Tanja Turunen

Optimising timsTOF Pro 2 Mass Spectrometer Sample Preparation

Metropolia University of Applied Sciences

Bachelor of Laboratory Services

Degree Programme in Laboratory Sciences

Bachelor's Thesis

27 May 2022

Abstract

Author:	Tanja Turunen
Title:	Optimising timsTOF Pro 2 Mass Spectrometer Sample Preparation
Number of Pages:	30 pages + 1 appendix
Date:	27 May 2022
Degree:	Bachelor of Laboratory Services
Degree Programme:	Degree Programme in Laboratory Sciences
Supervisors:	Salla Keskitalo, Ph.D, Senior Scientist Markku Varjosalo, Ph.D, Research Director Tiina Soininen, Senior Lecturer

Varjosalo Research laboratory recently acquired a timsTOF Pro 2 mass spectrometer from Bruker. This instrument provides a much higher sensitivity than other mass spectrometers in the laboratory, thus optimising sample preparation became necessary. The goal of this project was to decrease the amount of starting material, i.e. the number of cells needed for a sample meant for analysis with the timsTOF Pro 2, and to determine suitable amounts of purification materials for sample preparation. Based on the original affinity purification sample preparation protocol (published in Nature Protocols in 2020) two modified protocols were designed. These protocols were protocol 3+2, which was used for samples made from three or two cell plates, and protocol 1 for samples made from one cell plate.

First, a trial run of three biological replicates purified from 1–5 cell plates was produced. The number of required interactors needed for a successful sample was decided from the five-plate samples diluted 1:20 according to the original protocol. For the other cell plate numbers, the trial samples were used to determine suitable resuspension volumes and dilutions to meet the criteria set by the five-plate samples. Results from four cell plates did not differ significantly from the three-plate sample results, thus no more four-plate samples were made. A new batch of samples was produced with six biological replicates for 1–3 and five cell plates to test for reproducibility and to allow for better statistical comparison of the identification results.

Protocol 3+2 made with three cell plates with a 1:4 dilution yielded the best overall results. According to statistical analysis the three-plate sample peptide spectrum match (PSM) values did not have statistically significant difference to the five-plate sample (original protocol) PSM values on a 95 % confidence level. Thus, it was concluded that the new protocol 3+2 produced with three cell plates per biological replicate, instead of the original five plates, is the best option for future affinity purification sample analysis with the timsTOF Pro 2. Protocol 3+2 also decreases the amount of purification materials needed for sample preparation, decreasing the overall cost of producing affinity purification samples.

Keywords: timsTOF Pro 2, proteomics, affinity purification

Tiivistelmä

Tekijä:	Tanja Turunen
Otsikko:	TimsTOF Pro 2 -massaspektrometrin näytteenkäsittelyn optimointi
Sivumäärä:	30 sivua + 1 liite
Aika:	27.5.2022
Tutkinto:	Laboratorioanalyytikko (AMK)
Tutkinto-ohjelma:	Laboratorioanalytiikka
Ohjaajat:	Vanhempi tutkija, FT Salla Keskitalo Tutkimusjohtaja, FT Markku Varjosalo Lehtori Tiina Soininen

Tutkimuslaboratorio Varjosalo hankki hiljattain timsTOF Pro 2 -massaspektrometrin Brukerilta. Laite tarjoaa paljon suuremman analyysiherkkyyden kuin muut laboratorion massaspektrometrit, joten näytteenkäsittelyn optimointi oli välttämätöntä. Projektin tavoitteena oli vähentää timsTOF Pro 2 -massaspektrometrille tarkoitetun näytteen aloitusmateriaalin eli solumaljojen määrää sekä määrittää näytteen valmistukseen sopivat määrät puhdistusmateriaaleja. Alkuperäiseen affiniteettipuhdistus-protokollaan (julkaistu Nature Protocolsissa vuonna 2020) perustuen suunniteltiin kaksi modifioitua näytteenvalmistusprotokollaa. Protokollat olivat 3+2 ja 1. Protokollaa 3+2 käytettiin kolmesta tai kahdesta valmistetuille näytteille, ja protokollaa 1 yhden maljan näytteille.

Ensiksi 1–5 solumaljan protokollille tuotettiin kolmen biologisen replikaatin koe-erä. Viiden maljan näytteistä päätettiin onnistuneelle näytteelle vaadittavien proteiinitunnistusten määrä käyttämällä alkuperäisen protokollan mukaista 1:20-laimennosta. Muiden solumaljamäärien koenäytteet käytettiin sopivien resuspensointitilavuuksien ja laimennosten määrittämiseen. Tulokset neljän ja kolmen maljan välillä eivät eronneet merkitsevästi, joten neljän maljan näytteitä ei tuotettu enempää. Seuraavaksi 1–3:een ja viiden solumaljan näytteille tuotettiin uusi erä näytteitä, jossa jokaiselle maljamäärälle tuotettiin kuusi biologista replikaattia parempaa tilastollista vertailtavuutta varten sekä tulosten toistettavuuden testaamiseksi.

Protokolla 3+2 kolmesta solumaljasta 1:4-laimennoksella valmistettuna tuotti parhaat tulokset. Tilastollisesti vertailtaessa kolmen maljan proteiinien suhteellisia määriä (PSM-arvo) viiden maljan (alkuperäinen protokolla) PSM-arvoihin tulokset eivät eronneet toisistaan merkitsevästi 95 %:n luottamustasolla. Täten voitiin päätellä uuden protokollan 3+2 kolmella maljalla tuotettuna olevan paras vaihtoehto tuleville affiniteettipuhdistetuille timsTOF Pro 2 -näytteille. Protokolla 3+2 vähentää affiniteettipuhdistuksessa kuluvien puhdistusmateriaalien määriä, pienentäen affiniteettipuhdistuksen kokonaiskustannuksia.

Avainsanat: timsTOF Pro 2, proteomiikka, affiniteettipuhdistus

Contents

List of Abbreviations

AP: Affinity purification.

BioID: Proximity-dependent biotin identification.

CCNC: Cyclin C.

CCNH: Cyclin H.

CDK7: Cyclin-dependent kinase 7.

CDK8: Cyclin-dependent kinase 8.

ESI: Electrospray ionisation.

HPLC: High-performance liquid chromatography.

LC: Liquid chromatography.

LC-MS/MS: Liquid chromatography tandem mass spectrometry.

MAC-tag: Multiple approaches combined -tag.

MAT1: CDK-activating kinase assembly factor (French. ménage à trois-1).

MED: Mediator complex subunit.

MED12: Mediator complex subunit 12.

MED13: Mediator complex subunit 13.

MS: Mass spectrometry.

MS scan: A mass spectrometry scan mode where all the precursor ions are displayed in the mass spectrum.

MS/MS scan: A mass spectrometry scan mode where all the product ions fragmented from precursor ions are displayed in the mass spectrum.

PASEF: Parallel accumulation-serial fragmentation.

PIC: Transcription preinitiation complex.

PoI: Protein of interest.

PPI: Protein-protein interaction.

PSM: Peptide spectrum match.

QTOF: Quadrupole time-of-flight.

TFIIH/TF2H: Transcription factor II H.

TIMS: Trapped ion mobility spectrometry.

timsTOF: Trapped ion mobility spectrometer-time of flight.

TOF: Time-of-flight.

1 Introduction

Optimising sample preparation is a worthy endeavour regardless of sample type or analysis method. In the case of proteomics, the difference between a sample made with an optimised protocol versus a general one, can either make or break the overall results. When studying protein interactions any step to ensure better sample quality is one worth taking.

University of Helsinki's Varjosalo Research Laboratory does proteomics research on a wide scale and is constantly looking for ways to improve analysis methods and upgrade its instruments. In March 2021, the laboratory acquired a timsTOF (trapped ion mobility spectrometer-time of flight) Pro 2 mass spectrometer from Bruker (Germany). The instrument provides a much higher resolution, sensitivity, and analytical capacity than the other mass spectrometers in the laboratory from a significantly smaller amount of starting material, thus optimising sample preparation became necessary.

In 2020 the laboratory published a protocol for studying protein interactions called "Combined proximity labelling and affinity purification−mass spectrometry workflow for mapping and visualizing protein interaction networks" in Nature Protocols [1]. The majority of the laboratory's samples are produced according to this protocol. This project aims to decrease the amount of starting material needed for a sample meant for analysis with the timsTOF Pro 2. When the starting material (i.e. the number of cells) is decreased, it is necessary to determine the adequate amounts of purification materials, the correct elution volumes, and so on needed for sample preparation. The samples produced with the modified protocols are compared to ones produced following the original protocol using statistical mathematics to maintain sample quality.

2 Theory

2.1 Proteomics

The human body consists of roughly 15 % of protein [2]. From skin and muscle to the tiny receptor molecules and cell organelles, everything is largely composed of protein. Proteins are made of amino acids that form chains. These chains twist and fold, and often bind to other proteins, which gives each protein its functional form. [3.] When proteins interact with other proteins, they form protein-protein interactions (PPI). PPIs are fundamental for the function of cells and consequently any living organism. Gene expression, cell growth, proliferation, apoptosis, and intercellular communication are among the many biological processes that are facilitated by proteins. Cells respond to stimuli by altering their protein expression. Different cell-types also express different kinds of proteins, therefore understanding protein function is a complex challenge, especially when trying to do so in the proper biological context. [4.]

Proteomics is the study of proteomes. Proteomes are sets of proteins produced in an organism, system, or biological context. For example, one can refer to the proteome of a species, e.g. the human proteome, or the proteome of an organ, e.g. the liver. Proteomics can be used to study when and where proteins are expressed, their rates of degradation and steady-state abundance, post-translational modifications, how proteins move between subcellular compartments, their involvement in metabolic pathways, and how they interact with each other. [5.]

Besides overall understanding of human biology, proteomics has created opportunities to study proteins in diseases. Disease proteomics aim to understand how altered protein expression, structure and function can cause illness. This provides opportunities for more accurate diagnostics, disease prevention, and more specific treatment. For example, by combining proteomics and genomics tools researchers have been able to find different markers to differentiate colon and ovarian cancer from each other. Previously these

diseases were difficult to distinguish. Correct diagnosis for the different cancers is important since their treatments vary markedly and wrong treatment can be ineffective or even dangerous. [6.]

