Visualization of vitamin A metabolism

Koenders, S.T.A.

Chapter 8

Design, Synthesis and Biological Evaluation of a Broadspectrum ALDH Probe

Published as S.T.A. Koenders *et al., ChemBioChem*, **21**, 1911-1917 (2020).

Introduction

Aldehyde dehydrogenases (ALDHs) are involved in many metabolic pathways and pathologies.[1–7] They are often upregulated in cancer and have been linked to cancer therapy resistance. For example, both ALDH1A1 and ALDH3A1 induce resistance to the commonly used chemotherapeutic, cyclophosphamide.[8,9]

A recent study has shown that only 7 out of 20 commonly used or recently developed ALDH1A1 inhibitors showed cellular activity (**Fig. 8.1**).[10] Two compounds (compound 5 and NCT-501) were selective for ALDH1A1, whereas five were pan-ALDH inhibitors (Disulfiram, Bay-11-7085, Aldi-2, CID 725345 and DEAB).[10] CB7 was claimed to be a selective ALDH3A1 inhibitor.[10]

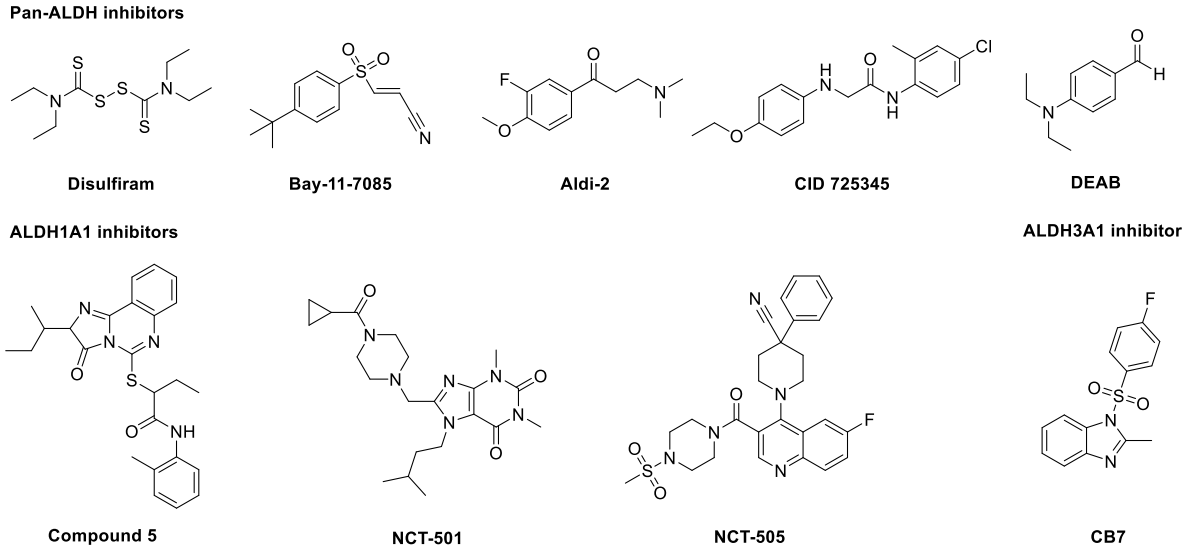


Fig. 8.1 | Chemical structures of *in situ* active ALDH inhibitors

Although *in vitro* assays are useful for the rapid screening of large compound libraries, the biochemical ALDH assays have two main drawbacks. First, the conditions in the biochemical assays differ from the cellular assays, which result in altered potency and selectivity profiles. For example, most biochemical assays are performed at pH > 7.4 and reducing agents, such as dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP), are added. This will change the reactivity of catalytic cysteine. Second, the biochemical assays do not take into account factors, such as cell permeability and metabolic stability.

