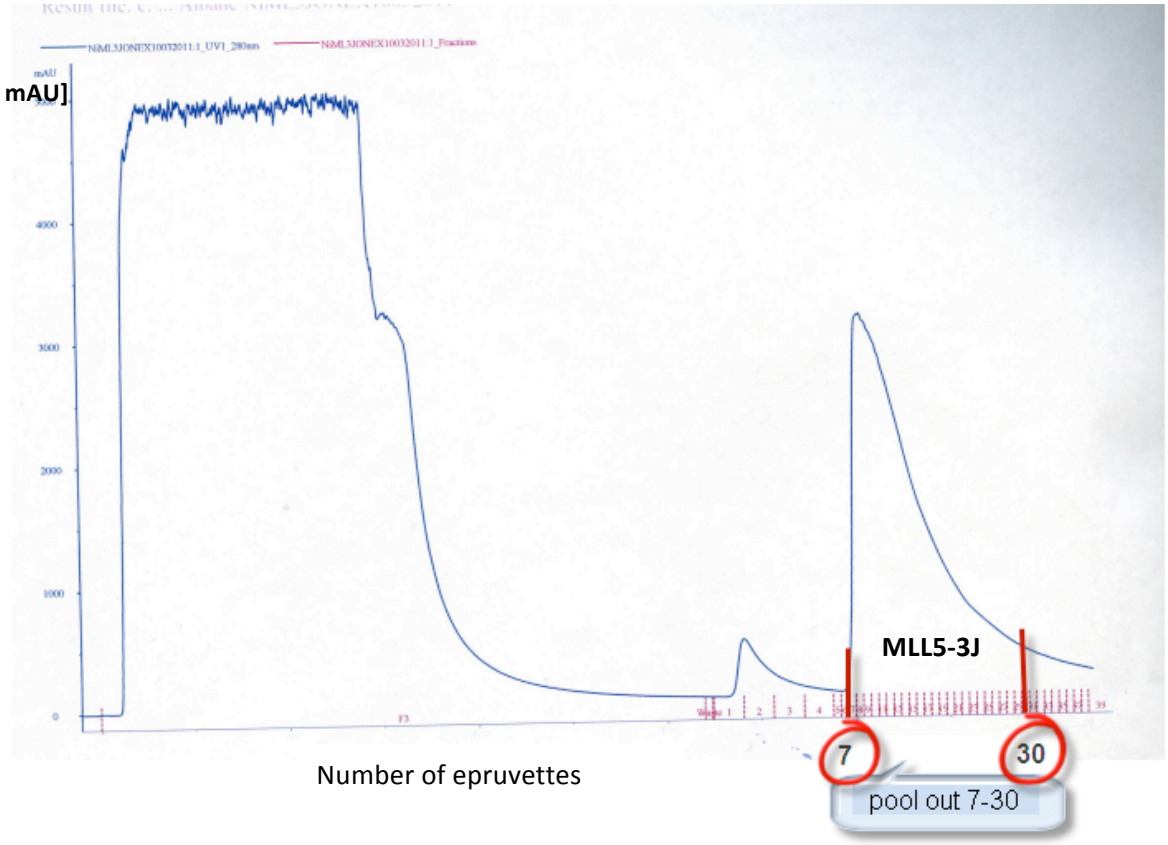

Figure 18: The 2nd trial, HisTrap MLL5-3J purification, Ni column, OnEx medium; conditions: FR (flow rate) 0.5 ml/min, start conc. 4%, volume of the sample 2000 ml, wash-out 20 ml, detector: UV detector.

The elution of the MLL-3J protein from the column was carried out using imidazol gradient. The graph presented in Figure 18 shows the result of the HisTrap chromatography using the Ni column. The eluted peak of our target protein is not proper. As marked we decided to pool the eluted fractions from 7 to 30 for gel filtration. The final volume of pooled fractions was 48 ml. We measured the concentration of 1.35 mg MLL-3J/ml by nano drop (Ɛ=55.405, MW=45.582), meaning that around 65 mg of the protein in total were obtained. As in the 1st trial we performed the GST column chromatography with the P3C protease cleavage. For subsequent GF we also used the first two eluates from the GST-column. What was different in comparison to the 1st trial was the fact that a different buffer was used for the gel filtration step, consisting of 100 mM Hepes, 150 mM NaCl and 2 mM beta-mercaptoethanol, pH 7.