2.2 Test Constructs: Cyclin-Dependent Kinases 7 and 8

In this project, two constructs were used to test the quality of the samples produced with the modified protocols. The constructs were cyclin-dependent kinase 7 (CDK7) and cyclin-dependent kinase 8 (CDK8). Both are involved in the formation of the transcription preinitiation complex (PIC). PIC is composed of several transcription factors, such as transcription factor II H (TFIIH) and the Mediator complex. Among other tasks, PIC recruits RNA polymerase II to the promoter of the transcribed gene, and thus takes the first step of initiating transcription. [7, p. 7.]

Protein kinases, such as CDK7 and CDK8, are enzymes that play a critical role in cell signal transduction and other cellular processes. They catalyse the transfer of a phosphate group from adenosine triphosphate (ATP) to serine, threonine, or tyrosine side chains on the target proteins. This reaction is called phosphorylation. Kinases are abundant in the human genome and important for the normal function of cells. Altered kinase activity is present in various human diseases, such as cancer. [8.]

CDK7 is a protein kinase that along with cyclin H and CDK-activating kinase assembly factor (MAT1) form a subcomplex to TFIIH. TFIIH has an important role at the start of transcription for it unwinds the template DNA, positions it correctly in the RNA polymerase II’s active site, and phosphorylates a serine at the C terminal domain of the RNA polymerase to initiate RNA production. [7, p. 7.] The position of CDK7 in TFIIH can be seen in Figure 1.

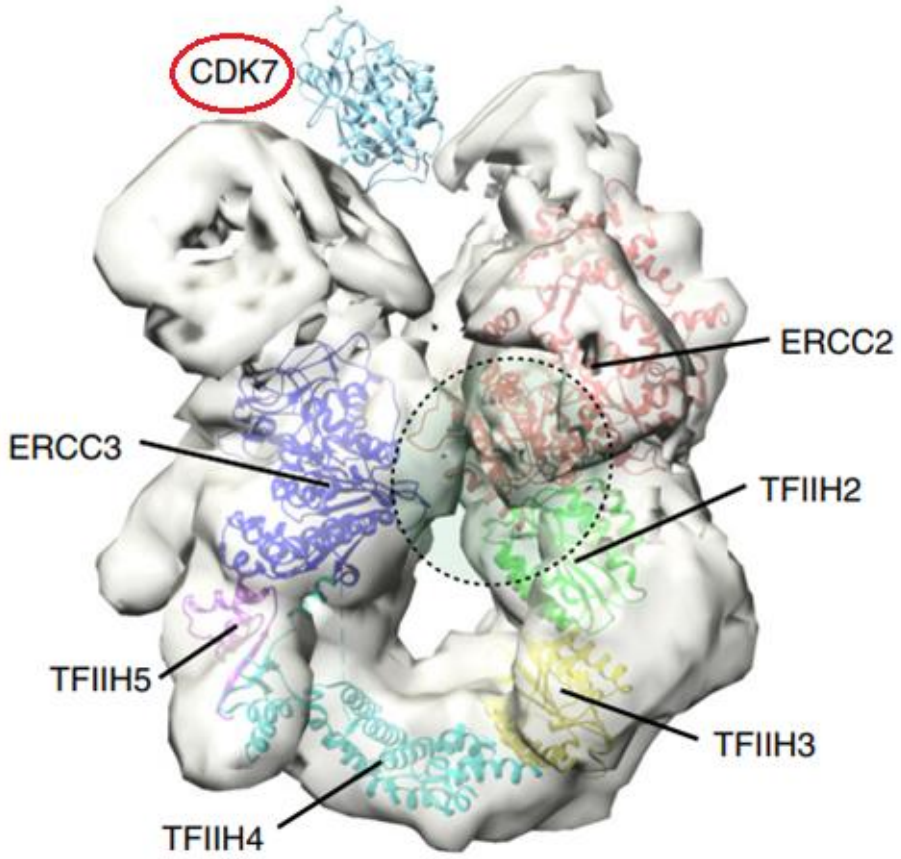

Figure 1. Structure of TFIIH. CDK7 is the upmost protein depicted in blue and circled in red [9].

CDK8 is a protein kinase and a subunit of the Mediator complex. The Mediator complex can be divided into four different parts: the head, middle, tail, and kinase module. CDK8 is located in the kinase module that connects to the middle part of the complex by the Mediator complex subunit 13 (MED13). The other two components of the kinase module are cyclin C and Mediator complex subunit 12 (MED12). [10.] The Mediator complex has multiple gene specific binding sites for transcription factors, and it can bind several transcription factors simultaneously. The binding sites are mainly located in the tail and the kinase module. [7, p. 18–19.] Structure of the Mediator complex is shown in Figure 2.

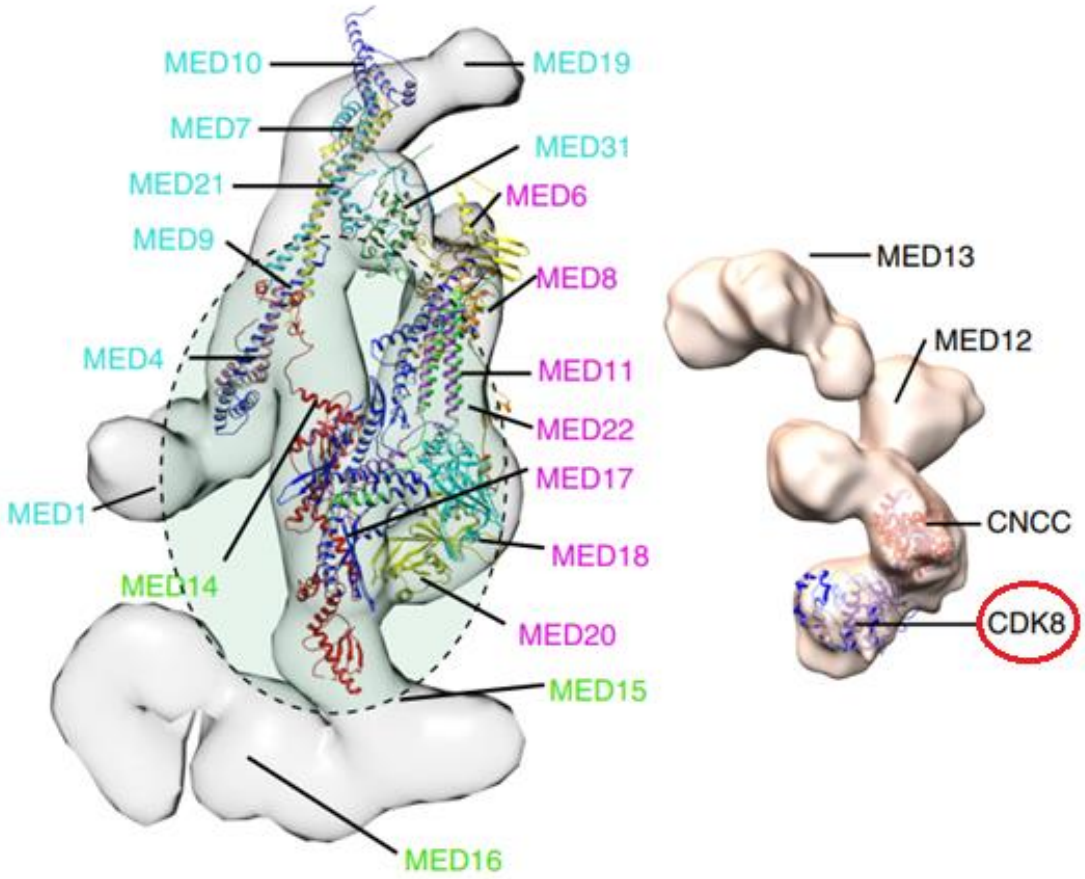

Figure 2. The Mediator complex and the kinase module. Subunits of the tail are indicated in green, subunits of the middle are cyan, and subunits of the head are magenta. Subunits of the kinase module are marked in black. CDK8 can be seen in the lower right corner, circled in red. [9.]

2.3 MAC-tag Based Affinity Purification Workflow for Analysing Protein Interactions

There are multiple approaches on how to produce a sample for PPI analysis. One of them is affinity purification (AP). AP is one of the most used methods of protein sample purification due to its robustness, especially in one-step purifications. AP relies on a matrix bound ligand that specifically, but reversibly, binds either the protein of interest (PoI) or a tag that has been added to the protein. [11.] The PoI acts as a bait to its stable interactors (preys). Varjosalo Research Laboratory's Nature Protocols workflow "Combined proximity labelling and affinity purification−mass spectrometry workflow for mapping and visualizing protein interaction networks" [1] is an example of an AP protocol. In Varjosalo's protocol [1], a multiple approaches combined (MAC)-tag is added to the PoI. The MAC-tag has a short amino acid sequence that binds to streptavidin beads packed in a Bio-Rad Chromatography Column. When a sample containing the PoI and other molecules is introduced to the column, the PoI binds (i.e. has an *affinity* to the ligand streptavidin) to the beads and is retained, while other impurities only pass through as waste. The PoI and the stable interactors that are bound to it are then eluted out of the column with an elution buffer that has a high concentration of biotin. Biotin acts as a competing binder that replaces the PoI. [1.] An overview of the purification step is in Figure 3.