So far, no methods exists to study the target-engagement of multiple ALDHs present in a biological sample simultaneously. Activity-based protein profiling (ABPP) is a powerful technique capable of determining the selectivity profile of drug candidate against an enzyme family in their native cellular environment.[11] This technique relies on activity-based probes, which are tailored to the enzyme family of interest and that react via its electrophilic warhead with a nucleophile within the active site of the enzyme. **Chapters 3-6** described the development of a first-in-class retinal-based probe **LEI-945** for the profiling of retinaldehyde dehydrogenases.[12] With **LEI-945**, six of the 19 enzymes in the ALDH family in humans, ALDH1A1, ALDH1A2, ALDH1A3, ALDH2, ALDH3A2 and ALDH3B1 can be detected from cell extract. The synthesis of **LEI-945**, which can be found in **Chapter 3**, is complex and challenging. It was envisioned that by modifying the reported covalent pan-ALDH inhibitor, Aldi-2[20] (**Fig. 8.1)**, an easily accessible broadspectrum probe for the ALDH family could be made.

This chapter describes the design, synthesis, biological validation and application of **STA-55** as a broadspectrum probe for the family of aldehyde dehydrogenases. Chemical proteomics showed that **STA-55** could be used to enrich both ALDH1A1 and ALDH3A1 in A549 lung cancer cells. Competitive proteomics with **STA-55** was performed to determine the selectivity profiles of known ALDH inhibitors DEAB[13], NCT-505[14] and CB7.[15] These results showed that **STA-55** can be used to identify therapy resistance biomarkers in cancer and to validate target-engagement of ALDH drug candidates.

Results and discussion

Synthesis and biological evaluation of a broadspectrum activity-based ALDH probe

An activity-based probe (ABP) consists of a reactive group (termed 'warhead' and often an electrophile), a recognition handle and a ligation handle. Aldi-2 already incorporates a masked warhead which after liberation reacts with the catalytic cysteine of the ALDH enzyme (**Fig. 8.2**).[16] Based on the reported structure-activity relationship studies, a synthetic strategy for the introduction of an azide ligation handle at the position of the methoxy-substituent was developed.

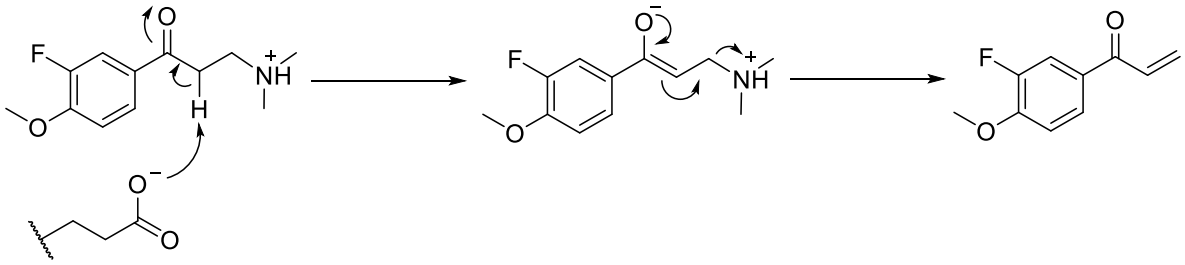
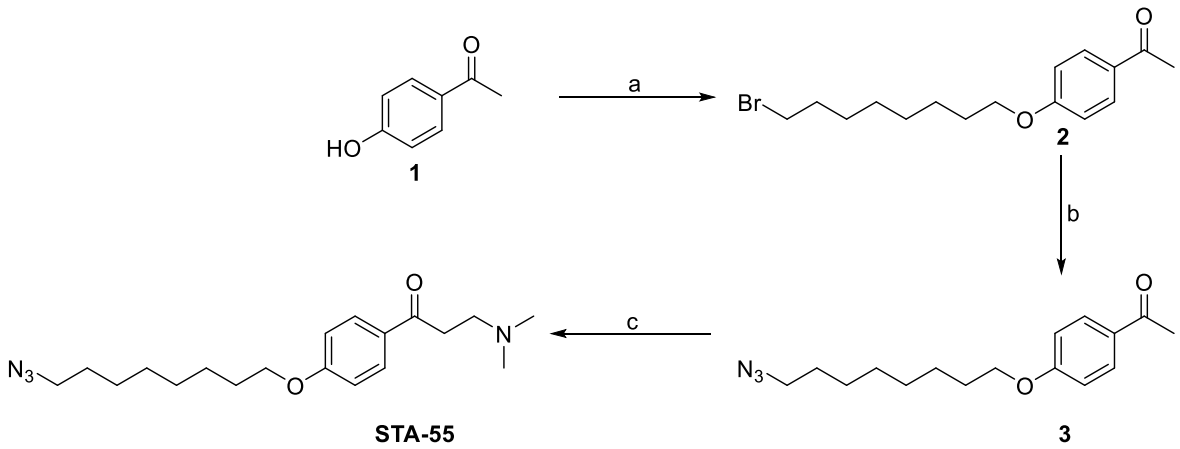