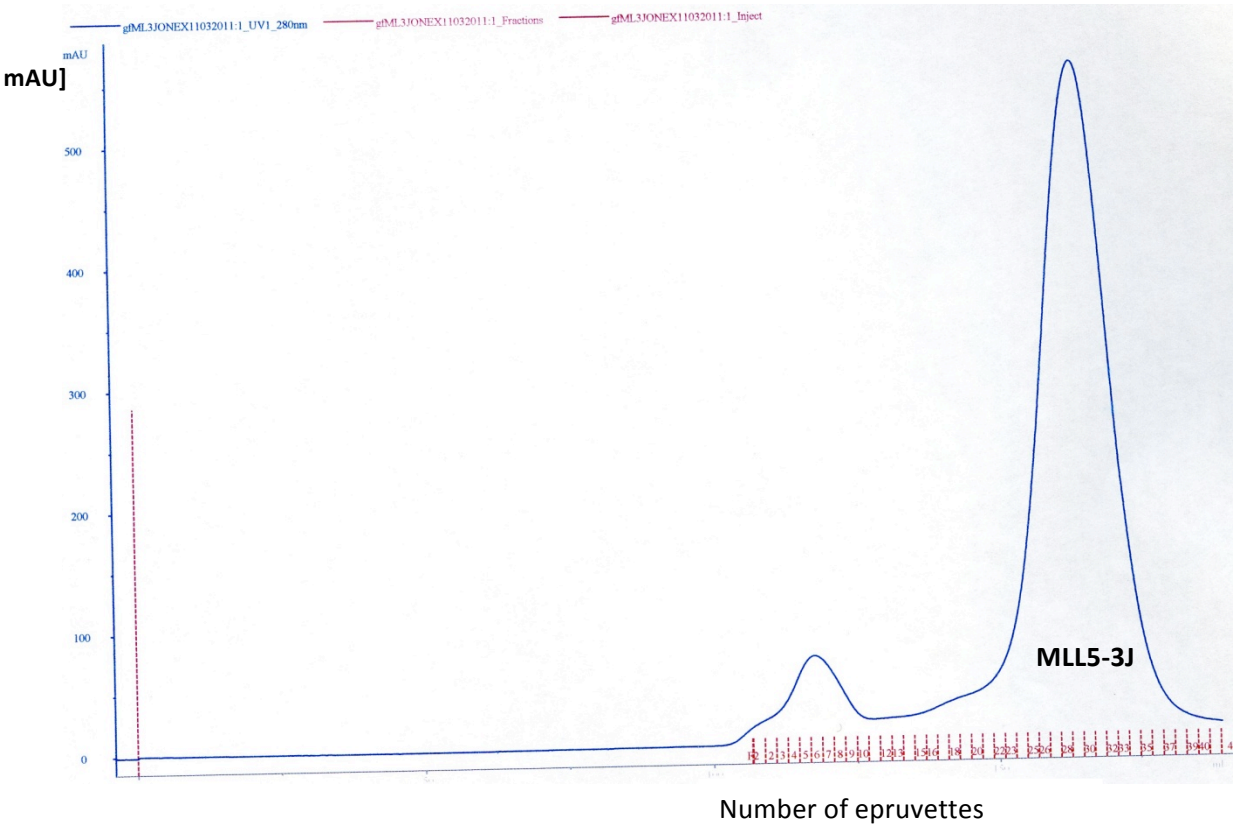

Figure 19: The 2nd trial, gel filtration (GF) of the MLL5-3J protein containing sample obtained after GST purification step; OnEx medium; conditions: FR 0.5 ml/min, detector: UV detector.

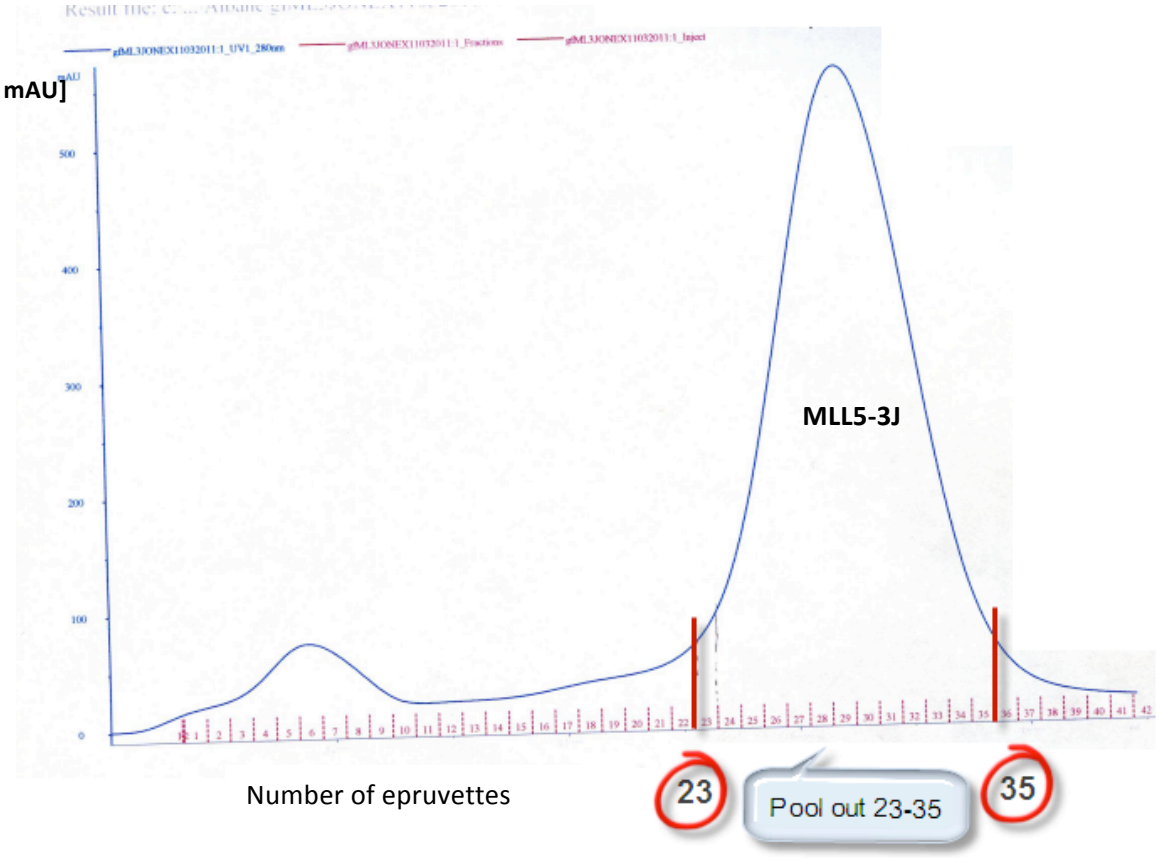

Figure 20: Same as Figure 19 but with the eluted MLL5-3J peak zoomed in and marked eluates that were pooled.