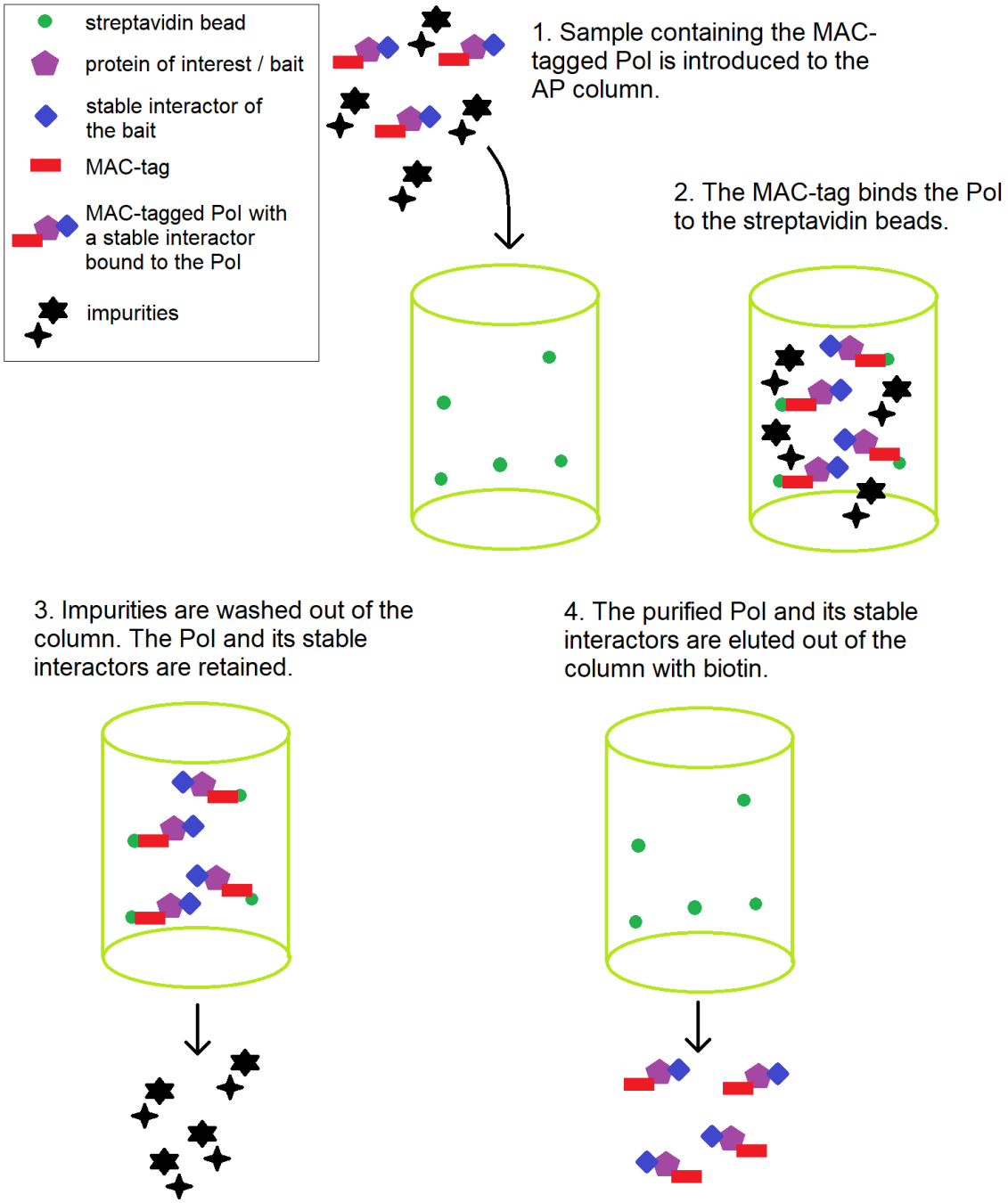

Figure 3. An overview of Varjosalo's MAC-tag based AP sample purification.

The eluted sample contains the PoI and its stable interactors. Purified protein samples are digested into peptides using trypsin, desalted, and subjected to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. In the LC-MS/MS analysis, the peptides are separated with high-performance liquid chromatography (HPLC) based on their hydrophobicity, then ionised, fragmented, and analysed with a mass spectrometer. In this project, the LC-MS/MS analysis was carried out with timsTOF Pro 2 mass spectrometer coupled with an EvoSep One HPLC. From MS data, the PoI and its stable interactors are identified by searching them against a protein library. The identification results are filtered from common contaminants and low-confidence interactors, leaving only the bait and the high-confidence interactors. Based on AP MS results, it is possible to identify all the stable interactors that a PoI has, and their relative abundance in the sample. [1.]

2.4 Evosep One and timsTOF Pro 2

Evosep One is a novel liquid chromatography (LC) instrument developed by Evosep Biosystems (Denmark). Instead of traditional sample vials, samples are loaded into Evotips (Figure 4). Evotips resemble regular 20–200 µl disposable pipet tips but they have a small C18 column at the end of the tip. Integrated de-salting on disposable tips improves sample purity and significantly reduces carry-over during sample loading [12]. The injector elutes the sample through the tip column into the LC’s sample loop. From there, the sample proceeds to a reverse-phase C18 analytical column.

Figure 4. Pictures of the Evosep One liquid chromatogram (bottom left), Evotips (upper left), and the timsTOF Pro 2 mass spectrometer (right).

After separation with the LC's analytical C18 column the peptides are ionised with electrospray ionisation (ESI). The ions enter the dual trapped ion mobility spectrometry (TIMS) tunnels through a glass capillary. The first TIMS tunnel acts mainly as an ion accumulation trap that stores all ions that enter the mass spectrometer, while the second TIMS tunnel performs trapped ion mobility analysis. In the second TIMS tunnel, ions simultaneously experience a drag from the incoming gas flow through the capillary and resistance from an electric field. The ions are separated based on their charge states and collisional cross

sections. Ions with high ion mobility come to rest closer to the entrance of the TIMS tunnel, while ions with low ion mobility travel closer to the exit. [13.]

The timsTOF Pro 2 system is a quadrupole time-of-flight (QTOF) spectrometer. After the ions are released sequentially from the dual TIMS tunnels as a function of decreasing electrical field strength, they pass through the quadrupole mass filter and are then accelerated into a collision cell. From there intact (MS scan) or fragmented daughter ions (MS/MS scan) are pushed into the V-shaped flight path of the time-of-flight (TOF) analyser where at the end, the ions impinge on a multichannel plate ion detector. [13.] TOF analyser determines the ions' mass-to-charge ratio (m/z) based on the time it takes for the ions to travel through the TOF flight path [14]. An overview of the function of the timsTOF Pro 2 mass spectrometer is shown in Figure 5.

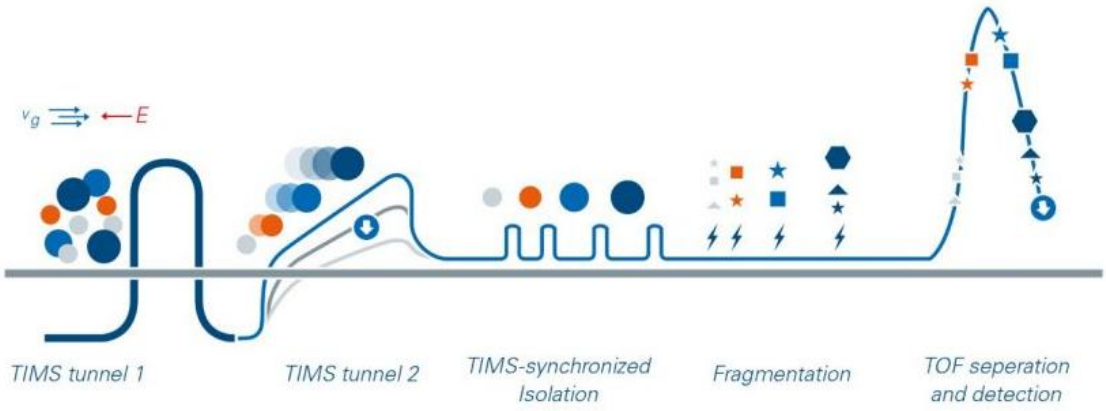

Figure 5. An overview of the timsTOF Pro 2 mass spectrometer. After ESI ions are accumulated in the first TIMS tunnel and separated in the second TIMS tunnel based on their ion mobility. From there, the ions are released sequentially, fragmented, separated in TOF, and finally detected. [15.]

The name "parallel accumulation-serial fragmentation" (PASEF) comes from the fact that with the dual TIMS tunnels it is possible to simultaneously collect and fragment ions. With previous mass spectrometers, there is a significant loss of sample, since when a selected precursor is fragmented and analysed the left over ions go to waste. This does not happen when ion accumulation and fragmentation happen parallel to each other so that no sample is lost.

Therefore, PASEF is a revolutionary MS method with a 100 % duty cycle that multiplies sequencing speed without sacrificing sensitivity. [13.]

3 Optimising Sample Preparation

3.1 Original Workflow

The “Combined proximity labeling and affinity purification−mass spectrometry workflow for mapping and visualizing protein interaction networks” [1] can be used to prepare AP and proximity-dependent biotin identification (BioID) samples. This project focuses on the optimisation of the AP workflow.

In the original Nature Protocols workflow [1], a PoI is cloned from a Gateway® recombination compatible human ORFeome collection (CCSB-DFCI hORFeome collection(51020@G07)) into an expression vector that adds the MAC-tag either to the N or C terminal end of the PoI. The expression vector is transfected into a Flp-In™ T-REx 293 host cell line (prod. no. R78007, Thermo Fisher). After a two week selection period with Hygromycin B, the cells are expanded and grown to confluency. Expression of the PoI is induced with tetracycline and after 24 hours the cells are harvested. After harvesting, the pelleted, flash-frozen cells are be stored in -80 °C until purification. After AP the purified sample is digested and analysed with mass spectrometry. [1.] An overview of the original Nature Protocols workflow [1] is in Figure 6.