Fig. 8.2 | Proposed Aldi-2 warhead deprotection mechanism

The synthesis of probe **STA-55** started from commercially available 4-hydroxyacetophenone **1** (**Scheme 8.1**). Reaction of **1** with 1,8-dibromooctane and potassium carbonate provided **2** in 82% yield. Treatment of **2** with sodium azide led to the substitution of the bromine for an azide, yielding compound **3** in 93% yield. Finally, Mannich reaction of **3** with dimethylamine and paraformaldehyde gave tertiary amine **STA-55** in 21% yield.



Scheme 8.1 | Synthesis of probe STA-55. Reagent and conditions: a) 1,8-dibromooctane, K_2CO_3, acetone, 56 °C, 18 h, 82%; b) NaN_3, DMF, 80 °C, 18 h, 93%; c) $(CH_3)_2NH{\cdot}HCl$, paraformaldehyde, HCl, EtOH, 90 °C, 18 h, 21%.

To determine whether **STA-55** interacts with the catalytic cysteine of an ALDH enzyme, recombinant ALDH1A3^{WT} and catalytically inactive ALDH1A3^{C314A} were overexpressed in human osteosarcoma U2OS cells. Treatment with **STA-55** (1 μM, 1 h), lysis and Cu(I)-catalysed azide-alkyne [2+3] cycloaddition with a Cy5-alkyne resulted in a fluorescent band around 55 kDa after resolving the samples by SDS-PAGE and fluorescent scanning, suggesting the ability of **STA-55** to interact with ALDHs. The disappearance of this band in the catalytically inactive C314A mutant indicates that **STA-55** has reacted covalently with the catalytic nucleophile of ALDH1A3 (**Supplementary Fig. 8.1**).[17]

To determine which members of the ALDH family are targeted by ABP **STA-55**, a label-free quantitative proteomics experiment was performed in the non-small-cell lung cancer cell line A549, which expresses high levels of ALDH activity.[18,19] **STA-55**-treated A549 cells (10 μM, 1 h) were harvested, lysed and the covalently bound enzymes conjugated with a biotin-alkyne. The probe labelled proteins were subsequently enriched using streptavidin-beads and several washing steps to remove unbound proteins. On-bead digestion was followed by protein identification and quantification by mass spectrometry. In this way, 259 were identified (fold-change > 2.0; p-value < 0.05) (**Fig. 8.3a, b**). Six identified proteins belong to the ALDH family: ALDH1A1, ALDH1B1, ALDH2, ALDH3A1, ALDH3A2 and ALDH3B1.