The chromatograph presented in Figure 20 shows a symmetric peak for MLL5-3J, with a proper shape which is not too big and not too small. We followed the same protocol that was used for the first trial. After being eluted, the fractions from 23 to 35 were pooled and the protein concentration of 0.46 mg/ml was defined by using nano drop (Ɛ=12.295, MW=17.8 kDa). The total volume of the pooled sample was 12 ml. Before the crystallization step we concentrated the sample to a volume of 1 ml by using filtration tube and centrifugation. It contained 10.8 mg of MLL5-3J/ml. As in the first trial, we were again unsuccessful in preparing crystals suitable for further analysis.

5.1.4 THE 3RD TRIAL MLL-3J (ROSETA, E. coli STRAIN) 3 x 700 ml OF LB MEDIUM

(*Cell culture growth in LB, pre-culture with addition of $ZnSO_4$, IPTG induction*)

In a third trial we used frozen sample containing MLL5-3J protein which was expressed in Rosetta cells, induced by IPTG. The pre-culture was prepared in the presence of $ZnSO_4$. Three cryotubes containing harvested cell pellets from 700 ml LB culture were thawed and their contents pooled. After sonication, DNAse I, together with its activator $MgCl_2$ was added in order to cut the DNA present in a sample into smaller pieces, to prevent its potential interference with subsequent analyses.

We proceeded with the HisTrap separation by loading the sample on the Ni-column using peristaltic pump with a flow rate of 0.75 ml/min. The washing step was performed manually, using 25 ml of PBS buffer with a flow rate of 5 ml/min. The second loading on the column was effectuated with the same flow rate and by applying the same conditions except that for the subsequent washing 50 ml of PBS buffer were used. We continued the washing with 50 ml of buffer containg 50 M Tris, 150 mM NaCl, pH 7.5, in order to prepare the column for the GST chromatography. After adding the P3C enzyme mix we proceeded with the elution. Measurements of the concentration of MLL5-3J in the first two eluates (El1 and El2) were performed by the Bradford method. Both eluted samples in a total volume of 12 ml were then used for GF.