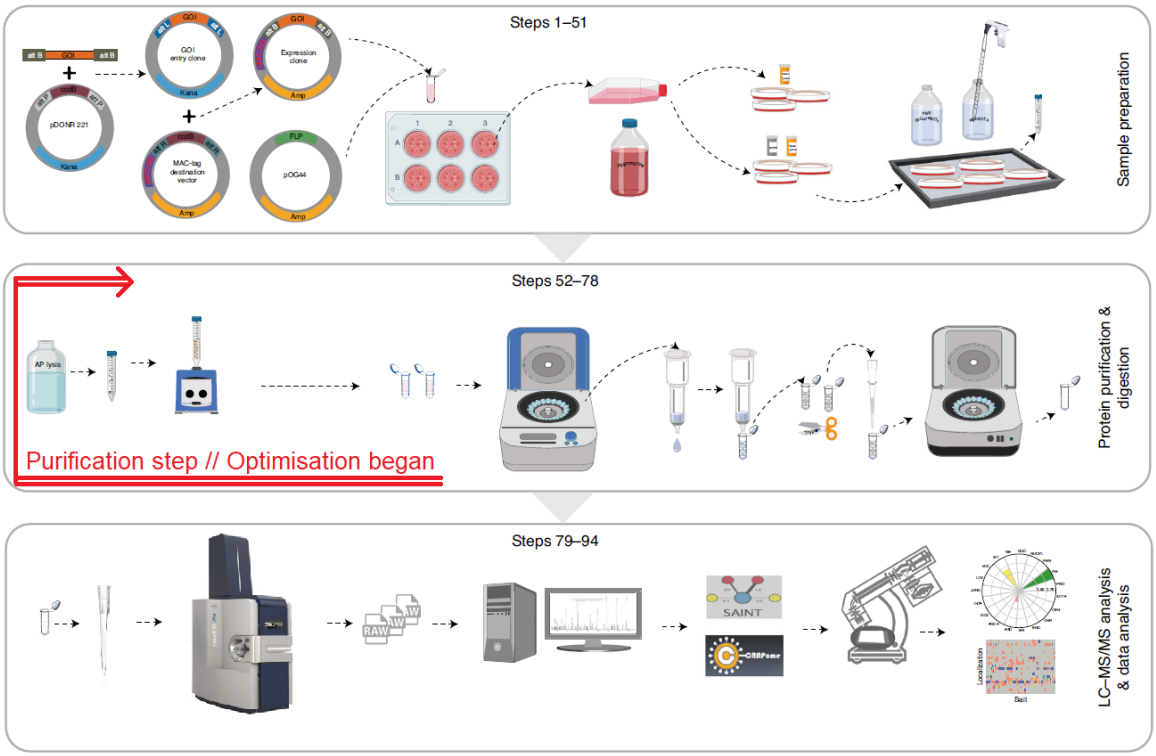

Figure 6. An overview of the steps of the Nature Protocols AP workflow. The point of where the optimisation project began is marked with a red arrow and text. [1.]

In this project the PoIs were CDK7 and CDK8. The CDK7 and CDK8 entry clones were MAC-tagged N terminally, and stable cell lines were generated according to the original protocol. Optimisation of the purification and sample preparation began from the harvesting step (step 47 “Generation of cell pellets for protein purification”) [1]. The step in question is visualised in Figure 6 with a red arrow. Before this step, the Nature Protocols workflow [1] was not altered.

3.2 Modifying the Original Protocol

Optimising sample preparation began by determining suitable modified protocols for cell pellets collected from 1–5 Ø 15 cm cell plates. It was decided that lysates that were collected from four (4) Ø 15 cm plates would be prepared the same way as lysates from five (5) plates using the original protocol. In a similar fashion, lysates collected from three (3) and two (2) cell plates were purified with their own protocol, and lysates made from one (1) cell plate had

their own protocol. All of the protocols are shown in Table 1. The protocols are addressed as protocol 5+4, protocol 3+2, and protocol 1 respectively, or by the number of plates used per sample.

Table 1. A summary of protocols 5+4, 3+2, and 1. Protocol 5+4 is the original protocol. Protocol 5+4 was used to purify samples collected from five or four cell plates, protocol 3+2 was used for three- or two-plate samples, and protocol 1 was used for samples made from only one cell plate. Protocol 1 was tested using columns from Bio-Rad and Thermo Fisher. The protocol used is marked with the corresponding column manufacturer.

Step:	**Protocol:**			
Affinity purification	**5+4 (original protocol)**	**3+2**	**1 (Bio-Rad)**	**1 (Thermo)**
No. of plates:	5 or 4	3 or 2	1	1
Lysis buffer	3 ml	2 ml	0,6 ml	0,6 ml
Centrifugation	16 000 x g, 15 min Move to a new tube 16 000 x g, 10 min	16 000 x g, 15 min Move to a new tube 16 000 x g, 10 min	16 000 x g, 15 min Move to a new tube 16 000 x g, 10 min	16 000 x g, 15 min Move to a new tube 16 000 x g, 10 min
Preparing the column	400 µl of 50 % bead slurry	250 µl of 50 % bead slurry	100 µl of 50 % bead slurry	25 µl of lysis buffer + 100 µl of 50 % bead slurry + 50 µl of lysis buffer Spin 100 x g, 2 min
Equilibration	1 ml of lysis buffer	1 ml of lysis buffer	1 ml of lysis buffer	2 x 0,15 ml of lysis buffer Spin 100 x g, 2 min
Loading the sample	≈ 4 ml by pouring	≈ 2,5 ml by pouring	≈ 1 ml by *careful* pipeting	≈ 1 ml by pipeting
Wash with lysis buffer	3 x 1 ml	3 x 1 ml	3 x 1 ml	3 x 0,3 ml Spin 200 x g, 1 min
Wash with HENN	4 x 1 ml	4 x 1 ml	4 x 1 ml	4 x 0,3 ml Spin 200 x g, 1 min
Elution	3 x 300 µl of 0,5 mM biotin	3 x 240 µl of 0,4 mM biotin	150 µl + 120 µl of 0,4 mM biotin	150 µl + 120 µl of 0,4 mM biotin Spin 100 x g, 1 min
In-solution digestion				
TCEP (50 mM)	100 µl	80 µl	30 µl	30 µl
IAA (100 mM)	111 µl	89 µl	33 µl	33 µl
Trypsin (0,5 µg/µl)	3 µl	2 µl	0,6 µl	0,6 µl

The purification steps, such as equilibration of the column, the number of washes etc. were not changed, but the volumes of streptavidin purification beads (Strep-Tactin Sepharose, prod. no. 2-1201-010, IBA Life Sciences), wash and elution buffers, and for protocols 3+2 and 1 the concentration of the elution buffer were modified. The modifications were made with careful consideration taking into account the existing protocol and IBA Life Sciences' instructions for the binding properties of the streptavidin beads without forgetting practicality.

Protocol 1 (made from one cell plate) had an obvious issue with only having 50 µl of streptavidin beads in the Bio-Rad Chromatography Column (prod. no. 7326008). The length of the bead bed affects the sample's affinity to the column, and if the bead bed is too short, the whole protein sample could possibly only pass through the column and nothing would be retained. Pierce™ Micro-Spin (prod. no. 89879) column by Thermo Fisher has an inner diameter of 4 mm in which 50 µl of streptavidin beads brings the bead bed length to roughly 1.5 cm. The length is sufficient compared to the bead bed length in the original protocol in a Bio-Rad column, which is around 1.0 cm. A picture showing the difference in the 50 µl bead bed length in a Bio-Rad and Thermo Fisher column can be seen in Figure 7. Unlike the Bio-rad columns, Thermo Fisher columns cannot be used with gravity flow. Thus suitable centrifugation speed and times needed to be determined as it is important that the beads do not dry during purification. The tests were done using 50 µl of expired streptavidin beads and water to act as a buffer. The final centrifugation speeds and times can be seen in Table 1.

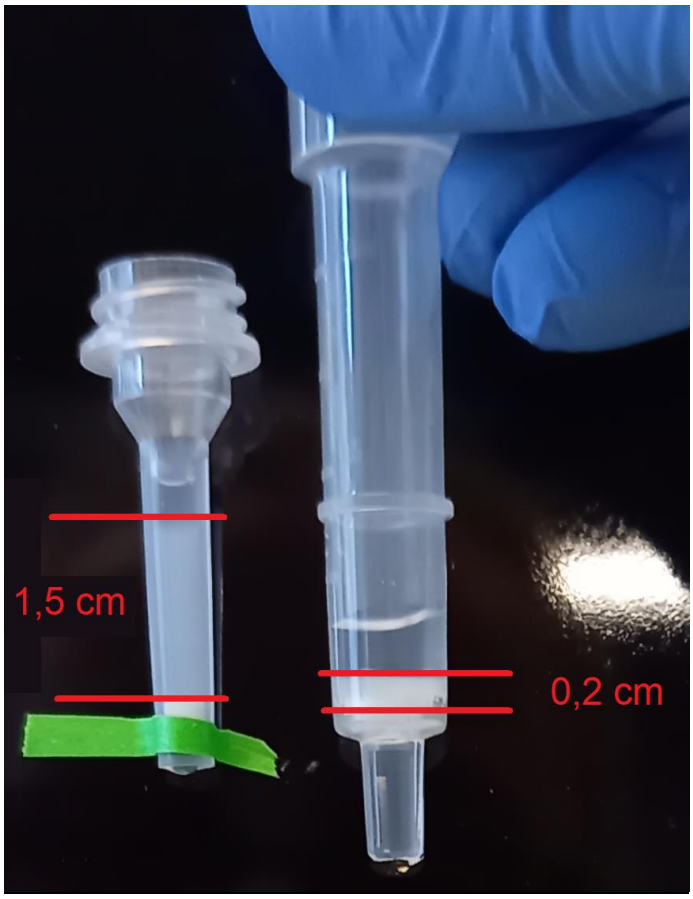

Figure 7. Difference in bead bed length using 50 µl of purification beads in a Thermo Fisher Pierce™ Micro-Spin column (left) and a Bio-Rad Chromatography Column (right). The original bead bed length in a Bio-Rad column using 200 µl of purification bead is 1.0 cm. The 1.5 cm bead bed in the Thermo Fisher column is adequate compared to this.

However, the Thermo Fisher columns proved to be quite difficult to handle and despite determining the suitable centrifuging conditions, some of the columns would still dry completely or all of the liquid would not have come through with the planned centrifugation speeds and times. Therefore, it was decided that despite the 0.2 cm bead bed length being short compared to the original 1.0 cm bead bed, protocol 1 was tested using both the Thermo Fisher and Bio-rad columns to see if either of them produced acceptable purification results. Pictures of the different bead bed lengths in Bio-Rad columns are in Figure 8.