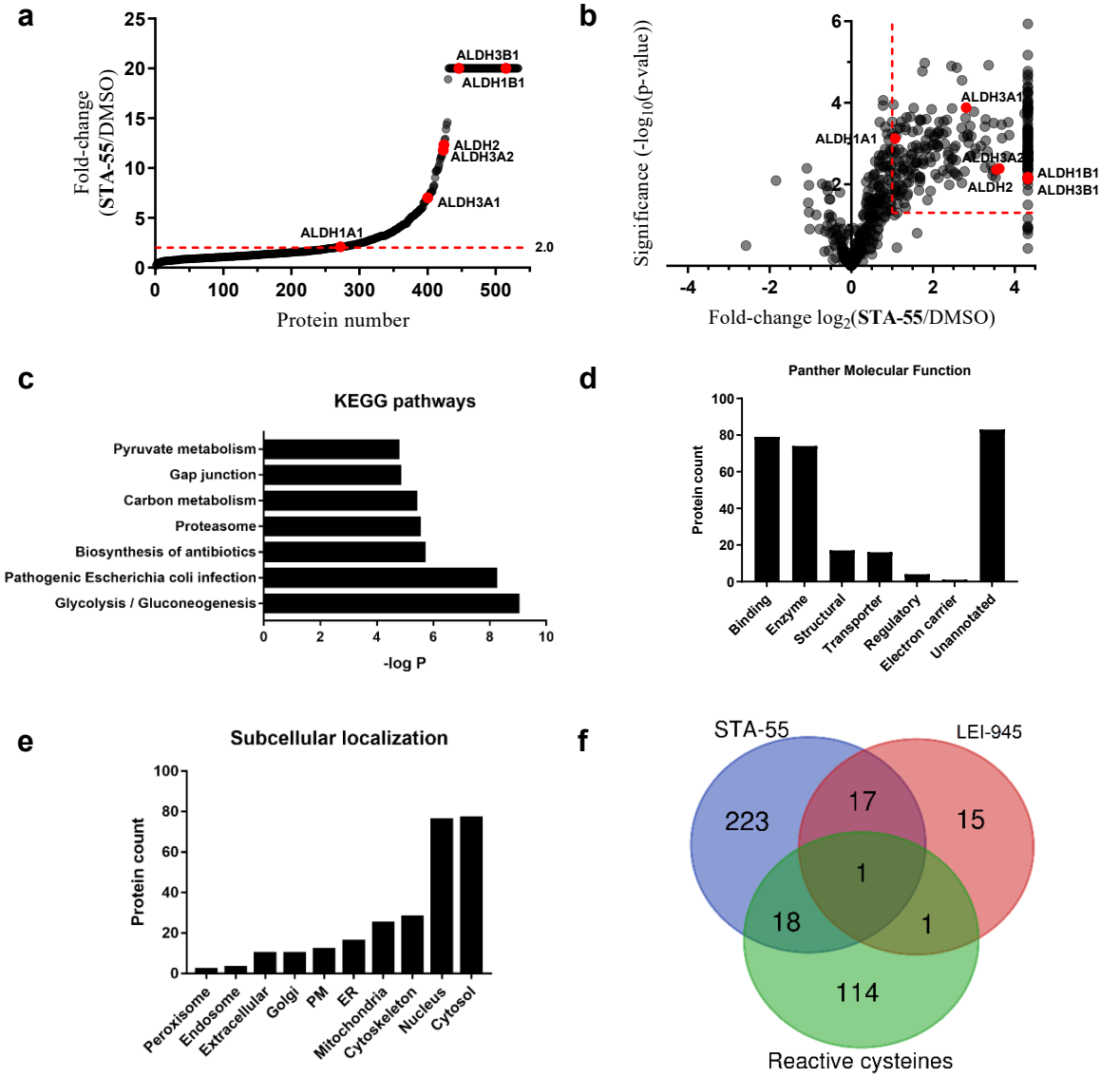


Fig. 8.3 | Chemical proteomics with broadspectrum ALDH probe STA-55. **a**, Fold-change (**STA-55**/DMSO) plot for total proteins identified in chemical proteomics experiment with probe **STA-55** (10 μM). Red lines indicate the threshold fold-change of 2-fold enrichment and the maximum fold-change is set at 20. Red dots represent significantly enriched ALDH enzymes. **b**, Volcano plot for total proteins identified in chemical proteomics experiment with probe **STA-55** (10 μM). Red lines indicate threshold values marking significantly enriched proteins. Red dots represent significantly enriched ALDH enzymes. For parts **a** and **b**, data are from $N = 3$ experiments (biological replicates). **c**, Top 7 pathways enriched in the group of significantly enriched proteins as determined by screening on the KEGG database. **d**, Biological functions annotated to significantly enriched proteins by the PANTHER database. **e**, Subcellular localization of significantly enriched proteins as annotated by the UniProt database. **f**, Venn diagram showing the distribution of proteins enriched by the broadspectrum **STA-55** probe, retinoid probe **LEI-945** and proteins containing highly reactive cysteines as determined by Weerapana *et al*.

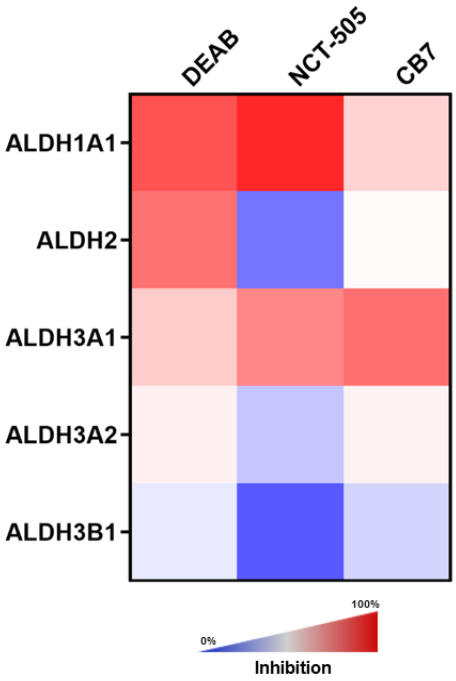