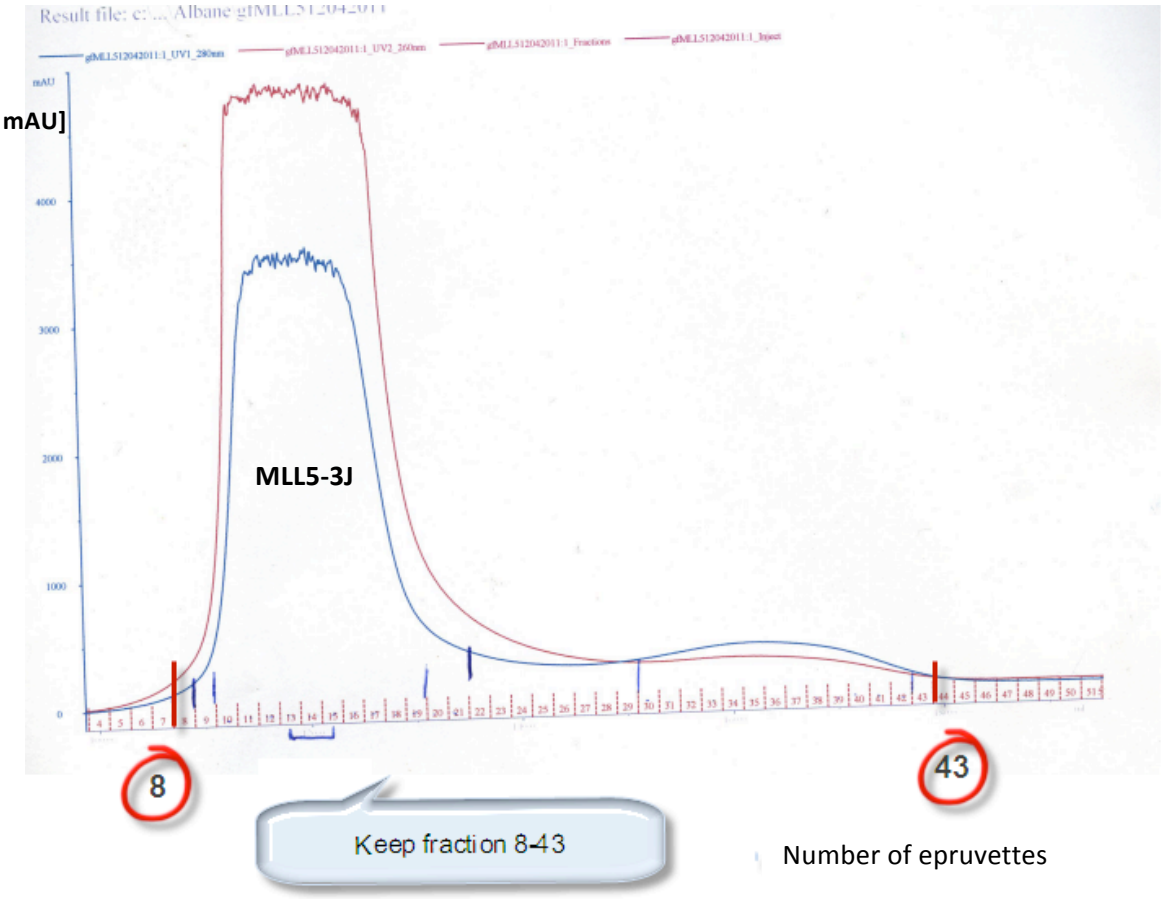

Figure 21: The 3rd Trial, gel filtration (GF) of the MLL5-3J protein, LB medium; conditions: FR 0.5 ml/min, detector: UV detector.

As it can be seen in Figure 21, the GF purification step was not so efficient as expected. Obviously the separation was not very efficient as the eluted MLL5-3J peak was too high to be properly sensed by detector. We decided to keep the eluted fractions from 8 to 43, as marked in the graph (Figure 21). The fractions were pooled, so the final volume containing our target protein was 70 ml. The pooled sample was concentrated in amicon cells under pressure, until its volume was reduced to 10 ml. Then it was reconstituted to original volume of 70 ml with dissociation bufer, because we assumed that there were multimers of protein present. The gel filtration of the reconstituted sample was then performed once again by using the GF apparatus.

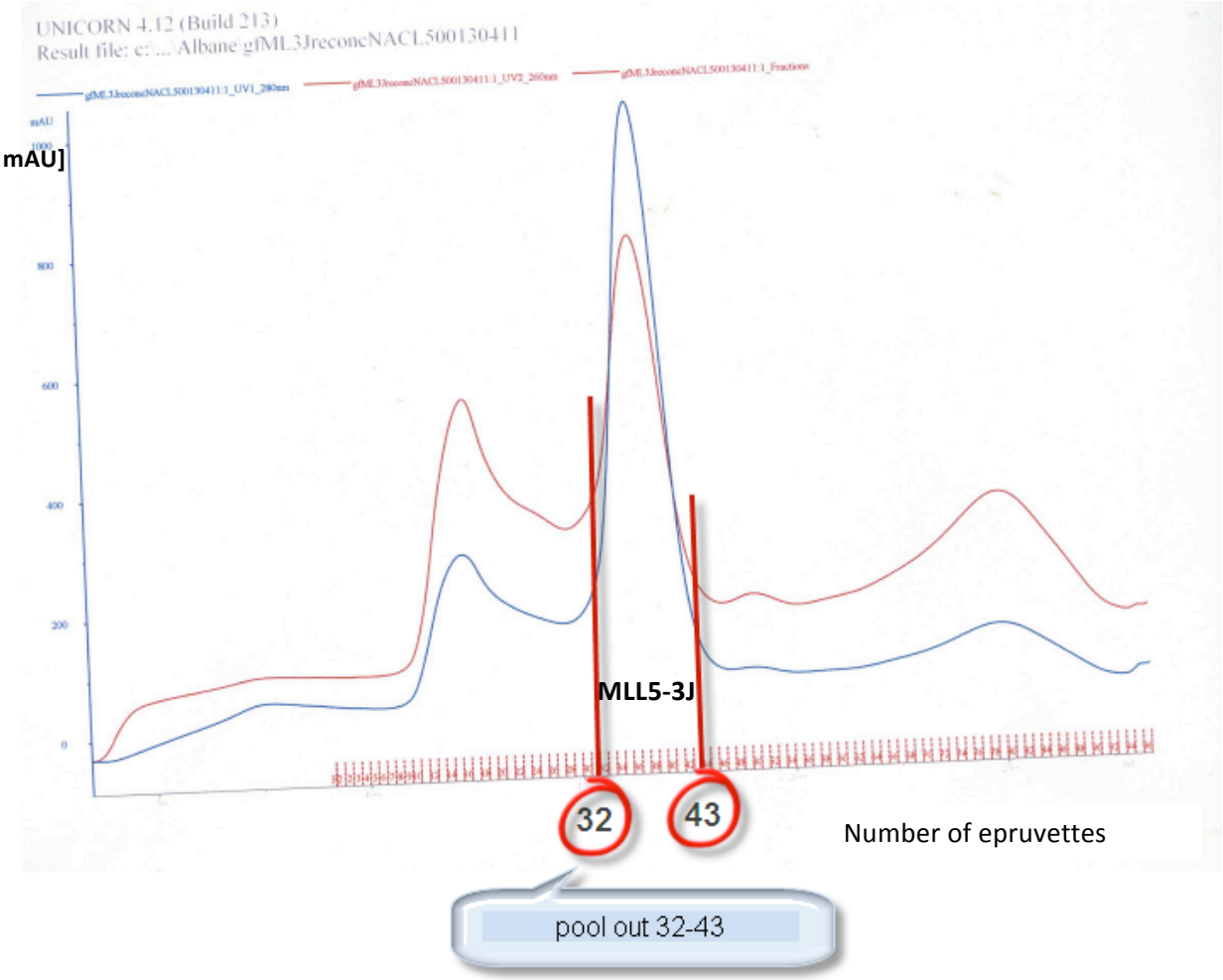

Figure 22: The 3rd trial, additional gel filtration (GF) of the MLL5-3J protein, LB medium; conditions: FR 0.5 ml/min, detector: UV detector.