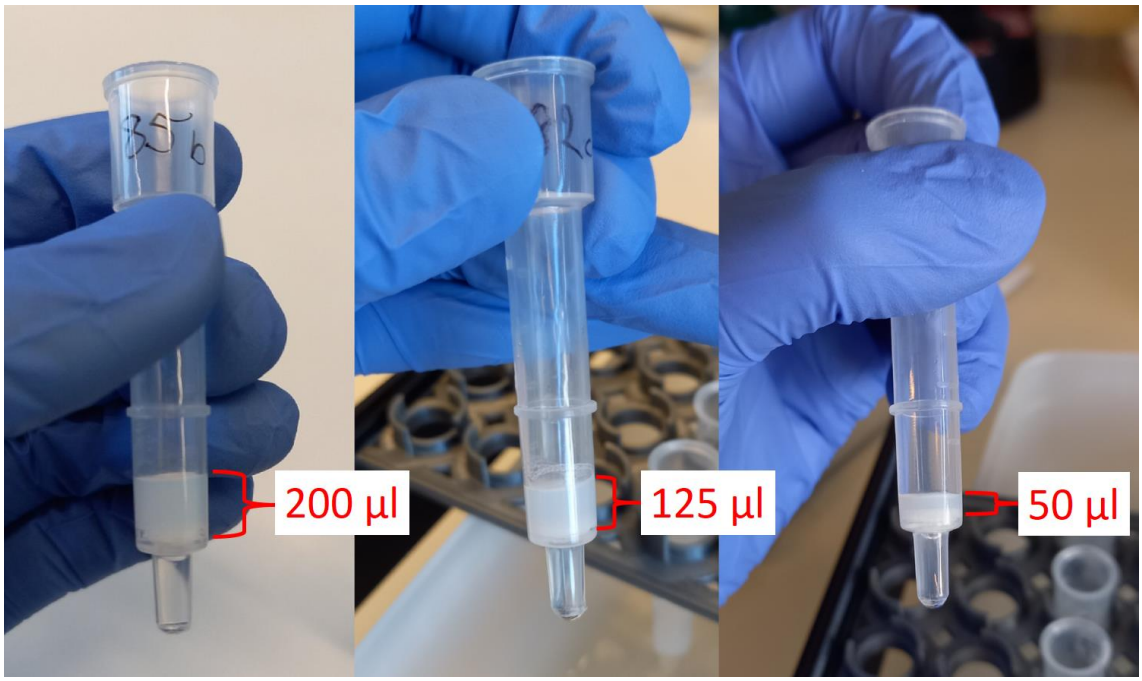

Figure 8. Different bead beds with different volumes of streptavidin beads in Bio-Rad Chromatography Columns. A column prepared according to the original Nature Protocols workflow (protocol 5+4) is on the left with 200 µl of beads, protocol 3+2 with 125 µl of beads is in the middle, and protocol 1 with 50 µl of beads is on the right.

3.3 Test Runs

Before a purified sample can be run on the timsTOF Pro 2, it needs to be diluted. The usual dilution for a sample produced according to the original protocol is 1:20 (1 µl of the purified sample + 19 µl of timsTOF buffer A). This leaves a wasteful surplus of purified sample since usually only 1–2 µl out of 30 µl is used. CDK7 and CDK8 were chosen to be the test samples (baits) because they have been well studied by Varjosalo laboratory. The stable interactors (preys) that CDK7 and CDK8 should have are known. The number of stable interactors identified from the test samples would act as a first step of quality control. Only if all of the required stable interactors were identified from a sample produced with a modified protocol, the sample would be considered a success. Otherwise, a more concentrated sample dilution would be tested, or the sample would be deemed a failure.

As a trial run to start the optimisation, the first batch of samples had pellets collected from 1–5 cell plates with three biological replicates each. These samples were used to determine the stable interactors that could be found with the original five-plate protocol, and to test for suitable dilutions for protocols 3+2 and 1. With these results, a second batch of purifications were made with six biological replicates to test for reproducibility and to allow for better statistical comparison of the results.

4 Results

4.1 Trial Runs

Based on the first five-plate test samples done according to the original protocol with the usual 1:20 dilution (1 µl of purified sample + 19 µl of timsTOF buffer A), it was decided that from a CDK7 sample the bait (CDK7) as well as stable interactors cyclin H (CCNH), MAT1, and at least three of the five TFIIH (TF2H1–5 in the tables) subunits should be identifiable after filtering for the sample to be considered successful. For CDK8 samples, the bait (CDK8) and stable interactors cyclin C (CCNC), and at least 15 out of 27 Mediator complex subunits (MED) should be identifiable after filtering. In addition to the number of identifications, peptide-spectrum match (PSM) values reflecting the proteins relative abundance were monitored.

The acquired MS data was analysed using FragPipe (ver. 17.1) with MSFragger (ver. 3.4) and Philosopher (ver. 4.1.1, build 1637179075). The results were searched against a reviewed human proteome (UP000005640, 20361 entries), downloaded from Uniprot.org on the 8th of March 2022. The results were filtered so that all identifications with a PSM ratio less than or equal to 20 % compared to matches found in the CRAPome 2.0 database were accepted and from those identifications the ones with a SaintScore equal or greater than 0.6 were accepted. The SaintScore cut-off value came from a SAINTexpress analysis performed at proteomics.fi which is an analysis platform maintained by the Varjosalo laboratory.

Every sample type was at first run with the usual 1:20 dilution to see when the number of stable interactors would start to decrease. The results are reported in Table 2 and Table 3. The difference between samples collected from four or three plates is not substantial, thus no more four-plate samples were produced after the initial trial batch. With the 1:20 dilution, the number of identifications and PSM values started to decrease in samples purified from three cell plates, but the difference was smaller than expected. On the other hand, with this dilution samples made from only one or two cell plates had lost all interactors, but this was according to expectations. Still, the PSM values of the baits looked promising that with the right dilution these samples could provide usable data.

Table 2. Results from the CDK7 trial run with all of the samples diluted 1:20 (1 µl of purified sample + 19 µl of timsTOF buffer A). Samples made from four or three plates have only minor differences in the identification results. The number of identifications from the three-plate samples is acceptable but the PSM values start to decrease. In the table, PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. The PSM values of different biological replicates are separated within the cell by a vertical line.

	CDK7					
No. of plates:	5	4	3	2	1 (Bio-Rad)	1 (Thermo)
Bait (PSM)	124|152|125	92|84|77	102|68|150	75|91|86	15|26|20	14|25|0
CCNH (PSM)	72|83|82	80|70|71	86|38|86	42|41|58	0	0
MAT1 (PSM)	53|20|75	42|57|42	48|66|70	30|34|33	2|7|0	0
TF2H1 (PSM)	40|37|22	11|5|11	15|0|17	0|4|6	0	0
TF2H2 (PSM)	0	0	0	2|6|7	0	0
TF2H3 (PSM)	6|8|11	6|6|4	6|0|8	0	0	0
TF2H4 (PSM)	22|19|17	12|11|12	11|0|10	8|11|14	0	0
TF2H5 (PSM)	7|6|4	5|2|6	2|0|2	0	0	0
Criteria met?	Yes	Yes	Yes	No	No	No

Table 3. Results from the CDK8 trial run with all the samples diluted 1:20 (1 µl of purified sample + 19 µl of timsTOF buffer A). Samples made from four or three plates have only minor differences in the identification results. The number of identifications from the three-plate samples is close to being acceptable but the PSM values start to decrease especially for the Mediator complex subunits 1–30. In the table, PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. In the "Total number of MED" row the initial number of filtered MED interactors is marked plainly, and the number of MEDs with the "saved" green MEDs added is marked in brackets. The PSM values of different biological replicates are separated within the cell by a vertical line.

	CDK8					
No. of plates:	5	4	3	2	1 (Bio-Rad)	1 (Thermo)
Bait (PSM)	176|177|184	178|177|179	135|190|190	67|123|135	15|51|59	12|12|16
CCNC (PSM)	176|177|184	178|177|179	135|190|190	67|123|135	15|51|59	12|12|16
Tot. No. MED	15 (17)	5 (12)	5 (11)	0	0	0
MED1	11|14|19	10|4|6	7|8|4			
MED4	14|15|12	6|7|6	4|7|3			
MED6	1|1|1	0|0|1				
MED8	2|4|3	1|1|1	0|1|0			
MED9	1|1|1					
MED10	5|4|5	2|2|2	2|4|4			
MED11	1|2|1	1|2|1	0|1|1			
MED12	51|49|51	16|14|13	6|19|14			
MED13	8|16|14	1|4|1				
MED14	17|29|23	4|1|0	0|6|0			
MED15	9|12|14	1|0|0	0|1|0			
MED16	14|13|9	4|8|1	2|5|3			
MED17	7|8|13	6|4|2	1|3|5			
MED19	2|1|1					
MED20	5|3|3	1|1|0	0|2|1			
MED21	5|7|8	2|1|3	2|1|2			
MED22	2|2|3	0|0|2	0|0|1			
MED23	9|11|8	6|7|4	4|4|4			
MED24	23|32|32	12|10|9	10|17|6			
MED25	1|2|1	2|0|0	1|4|2			
MED27	4|5|4					
MED28	2|0|0	0|1|0				
MED29	1|0|0		0|1|2			
MED30	6|4|5	2|5|2	3|3|3			
Criteria met?	Yes	No	No	No	No	No

4.2 Dilution Tests

The dilution tests were only made with CDK8 samples to save time and samples. First, a 1:4 dilution (5 µl of the purified sample + 15 µl of timsTOF buffer A) was made for samples produced with protocol 3+2, and 1. The test results can be seen in Table 4. With this dilution, samples purified from three cell plates yielded excellent results even after filtering but results from two cell plates were still inadequate. After filtering, many of the stable interactors were lost from the two-plate samples, thus sample quality was not yet satisfactory. The same applied for the one-plate samples.

Table 4. Results of the 1:4 (5 µl of the purified sample + 15 µl of timsTOF buffer A) dilution tests. Identifications are satisfactory for the three-plate samples but the two- and one-plate samples do not have enough of the required identifications found. The numbers in brackets depict the number of MED found before filtering, and plain one is the total after filtering.

	CDK8			
No. of plates:	3	2	1 (Bio-Rad)	1 (Thermo)
Bait (PSM)	252	117	89	73
CCNC	10	0	3	0
Tot. No. MED	22 (26)	2 (6)	2 (9)	1 (5)
Criteria met?	Yes	No	No	No

For the two-plate samples, a 1:2 dilution (10 µl of the purified sample + 10 µl of timsTOF buffer A) was made but this did not solve the issue. In a total ion chromatogram (TIC) from one of the 1:2 dilution tests there is a significant drop in signal intensity at 7.0 minutes (Figure 9). This drop has been more pronounced when a more concentrated sample of any cell plate number was made. The drop could be caused by the increased amount of biotin in the sample, since it is known that too much biotin in a sample will start to cover any other signals, thus it could explain the loss of interactors.