Comparison with the Expression Atlas (**Supplementary Fig. 8.2**)[20,21] showed that all the ALDH enzymes expressed in the A549 cell line were detected by the **STA-55** broadspectrum probe, whereas the more specific **LEI-945** probe did not detect ALDH1B1 and ALDH3A1 in this cell line. The proteins significantly enriched by **STA-55** were further analysed using the KEGG[22], PANTHER[23] and UniProtKB databases.[24] Proteins in the glycolysis/gluconeogenesis, biosynthesis of antibiotics and carbon metabolism (ALDHs, lactate dehydrogenases and phosphofructokinases), pathogenic *E. Coli* infection and gap junction (tubulins) and the proteasome were identified via the KEGG pathway database (**Fig. 8.3c**).

The majority of proteins identified possess enzyme activity or have specific protein interaction partners (**Fig. 8.3d**). The subcellular localization showed that the bulk of the enriched proteins are derived from the cytosol and nucleus (**Fig. 8.3e**). Comparing the probe targets of **STA-55** and **LEI-945** shows that 18 proteins enriched by **LEI-945** are also detected by **STA-55** (**Fig. 8.3f**). Although Aldi-2 was claimed to be selective for ALDHs over other enzymes containing reactive cysteines, such as cysteine proteases, alcohol dehydrogenases and tyrosine phosphatases, **STA-55** did enrich 19 enzymes containing reactive cysteines, including ATPases, hydrolases, cysteine proteases, disulphide isomerases and kinases (**Fig. 8.3f**).[16] Taken together, these data indicate that **STA-55** can be used as a broadspectrum activity-based probe for aldehyde dehydrogenases, including the known cancer resistance biomarkers, ALDH1A1 and ALDH3A1.[8,9]