The eluted peak of MLL5-3J was once again not as expected to be, but anyhow much better than after the first GF. We assumed that the target protein has been eluted between fractions 32 and 43. Therefore we have pooled them, which resulted in a total volume of 25 ml and measured the MLL5-3J concentration by nano drop (Ɛ=12.3, MW=17.8). The pooled sample contained 1.089 mg of MLL5-3J /ml (27,225 mg in total). Subsequently we concentrated the sample to the final volume of 1.8 ml, containing 10.4 mg of MLL5-3J /ml.

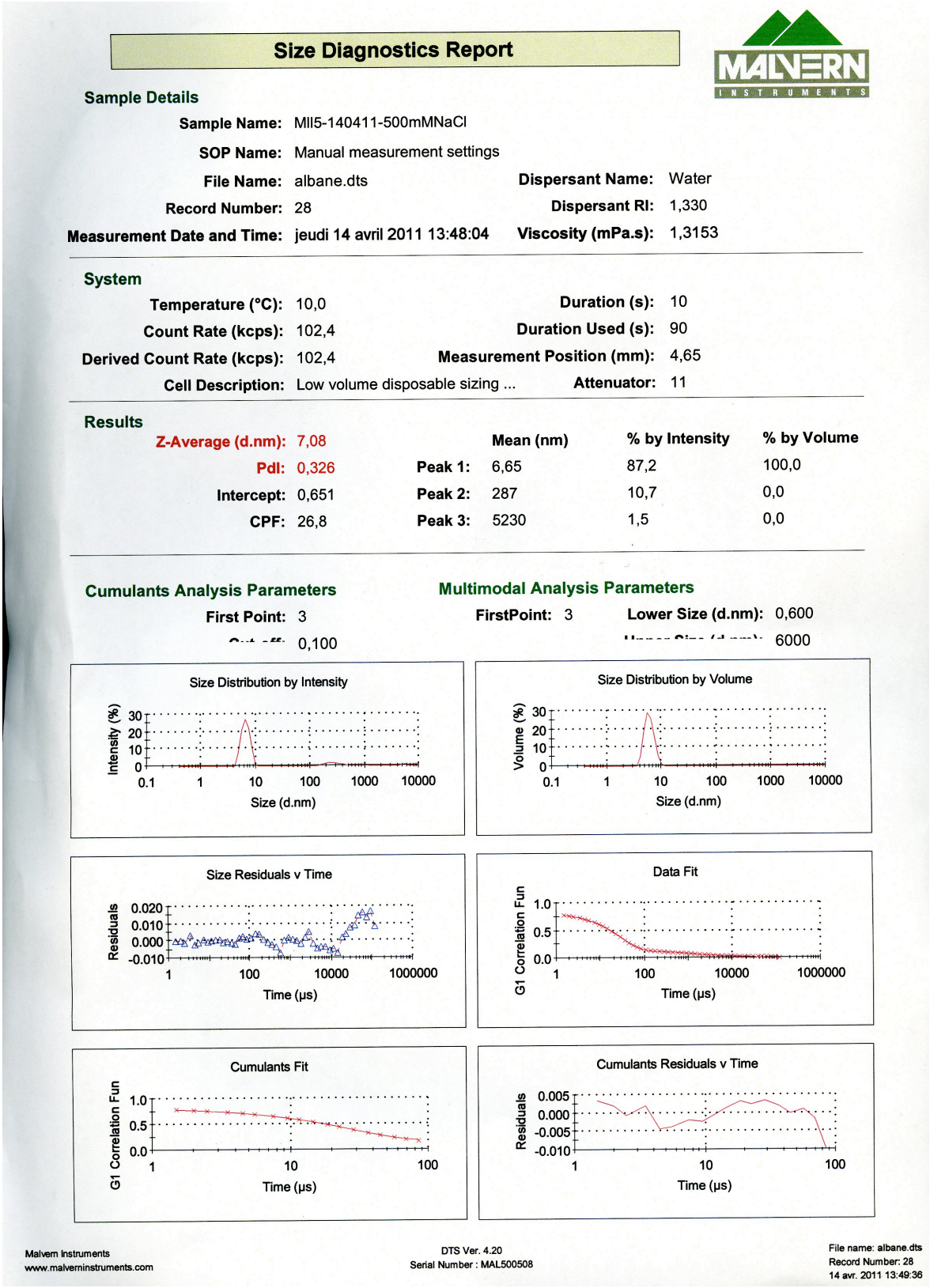

Figure 23: The 3rd trial, following the lastGF: the results of the DLS analysis of MLL5-3J (LB medium).

In Figure 23 the results of dynamic light scattering (DLS) analysis are presented. Our sample contained a protein (MLL5-3J) with a Z-average value of 7,08, meaning that there were no multimers present. Therefore we could continue with the crystallization step.