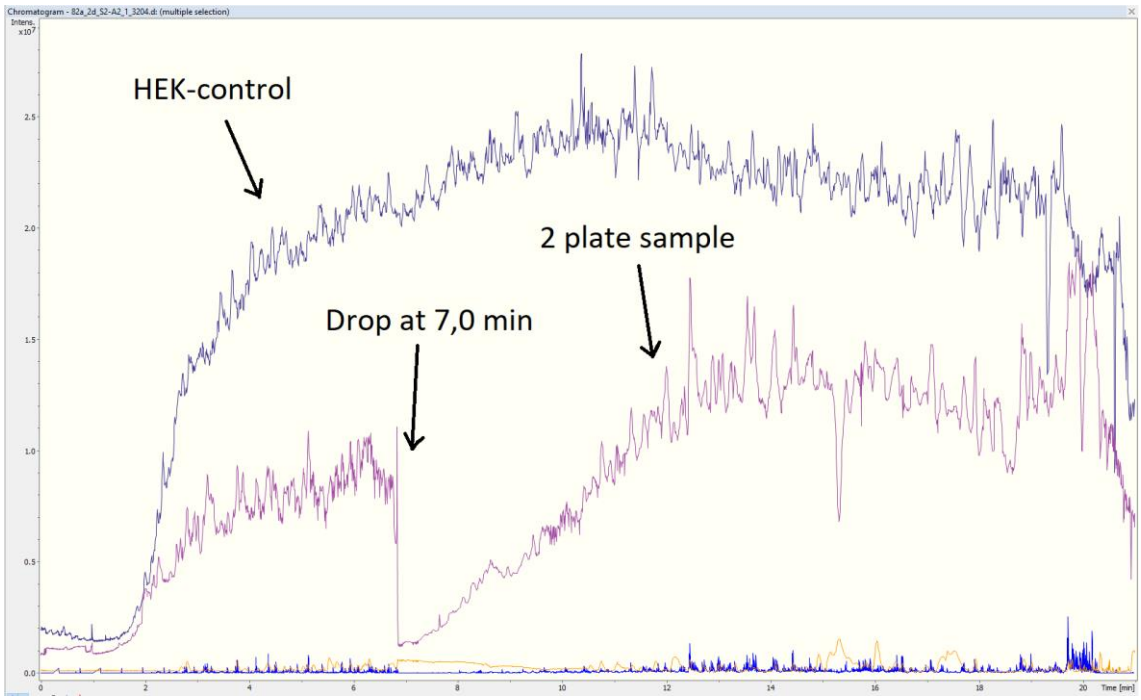

Figure 9. A chromatogram with a suspected drop in the TIC signal intensity due to a high biotin concentration. The chromatogram's x-axis is time and y-axis is signal intensity. The upper dark blue graph shows a normal HEK control run with the correct shape to a timsTOF graph. In comparison a two-plate sample with a 1:2 dilution (10 µl of purified sample + 10 µl of timsTOF buffer A), marked in pink, has a significant drop in signal intensity at 7.0 minutes.

Next, a different approach was taken for the dilution. After desalting the sample, instead of resuspending the sample in 30 µl of buffer A, a two-plate sample was resuspended in 50 µl of buffer A. From this, 20 µl of the resuspension was loaded directly into an Evotip without any further dilution. The TIC (Figure 10) signal intensity is overall higher and shows that while there still is a significant drop in signal intensity at 7.0 minutes, the instrument is able to recover much faster compared to the 1:2 dilution.

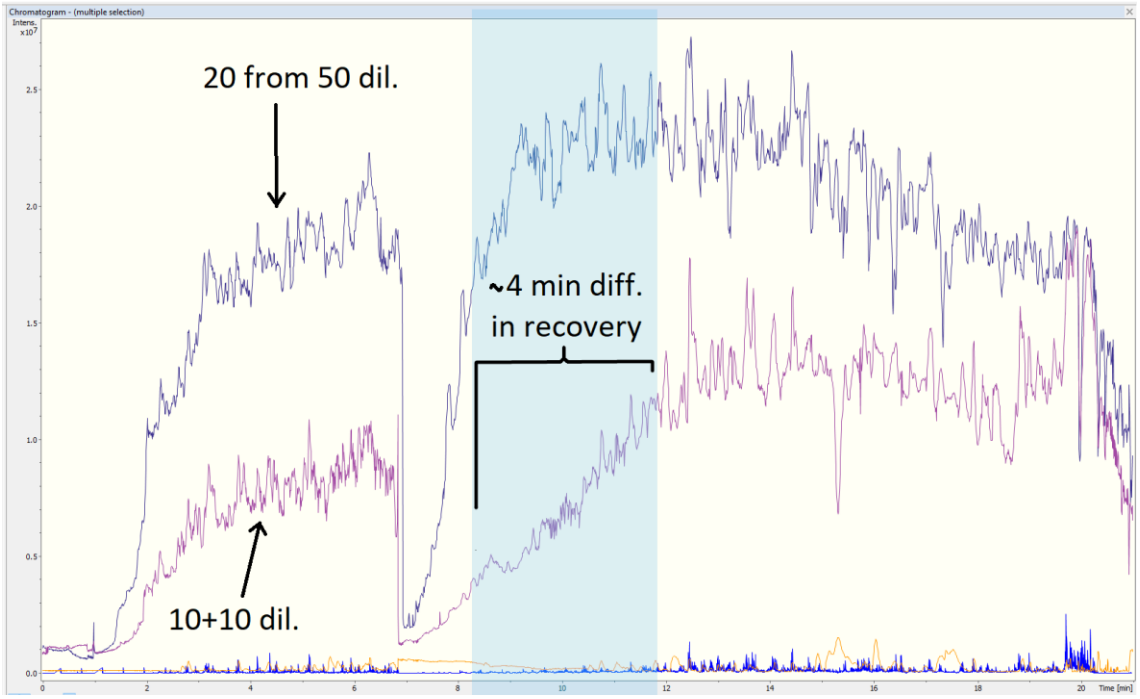

Figure 10. A chromatogram of two two-plate samples with different dilutions. The chromatogram's x-axis is time and y-axis is signal intensity. A two-plate sample that was resuspended in 50 µl of timsTOF buffer A and had 20 µl loaded straight into an Evotip (dark blue) has the same drop in signal intensity as the sample diluted 1:2 (pink), but the instrument is able to recover from the suspected over-abundance of biotin almost four minutes faster. The difference in recovery periods is highlighted in light blue.

The better TIC translates to the results as well. From the two-plate test samples, the required stable interactors could be found even after filtering, and thus were deemed successful. This dilution was also tested with protocol 1 samples. The PSM values were not as good as with the two-plate samples, but the required stable interactors were visible before filtering. Filtering took out four of the found 15 interactors from the one test sample done, thus the sample did not meet the required sample criteria. Still this dilution was chosen for the six replicate runs because signal intensity and the bait PSM values were promising. The results from these test runs can be seen in Table 5.

Table 5. Results of the different resuspension volume tests with samples made from two or one cell plates. After desalting, these samples were resuspended in 50 µl of timsTOF buffer A, and from that, 20 µl were loaded directly into an Evotip. The numbers in brackets depict the number of MEDs found before filtering, and plain one is the total after filtering.

	CDK8	
No. of plates:	2	1 (Bio-Rad)
Bait (PSM)	291	189
CCNC	6	9
Tot. No. MED	22 (25)	11 (15)
Criteria met?	Yes	No

Six more biological replicates were produced for protocols 3+2 and 1. Samples produced from three cell plates were diluted 1:4, and samples from two or one cell plates were resuspended in 50 µl from which 20 µl was loaded directly to an Evotip. The detailed results of the individual runs are listed in Appendix 1 and summarised in Table 6. The results match the corresponding dilution tests. Only the one-plate samples did not produce enough identifications required for TFIIH and MED subunits but the three- and two-plate samples fulfill the set criteria.

Table 6. Summary of the six replicate runs. Results from three and two cell plates meet the set criteria. After filtering, the one-plate samples had only two transcription factor II H subunits and 12 MED subunits identifiable from the data, and therefore the samples were deemed a failure, even if the required subunits were visible before filtering. The PSM values were too low to be considered reliable. In rows “Total number of TF2H” and “Total number of MED” the plain number is the number of identifications after filtering, and the number in brackets is the total before filtering. For CDK7 samples, the maximum number of TFIIH subunits possible is five, and for CDK8 samples the maximum number of MED subunits is 27.

	CDK7		
No. of plates:	3 (1:4)	2 (50→20)	1B (50→20)
Bait (PSM)	187|172|154|171|138|115	182|128|135|143|170|35	88|97|103|98|100|104
CCNH (PSM)	88|83|94|80|72|80	87|87|72|82|94|94	69|66|63|66|66|58
MAT1 (PSM)	74|53|69|73|63|58	55|45|68|57|67|75	47|52|41|42|39|41
Tot. No TF2H	5	3 (5)	2 (5)
Criteria met?	Yes	Yes	No
	CDK8		
No. of plates:	3 (1:4)	2 (50→20)	1B (50→20)
Bait (PSM)	252|150|159|180|151|166	291|198|110|118|204|171	189|187|174|167|140|157
CCNC (PSM)	10|6|6|7|4|7	6|9|8|3|7|8	11|9|3|3|6|2
Tot. No. MED	19 (27)	19 (27)	12 (25)
Criteria met?	Yes	Yes	No

4.3 Comparison of Six Replicate Runs with Statistical Mathematics

Based on the dilution testing, six more biological replicates were produced for the original five-plate protocol, protocols 3+2 and 1 (using the Bio-Rad columns only). The 1:20 diluted five-plate samples acted as the comparison point for protocols 3+2, and 1. The three-plate samples were diluted 1:4 from the usual 30 µl resuspension, while the two- and one-plate samples were resuspended in 50 µl of buffer A and from that 20 µl was loaded directly into an Evotip. CDK7 samples were compared to each other using the PSM values from the bait, cyclin H, MAT1, and TF2H4. For the CDK8 samples, the bait, cyclin C, and MED12 were used for the comparison.