Competitive ABPP of ALDH inhibitors using STA-55

Having established that **STA-55** significantly enriches a broad range of ALDHs in the A549 lung cancer cell extracts, *in situ* selectivity profiling of three known ALDH inhibitors (DEAB[10,13], NCT-505[14] and CB7[15]) was performed. DEAB is a pan-ALDH inhibitor regularly used as control compound in ALDH activity assays.[10,13] NCT-505 is a recently reported selective ALDH1A1 inhibitor[14] and CB7 is reported as a selective ALDH3A1 inhibitor.[15]

a



b

	DEAB	NCT-505	CB7
ALDH1A1	7.2±0.07	8.2	NI
ALDH2	6.0±1.2	4.7	4.7
ALDH3A1	5.2	NI	6.6±0.2
ALDH3A2	ND	ND	ND
ALDH3B1	ND	ND	ND

Fig. 8.4 | Competitive ABPP of STA-55 with ALDH inhibitors. a, Heatmap showing the selectivity profile of pan-ALDH inhibitor DEAB, ALDH1A1 selective inhibitor NCT-505 and ALDH3A1 selective inhibitor CB7 as determined by competitive ABPP with **STA-55** (1 µM). *N* = 4 experiments (biological replicates). **b**, Reported pIC50 values of DEAB[10,13], NCT-505[14] and CB7[10,15] for the ALDHs identified using **STA-55**. When divergent values have been reported, data are represented as mean values ± SD. NI = no inhibition and ND = no data.

A549 cells were preincubated for 30 minutes with the inhibitor (10 µM) after which **STA-55** (1 µM) was added and incubated for one hour. After which the chemical proteomics protocol described before was followed. The ALDH selectivity profiles obtained are shown in **Fig. 8.4a**. Of note, ALDH1B1 was not detected in the competitive ABPP experiment, probably because a 10-fold lower concentration of **STA-55** was used and ALDH1B1, one of the lowest abundant ALDH enzymes in this cell line, did not reach the detection threshold (**Supplementary Fig. 8.3**).

DEAB inhibited ALDH1A1, ALDH2, ALDH3A1 and ALDH3A2, which is in agreement with previously reported results (**Fig. 8.4b**).[13] NCT-505 inhibited ALDH1A1, ALDH1A3 and ALDH3A1. CB7 appeared to be more promiscuous than reported and inhibited ALDH3A1, ALDH1A1, ALDH2 and ALDH3A2 under these conditions. These results challenge the reported selectivity for ALDH1A1 and ALDH3A1 of NCT-505[14] and CB7[15], respectively.

For both inhibitors these selectivity claims are based on biochemical substrate assays using purified enzymes. However, the activity of isolated enzymes in a biochemical assay does not necessarily reflect their activity in a cellular environment. It can therefore be argued that the selectivity profiles derived via an *in situ* competitive ABPP method provide a more accurate representation of the cellular target-engagement of a drug candidate.[25] The ability of CB7 to sensitize lung cancer cells to cyclophosphamide treatment,[15] can therefore not be solely attributed to its inhibition of ALDH3A1, but might actually require dual inhibition of ALDH1A1 and ALDH3A1. From this point of view, NCT-505 has therapeutic potential as a dual inhibitor of these cancer therapy resistance biomarkers.

These results show the ability of **STA-55** to be used for target identification and engagement studies using competitive ABPP of new drug candidates for ALDHs. **STA-55** is easily accessible compared with the synthetically challenging **LEI-945**. **STA-55** is also capable of enriching all ALDH enzymes present in the A549 lung cancer cells, including the cancer biomarker ALDH3A1, where **LEI-945** is more specifically targeted at the retinaldehyde dehydrogenases.

Conclusion

In summary, this chapter describes the design, development and biological validation of **STA-55**, a broadspectrum activity-based probe for the family of aldehyde dehydrogenases. Certain ALDHs are often upregulated in cancer and confer therapy resistance. This probe enables the identification and quantification of these cancer biomarkers using chemical proteomics. Furthermore, it shows the ability of **STA-55** to be used in cellular target identification and target engagement studies, and proposes that the probe may be used to facilitate early drug discovery studies aimed at the identification of selective and tissue-permeable ALDH inhibitors.