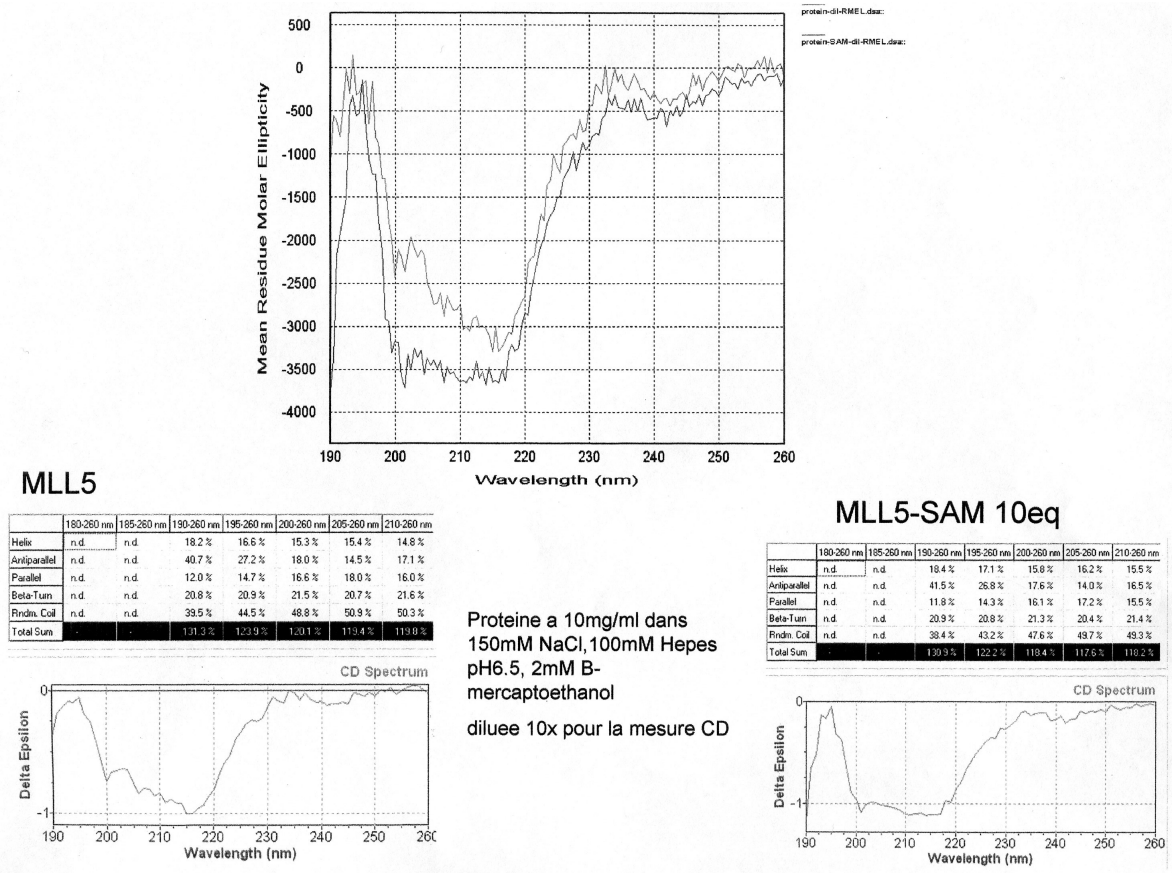

Figure 24: The 3rd trial, CD results, MLL5, (LB medium)

We also made the CD (circular dichroism analysis). The protein was mainly in a beta sheet configuration, what is expected for the SET domain, shown in Figure 24. Once again we had no remarkable sucess with the crystallization of our protein. All we were able to obtain were obviously just small amorphic shape crystals.

5.1.5 THE BRADFORD METHOD

Table IV: Measurements of the MLL5-3J protein concentration with the Bradford method.

BSA	0	1	2	5	10
A	0	0,04	0,11	0,27	0,45

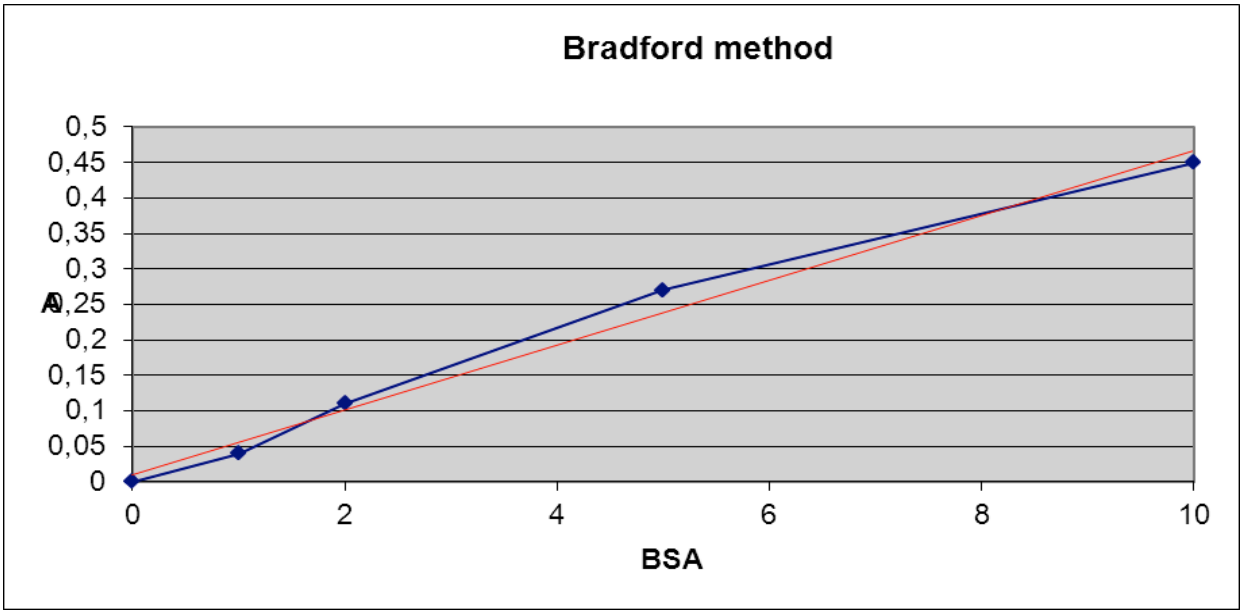

Figure 25: The calibration curve (red) and the graphical presentation of results (blue) obtained by the Bradford method; A – absorbance; BSA – bovine serum albumin.