First, the standard deviations of the different plate number PSMs were compared to the deviations of the five-plate samples with an F-test. The risk factor, α, was set to 0,05. Depending on the F-test results, either a t-test of equal variances or a t-test of unequal variances was performed. All the tests were done using Microsoft® Excel® 2016 (16.0.5278.1000). An overview of the results is in Table 7.

Table 7. A summary of the statistical comparison of five-plate sample PSM values to the other sample types. The numbers listed are p-values gained from either an F-test for equal variances, or a t-test (equal or unequal variances). The four p-values gained from unequal t-tests are marked with an asterisk, all the other t-tests are with equal variances. The un-highlighted p-values are not significantly different when compared to the original protocol. Values highlighted in green are significantly different, but the PSM values are better than in the five-plate samples, or there is less variance, and therefore the result is still acceptable. The values highlighted in red are significantly worse than the comparison point.

	3 (1:4)		2(50→20)		1(50→20)	
	F-test	t-test	F-test	t-test	F-test	t-test
CDK8	0,670	0,402	0,475	0,428	0,062	0,501
MED12	0,097	0,022	0,057	0,113	0,301	0,009
CCNC	0,273	0,086	0,354	0,078	0,835	0,346
CDK7	0,167	0,126	0,088	0,129	1,8E-04*	0,421
CCNH	0,228	0,225	0,307	0,110	0,014*	0,131
MAT1	0,181	0,348	0,402	0,669	0,020*	0,089
TF2H4	0,523	0,499	0,062	0,009	0,029*	0,029

Samples made from three plates did not have statistically significant differences, or in the case of MED12 had significantly better PSM values than the five-plate samples. With two plates, only in the case of TF2H4 were the PSM values significantly lower than in the samples made with the original protocol. The one-plate samples had very little variance in the PSM values, thus for the CDK7, CCNH, MAT1, and TF2H4 a t-test of unequal variances was performed. The results were similar to the two-plate samples, only the TF2H4

PSM values were significantly lower compared to the five-plate sample PSM values. Overall based on these tests, it is possible to conclude that on a 95 % confidence level the three-plate PSM values do not have statistically significant difference to the original five-plate samples, and therefore can be considered reliable. The two- and one-plate samples fail one aspect of the statistical comparison, but in general, especially the two-plate sample can be considered a good option if one has only limited resources for sample preparation.

5 Conclusions

The aim of this project was to find the smallest number of Ø 15 cm plates needed to produce an AP sample for the timsTOF Pro 2 mass spectrometer that would still be equal in quality when compared to a sample produced according to the original protocol. For this, modified protocols were designed and tested, and two options for the final amount of cell plates were found. Using protocol 3+2 (the protocol modified for samples made from three or two plates), samples produced from three Ø 15 cm cell plates yielded the best results out of all the experiments. Samples produced from two plates still passed all requirements set for an acceptable sample but the identification results were not as reliable as with three plates. Therefore it was concluded that protocol 3+2 is the best option for future timsTOF Pro 2 AP analysis. This means that instead of five Ø 15 cm cell plates, only three are needed per biological replicate. Considering purification materials, compared to the original Nature Protocols workflow protocol 3+2 uses 125 µl of streptavidin beads instead of 200 µl. Cell lysis is done with 2 ml instead of 3 ml of lysis buffer, and elution is done using 240 µl of 0,4 mM biotin solution, instead of 300 µl of 0,5 mM biotin solution. Wash buffers, and the number of washes and elutions remain unchanged.

One-plate samples produced with protocol 1 using the Bio-Rad columns were quite good, but did not meet all of the set sample criteria. The Pierce™ Micro-Spin columns from Thermo Fisher were ruled out of the experiments after the initial trial batch. The columns were difficult to handle and overall not suited for this AP protocol. The identification results from the Thermo Fisher column

samples were equal or worse than the ones produced with Bio-Rad columns, therefore no further testing was done using them. The main concern in performing AP in the Bio-Rad columns when using only 50 µl of streptavidin beads was that due to the short bead bed the sample would only flow through the column and the PoI would not be retained. Based on the results, the bead bed length was sufficient but not optimal for AP. An important note from protocol 1 was that instead of pouring the sample into the column, it was better to gently pipet the lysate to avoid disturbing the short bead bed.

The results of this optimisation will bring down the cost and labour of the AP protocol. The overall material savings are estimated to be around 40 %. For example, with a medium sized batch of ten samples with just the consumable plastics and purification beads the savings will be around 200 €. When this is scaled for one year (1000+ purifications), the savings from these materials alone will be 15 000–20 000 €. Decreasing the cell plate number allows for more samples to be produced at the same time as incubator space and work time can be utilised more efficiently.

In the future, an interesting idea for streamlining the protocol even more would be to find a way to utilise the brand new Biomek i5 Automated Workstation that was recently purchased from Berner (Finland). If for example the lysate purification step could be automatised, it would be a huge time saver. If the purification could be modified to fit a 96-well format, this would multiply the amount of samples purified at once and relieve the researchers to do other tasks in the meanwhile. In order for this to work, either a suitable commercial option with the purification material already integrated into the 96-well plate would need to be found, or another option would be to find a plate that could be loaded with the streptavidin purification beads with the Biomek. These options have already been looked at, but nothing interesting has yet surfaced.

Another point of interest for further optimisation of the AP protocol would be to optimise the amount of biotin used for sample elution during the lysate purification step. The current concentration of biotin in the elution buffer causes

a significant drop in signal intensity 7.0 minutes into the MS run. This problem was encountered particularly while running the more concentrated two- and one-plate samples. During maintenance procedures done after the practical part of this project was finished, Bruker confirmed the suspicions that the TIC signal intensity drop was indeed caused by an excess of biotin by inspecting extracted ion chromatograms from the two-plate runs. Decreasing the amount of biotin will affect how well the PoI elutes out of the purification column, thus optimising this would need careful testing and thorough validation. There was not enough time left for these tests in the schedule of this project, but they will be done later as I continue working for the Varjosalo laboratory.

To conclude, the three-plate protocol with a 1:4 dilution (5 µl of desalted sample resuspended in 30 µl of buffer A + 15 µl of timsTOF buffer A) will be the new standard purification protocol for all AP timsTOF Pro 2 samples. In case there is a sudden shortage in resources, for example in cell plates due to the global COVID-19 situation or otherwise, the two-plate method is a valid option but might not yield the best results. Using only one cell plate for sample preparation is not recommended as loss in identification results is likely. A simplified overview of the user-friendliness of the modified protocols is in Table 8.

Table 8. A summary of the user-friendliness and overall usefulness of the protocols. A green thumbs up sign means the cell plate amount performs well in the parameter, thumbs down means bad performance. A plus sign means there is room for improvement, the performance is not good or bad, and the skull sign means terrible performance.

	No. of plates:					
Parameter	5	4	3	2	1 (Bio-Rad)	1 (Thermo)
Time required to grow cells to confluency	+	+	+/👍	👍	👍	👍
Material efficiency (plastics, beads, etc.)	👎	👎	👍	👍	👍	👍
Purification practicality	+	+	👍	👍	👍	💀
No. of identifications	👍	👍	👍	👍	+	+
Wasteful surplus of purified sample?	👎	👎	👍	👍	👍	👍
Statistically similar to original protocol	👍	👍	👍	+	+	👎
Overall score	+	+	BEST! 👍	👍	+	👎

References

1 Liu, Xiaonan; Salokas, Kari; Weldatsadik, Rigbe; Gawriyski, Lisa & Varjosalo, Markku. 2020. Combined proximity labeling and affinity purification – mass spectrometry workflow for mapping and visualizing protein interaction networks. Online material. Nature Protocols. <https://doi.org/10.1038/s41596-020-0365-x>. Published 10.8.2020.

2 Helmenstine, Anne. 2019. Chemical Composition of the Human Body. Online material. ThoughtCo. <https://www.thoughtco.com/chemical-composition-of-the-human-body-603995>. Updated 2.12.2019. Read 20.3.2022.

3 What are Proteins? Online material. Genetic Science Learning Center. <https://learn.genetics.utah.edu/content/basics/proteins>. Read 20.3.2022.

4 Overview of Protein-Protein Interaction Analysis. Online material. Thermo Fisher Scientific. <https://www.thermofisher.com/fi/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-protein-protein-interaction-analysis.html>. Read 17.3.2022.

5 Proteomics. Online material. EMBL's European Bioinformatics Institute. <https://www.ebi.ac.uk/training/online/courses/proteomics-an-introduction/what-is-proteomics/>. Read 17.3.2022.

6 Kavallaris, Maria & Marshall, Glenn. 2005. Proteomics and disease: opportunities and challenges. Online material. The Medical Journal of Australia. <https://www.mja.com.au/journal/2005/182/11/proteomics-and-disease-opportunities-and-challenges>. Published 6.6.2005. Read 17.3.2022.

7 Keskitalo, Salla. 2017. CMGC Kinases and Cancer. Academic dissertation. Helsinki: Hansaprint. 37/2017.

8 Wang, Zhihong & Cole, Philip. 2015. Catalytic Mechanisms and Regulation of Protein Kinases. Online material. PubMed Central. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4373616/>. Published 25.3.2015.

9 Liu, Xiaonan; Salokas, Kari; Tamene, Fitsum; Jiu, Yaming; Weldatsadik, Rigbe; Öhman, Tiina & Varjosalo, Markku. 2018. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. Online material. Nature Communications. <https://doi.org/10.1038/s41467-018-03523-2>. Published 22.3.2018.

10 Allen, Benjamin & Taatjes, Dylan. 2015. The Mediator complex: a central integrator of transcription. Online material. Nature Reviews Molecular Cell Biology. <https://doi.org/10.1038/nrm3951>. Published 18.2.2015.

11 Protein Purification by Affinity Chromatography. Online material. Sino Biological. <https://www.sinobiological.com/resource/protein-review/protein-purification-by-ac>. Read 17.3.2022.

12 Evotips. Online material. Evosep. <https://www.evosep.com/evotip/>. Read 20.3.2022.

13 Meier, Florian; Brunner, Andreas-David; Koch, Scarlet; Koch, Heiner; Lubeck, Markus; Krause, Michael; Goebeck, Niels; Decker, Jens; Kosinski, Thomas; Park, Melvin; Bache, Nicolai; Hoerning, Ole; Cox, Jürgen; Räther, Oliver & Mann, Matthias. 2018. Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Online material. Science direct. <https://doi.org/10.1074/mcp.TIR118.000900>.