We can read out the concentration of the sample with help of this graph. The absorption of the sample elution 1 is 0.30 and of elution 2 is 0.05. Then we check with table and graph and calculate that concentration of elution 1 is 1.25 mg/ml and elution's two is 5.5 mg/ml.

1…0.04 5…0.27

X (el1)…0.05 x…0.30

$$x = \frac{0.05}{0.04} = 1.25 \qquad x = \frac{5 * 0.30}{0.27} = 5.5$$

5.1.6 THE RESULTS OF OPTIMISING CRYSTALLIZING BUFFER CONDITIONS

Table V contains the results of drop clearness analysis of different crystallization buffers tested. In Table VI, the protein dispersibility index (PDI) values and peaks of buffers where the drops were found to be clear, are shown.

Legend:

/ - no clear drops;

Ø –clear drop containing 10 mg/ml of MLL5-3J;

· - clear drop containing 10 mg/ml of MLL5-3J with SAM;

·· - clear drop containing 25 mg/ml of MLL5-3J;

All - all drops clear.

Table V: The list of buffers and the radings of drop clearness.

#	Buffer (100 mM)	pH	Clearness	#	Buffer (100 mM)	pH	Clearness
1	Glycine	3	/	12	Cacodylate	6.5	/
2	Citric acid	3.2	Ø ·	13	Ammonium acetate	7	All
3	Citric acid	4	/	14	MOPS	7	All
4	Sodium acetate	4.5	/	15	Sodium phosphate	7	All
5	Sodium phosphate	5	All	16	HEPES	7.5	All
6	Sodium citrate	5.5	All	17	Tris HCl	7.5	All
7	Sodium phosphate	6	All	18	Imidazole	8	/
8	Bis-tris	6	All	19	Tris HCl	8	Ø ·
9	MES	6.2	All	20	CHES	9	Ø ·
10	ADA	6.5	All	21	CHES	9.5	Ø
11	Bis-tris-propane	6.5	All	22	CAPS	10	All

Table VI: The list of tested buffers and the results of measuring protein dispersibility index (PDI) and peaks of buffers where the drops were clear.

	135 µl of buffer + 15 µl of MLL5-3J		**2nd measurement**		**2nd measurement: sample + 15 µl of 5M NaCl**	
#	**PDI**	**PEAK**	**PDI**	**PEAK**	**PDI**	**PEAK**
5	0.460	6.42	precipitant	precipitant	0.747	7.73
6	0.534	8.62	0.840	9.28	1.000	18.9
7	0.397	8.33	0.666	6.80	0.989	8.48
8	0.538	10.6	0.831	7.05	1.000	7.20
9	0.619	7.47	0.710	4.82	1.000	8.11
10	0.653	14.2	0.391	7.58	0.954	7.53
11	1.000	6.20	1.000	7.67	1.000	6.49
13	0.656	7.81	1.000	6.45	0.598	0.64
14	0.543	9.05	precipitant	precipitant	1.000	7.79
15	0.980	7.15	1.000	7.16	1.000	6.67
16	0.322	11.6	1.000	7.68	precipitant	precipitant
17	0.679	8.40	1.000	4.92	0.694	7.06
19	0.515	10.2	0.547	11.1	1.000	9.74
20	0.733	9.35	0.449	24.8	precipitant	precipitant
21	0.410	7.26	0.432	14.3	0.495	14.9
22	0.454	7.92	0.496	17.7	1.000	15.7

The buffers that didn't cause protein precipitation were selected for further analysis. Those with PDI values (the protein dispersibility index is a mean of comparing the water solubility of a protein) around 0,5 and with buffer peaks between 6-10 are better than the rest.

6 CONCLUSION

During the whole process of expression, induction, isolation, purification, concentration and attempted crystallization of our target protein we were able to address several questions by changing and testing experimental protocols. In order to do so, we have carried out three different experiments where we have tested several changes on our experiment and procedure to make the isolation process more efficient, that we would have greater possibility to make crystals in crystallization part.

For example in preparing the optimal Rosetta E. coli strain production cell cultures and harvesting the recombinant protein, two different protocols were compared in the 2nd and the 3rd experimental trial. So in the 3rd trial 100 µL of $ZnSO_4$ solution (10 µM $ZnSO_4$) was added to cell cultures to increase the protein induction. After these were sonicated to harvest the recombinant product (MLL5-3J), DNAse I was added together with its activator $MgCl_2$ to cut bacterial DNA into small pieces, non-interfering with subsequent analyses. Additionally the buffers were changed Tris to Hepes and PBS. Up to this experimental point the procedures were quite efficient, as witnessed by the results of various analyses that were done during the experimental process. However, as stated before the last, i.e. the crystallization step was unsuccessful in all three experimental trials. Therefore we decided to approach this problem as well.

We tried to find a perfect or at least close to perfect MLL5-3J protein isolation and crystallization buffer. For this purpose we analysed an array of different buffers that would enable us to optimize both procedures. We finished with a list of those that seem to have good potentials for achieving this goal. Nevertheless other factors should also be addressed and tested in order to optimize the isolation and crystallization of our target protein.

Another important issue is the expression of the recombinant MLL5-3J protein in Rosetta E. coli strain. According to our results this part of experiment was successful. Unfortunately we were not able to test the stabilizing effect of Zn^{2+} ions on the conformational stage of our protein; which would be very interesting as well.

Even though, we didn't succeed in preparing appropriate crystals of the recombinant MLL5-3J protein, that would allow further structural analysis of its SET domain, we think that we have made some important contributions with testing new approaches and different experimental conditions which might help to obtain a positive result in the future.

7 LITERATURE

1. Deng L-W, Chiu I, Strominger JL. MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. Proc Natl Acad Sci USA, 2004;101:757-762.
2. Cheng F, Liu J, Zhou SH, Wang XN, Chew JF, Deng L-W. RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. Int J Biochem Cell Biol 2008;40:2472-2481.
3. Fujiki R, Chikanishi T, Hashiba W, Ito H, Takada I, Roeder RG, Kitagawa H., Kato S. GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis.Nature 2009;459:455-459
4. http://www.uniprot.org/uniprot/Q8IZD2#ref9 (14.03.2012)
5. Emerling BF, Bonifas J, Kratz CP, Donovan S, Taxlor BR, Green ED, Le Beau MM, Shannon KM. MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. Oncogene 2002; 21(31):4849-4854.
6. http://theses.cz/id/zly82p/downloadPraceContent_adipIdno_9684 (14.03.2012; M.Sc. Ivana Tomčová; New alternative crystallization techniques)
7. http://www.ruppweb.org/Xray/tutorial/protein_structure.htm (14.03.2011)
8. http://www2.bioversityinternational.org/Publications/Molecular_Markers_Volume_1_en/molmarkers/PDF/Protein%20basics.pdf (13.02.2012)
9. http://www.ncbi.nlm.nih.gov/pubmed/10940251 (13.02.2012)
10. Adrian Bird Introduction Perceptions of epigenetics Nature 447, 396-398 (24 May 2007); doi:10.1038/nature05913; Published online 23 May 2007
11. Aasland R, Gibson T, Stewart AF . The PHD finger: implications for chromatin-mediated transcriptional regulation. TIBS 1995, 20, 56-59
12. http://en.inforapid.org/index.php?search=ASH1L (14.02.2012)
13. Pascual J, Martinez-Yamout M, et al. Structure of the PHD Zinc Finger from Human Williams-Beuren Syndrome Transcription Factor. Journal of Molecular Biology 2000;304:723-729.
14. http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=2.1.1.43, (14.02.2012)
15. http://www.kreatech.com/products/repeat-freetm-poseidontm-fish-probes/hematology/on-mds-7q-7q227q36-se-7-tc.htm (20.02.2012)
16. http://www.phosphosite.org/proteinAction.do?id=16432&showAllSites=true (20.02.2011)
17. http://wolfson.huji.ac.il/expression/bac-strains-prot-exp.html (15.04.2012; Dr. Nurit Doron,Bacterial Strains for Protein Expression 2007)
18. "GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis." Fujiki R., Chikanishi T., Hashiba W., Ito H., Takada I., Roeder R.G., Kitagawa H., Kato S. Nature 459:455-459(2009) [PubMed: 19377461]

19. Han Liu, Todd D. Westergard and James J.-D. Hsieh; (2009). MLL5 governs hematopoiesis: a step closer, (113: 1395-1396;)

20. Bernt KM, Armstrong SA. Targeting Epigenetic Programs in MLL-Rearranged Leukemias. Hematology Am Soc Hematol Educ Program 2011;2011:354-360.

21. http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/sl/GELifeSciences-SI/products/AlternativeProductStructure_17325/ (15.03.2012)

22. http://www.sigmaaldrich.com/life-science/proteomics/protein-chromatography/gel-filtration-chromatography.html (15.04.2011)

23. http://www.embl-hamburg.de/biosaxs/reprints/nanodrop.pdf (15.06.2012)

24. http://www.ncbi.nlm.nih.gov/pubmed/17406547 (20.04.2011)

25. http://en.wikipedia.org/wiki/Circular_dichroism (20.04.2011)

26. http://en.wikipedia.org/wiki/Dynamic_light_scattering (20.04.2011)

27. http://www.p212121.com/2009/11/15/the-phase-diagram-in-protein-crystallization/ (15.04.2011)

28. Ranjini Raghunandan, K. Muralidhar, Anamika Sethia Gupta (2005). Imaging and interferometric analysis of protein crystal growth; Center of Excellence in Lasers and Optoelectronic Sciences, Cochin University of Science and Technology, Cochin, Kerala, India;

29. http://www.bio.davidson.edu/courses/molbio/molstudents/spring2003/kogoy/protein.html

30. http://www.qiagen.com/products/easyxtal15-welltools.aspx#Tabs=t0 (15.04.2011)

31. http://www.p212121.com/2009/11/15/the-phase-diagram-in-protein-crystallization/ (15.04.2011)