14 Gavina, Jennilee; Yao, Chunhw & Feng, Yong-Lai. 2014. Recent developments in DNA adduct analysis by mass spectrometry: A tool for exposure biomonitoring and identification of hazard for environmental pollutants. Online material. Science direct. <https://doi.org/10.1016/j.talanta.2014.06.050>.

15 Introduction to timsTOF and timsTOF Pro (Revision B). 2019. Manufacturer’s introduction material. Bruker.

Individual Results of the Six Replicate Runs

Table 1. Individual results from five-plate samples with a 1:20 dilution (1 µl of purified sample + 19 µl of timsTOF buffer A). PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. Red highlight indicates an interactor that was found in the original Nature Protocol publication but is not present in these samples. In the row “Total number of MED” the plain number is the number of identifications after filtering, and the number in brackets is the total before filtering.

	CDK7
Bait (PSM)	78|104|171|105|52|68
CCNH (PSM)	80|62|57|72|83|93
MAT1 (PSM)	78|47|37|47|67|70
TF2H1 (PSM)	24|20|31|16|3|3
TF2H2 (PSM)	1|15|15|11|5|1
TF2H3 (PSM)	8|7|7|5|1|0
TF2H4 (PSM)	10|15|22|15|8|5
TF2H5 (PSM)	4|6|10|3|1|2

	CDK8
Bait (PSM)	187|173|61|152|180|174
CCNC (PSM)	3|1|3|0|6|9
Tot. No MED	9 (20)
MED1 (PSM)	12|0|0|0|3|5
MED4 (PSM)	7|2|0|0|6|7
MED6 (PSM)	
MED7 (PSM)	
MED8 (PSM)	3|0|0|0|1|2
MED9 (PSM)	
MED10 (PSM)	4|0|0|0|3|2
MED11 (PSM)	1|0|0|0|1|3
MED12 (PSM)	20|17|15|0|13|17
MED13 (PSM)	4|0|0|0|0|2
MED14 (PSM)	12|7|3|0|4|12
MED15 (PSM)	3|0|0|0|0|1
MED16 (PSM)	4|5|3|0|8|14
MED17 (PSM)	6|4|4|0|1|8
MED18 (PSM)	
MED19 (PSM)	2|0|0|0|0|2
MED20 (PSM)	3|3|2|0|2|4
MED21 (PSM)	3|2|3|0|2|3
MED22 (PSM)	2|0|0|0|2|1
MED23 (PSM)	7|4|2|0|6|8
MED24 (PSM)	18|6|6|0|9|16
MED25 (PSM)	1|0|0|0|0|0
MED27 (PSM)	2|0|0|0|0|0
MED28 (PSM)	
MED29 (PSM)	
MED30 (PSM)	3|2|1|0|2|1
MED31 (PSM)	

Table 2. Individual results from three-plate samples with a 1:4 dilution (5 µl of purified sample + 15 µl of timsTOF buffer A). PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. The PSM values of different biological replicates are separated within the cell by a vertical line. In the row "Total number of MED" the plain number is the number of identifications after filtering, and the number in brackets is the total before filtering.

	CDK7
Bait (PSM)	187|172|154|171|138|115
CCNH (PSM)	88|83|94|80|72|80
MAT1 (PSM)	74|53|69|73|63|58
TF2H1 (PSM)	41|35|28|29|31|30
TF2H2 (PSM)	24|8|11|13|7|10
TF2H3 (PSM)	12|10|10|8|9|5
TF2H4 (PSM)	19|17|19|14|11|8
TF2H5 (PSM)	11|9|13|5|5|7

	CDK8
Bait (PSM)	252|150|159|180|151|166
CCNC (PSM)	10|6|6|7|4|7
Tot. No MED	19 (27)
MED1 (PSM)	40|14|12|12|22|16
MED4 (PSM)	12|9|8|9|10|9
MED6 (PSM)	6|2|3|3|2|3
MED7 (PSM)	1|1|2|1|0|0
MED8 (PSM)	3|10|7|9|5|4
MED9 (PSM)	2|2|0|3|4|0
MED10 (PSM)	1|5|8|6|7|7
MED11 (PSM)	4|5|2|3|3|5
MED12 (PSM)	65|29|21|24|30|29
MED13 (PSM)	18|26|28|24|17|16
MED14 (PSM)	38|27|24|22|23|30
MED15 (PSM)	12|7|5|7|9|7
MED16 (PSM)	13|15|11|14|12|11
MED17 (PSM)	12|10|8|9|15|9
MED18 (PSM)	2|0|0|0|0|0
MED19 (PSM)	6|1|1|2|1|2
MED20 (PSM)	4|1|2|1|4|1
MED21 (PSM)	6|4|5|9|6|4
MED22 (PSM)	2|5|3|3|4|2
MED23 (PSM)	20|15|14|12|16|9
MED24 (PSM)	28|27|30|23|22|15
MED25 (PSM)	6|1|3|3|3|4
MED27 (PSM)	7|5|12|11|12|3
MED28 (PSM)	1|3|1|4|0|1
MED29 (PSM)	2|0|0|0|0|0
MED30 (PSM)	8|4|6|5|8|5
MED31 (PSM)	2|1|1|0|2|2

Table 3. Individual results from two-plate samples, where the purified sample was resuspended in 50 µl of buffer A, from which 20 µl loaded directly into an Evotip. PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. The PSM values of different biological replicates are separated within the cell by a vertical line. In the row "Total number of MED" the plain number is the number of identifications after filtering, and the number in brackets is the total before filtering.

	CDK7
Bait (PSM)	182|128|135|143|170|170
CCNH (PSM)	87|87|72|82|94|94
MAT1 (PSM)	55|45|68|57|67|75
TF2H1 (PSM)	3|8|9|8|3|2
TF2H2 (PSM)	6|1|7|1|6|10
TF2H3 (PSM)	2|1|4|1|2|5
TF2H4 (PSM)	5|4|5|7|9|9
TF2H5 (PSM)	0|2|2|2|1|1

	CDK8
Bait (PSM)	291|198|110|118|204|171
CCNC (PSM)	6|9|8|3|7|8
Tot. No MED	19 (27)
MED1 (PSM)	28|14|11|8|6|6
MED4 (PSM)	14|7|8|7|11|2
MED6 (PSM)	2|1|0|0|1|0
MED7 (PSM)	0|1|0|2|2|1
MED8 (PSM)	4|7|5|3|3|2
MED9 (PSM)	7|7|0|0|1|0
MED10 (PSM)	5|7|5|6|4|5
MED11 (PSM)	2|2|1|4|2|3
MED12 (PSM)	58|43|15|17|19|14
MED13 (PSM)	24|26|13|19|11|1
MED14 (PSM)	44|35|16|19|19|9
MED15 (PSM)	12|9|7|7|3|2
MED16 (PSM)	16|18|6|10|13|9
MED17 (PSM)	15|12|7|11|8|2
MED18 (PSM)	2|0|0|0|0|0
MED19 (PSM)	0|3|2|3|2|2
MED20 (PSM)	5|1|2|3|3|3
MED21 (PSM)	3|3|2|3|0|1
MED22 (PSM)	3|3|3|3|3|4
MED23 (PSM)	17|24|18|24|19|15
MED24 (PSM)	31|28|20|19|23|17
MED25 (PSM)	5|8|0|2|2|1
MED27 (PSM)	14|9|4|4|3|3
MED28 (PSM)	5|2|0|2|0|0
MED29 (PSM)	2|0|0|0|0|0
MED30 (PSM)	3|3|2|3|4|4
MED31 (PSM)	5|3|1|0|1|0

Table 4. Individual results from one-plate samples, where the purified sample was resuspended in 50 µl of buffer A, from which 20 µl loaded directly into an Evotip. PSMs written in plain black are identifiable after filtering, but the ones highlighted with yellow are filtered out. Ones highlighted with green are initially filtered out but with further inspection based on their PSM values they can be accepted as real results. Red highlight indicates an interactor that was found in the original Nature Protocol publication but is not present in these samples. The PSM values of different biological replicates are separated within the cell by a vertical line. In the row "Total number of MED" the plain number is the number of identifications after filtering, and the number in brackets is the total before filtering.

	CDK7
Bait (PSM)	88|97|103|98|100|104
CCNH (PSM)	69|66|63|66|66|58
MAT1 (PSM)	47|52|41|42|39|41
TF2H1 (PSM)	3|0|0|2|0|1
TF2H2 (PSM)	6|5|6|7|5|4
TF2H3 (PSM)	0|1|2|0|1|0
TF2H4 (PSM)	6|3|8|3|6|4
TF2H5 (PSM)	2|1|4|2|1|0

	CDK8
Bait (PSM)	189|187|174|167|140|157
CCNC (PSM)	11|9|3|3|6|2
Tot. No MED	12 (25)
MED1 (PSM)	15|11|7|6|4|2
MED4 (PSM)	12|18|10|12|9|8
MED6 (PSM)	2|2|1|1|0|1
MED7 (PSM)	
MED8 (PSM)	2|0|1|0|0|0
MED9 (PSM)	0|1|1|0|0|0
MED10 (PSM)	6|7|4|4|3|1
MED11 (PSM)	1|2|1|1|0|1
MED12 (PSM)	51|40|29|26|24|20
MED13 (PSM)	6|9|6|5|1|0
MED14 (PSM)	19|21|3|8|3|3
MED15 (PSM)	6|8|0|2|2|0
MED16 (PSM)	7|10|6|7|7|6
MED17 (PSM)	10|10|4|6|5|6
MED18 (PSM)	2|1|0|0|0|0
MED19 (PSM)	
MED20 (PSM)	2|1|2|2|1|1
MED21 (PSM)	4|2|2|2|2|2
MED22 (PSM)	2|3|2|3|1|0
MED23 (PSM)	9|8|5|5|9|7
MED24 (PSM)	20|15|13|8|11|6
MED25 (PSM)	1|1|1|1|3|3
MED27 (PSM)	3|1|0|0|0|0
MED28 (PSM)	5|3|0|1|0|0
MED29 (PSM)	3|2|0|1|1|0
MED30 (PSM)	4|2|3|2|2|1
MED31 (PSM)	1|2|2|1|3|1