Acknowledgements

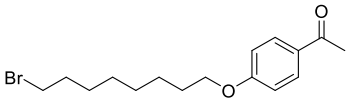
Eva van Rooden and Bogdan Florea are kindly acknowledged for the mass spectrometry analysis.

Experimental procedures

Synthetic methods

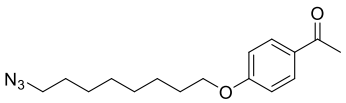
General remarks. All reactions were performed using oven or flame-dried glassware and dry solvents. Reagents were purchased from Sigma Aldrich, Acros, Biosolve, VWR, Fluka, Fischer Scientific and Merck and used as received unless stated otherwise. Inhibitors NCT-505 and CB7 were prepared as previously described.[14,15] Tetrahydrofuran (THF) and *N,N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves before use. All moisture sensitive reactions were performed under a nitrogen atmosphere. TLC analysis was performed using Merck aluminum sheets (TLC silica gel 60/Kieselguhr F_{254}). Compounds were visualized using a solution of $KMnO_4$ (7.5 g), K_2CO_3 (50 g), 10% NaOH (6 mL) in H_2O (1 L). Column chromatography was performed using Screening Device B.V. silica gel (particle size 40 – 63 µm, pore diameter of 60 Å) with the indicated eluents. ^{1}H- and ^{13}C-NMR spectra were recorded on Bruker AV-400 (400 MHz and 101 MHz, respectively) or Bruker AV-500 MHz (500 MHz and 125 MHz, respectively) using $CDCl_3$ as solvent. Chemical shifts are reported in ppm (δ) relative to the residual solvent peak or tetramethylsilane. Coupling constants are given in Hz. High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electrospray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL/min, capillary temperature 250 °C) with resolution R = 60000 at m/z 400 (mass range m/z = 150 – 2000) and dioctyl phthalate (m/z = 391.28428) as a "lock mass", or with a Synapt G2-Si (Waters), equipped with an electrospray ion source in positive mode (ESI-TOF), injection via NanoEquity system (Waters), with LeuEnk (m/z = 556.2771) as "lock mass". Eluents used: MeCN:H_2O (1:1 v/v) supplemented with 0.1% formic acid. The high-resolution mass spectrometers were calibrated prior to measurements with a calibration mixture (Thermo Finnigan).

1-(4-((8-Bromooctyl)oxy)phenyl)ethan-1-one (2):



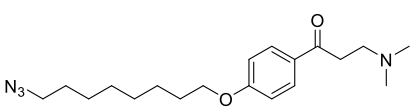
To a mixture of 1,8-dibromooctane (4.1 mL, 22 mmol) and K_2CO_3 (1.1 g, 8.1 mmol) in acetone (40 mL) at 56 °C was added dropwise 4-hydroxyacetophenone (1.0 g, 7.3 mmol) in acetone (20 mL). The reaction mixture was stirred for 18 hours at 56 °C. The reaction mixture was then allowed to cool down, filtered and concentrated under reduced pressure. The residue was then dissolved in EtOAc, the organic layer washed with H_2O and brine, dried with Na_2SO_4, filtered and concentrated under reduced pressure. Purification of the residue by column chromatography (DCM/pentane) afforded the title compound **2** (2.0 g, 6.0 mmol, 82%) as a white solid. R_f(50% DCM in pentane) = 0.5. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.93 (d, *J* = 4.8 Hz, 2H), 6.92 (d, *J* = 4.8 Hz, 2H), 4.02 (t, *J* = 6.8 Hz, 2H), 3.41 (t, 6.8 Hz, 2H), 2.56 (s, 3H), 1.83 (m, 4H), 1.47 (m, 4H), 1.39 (m, 4H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 196.8, 163.1, 130.6, 130.1, 114.1, 68.1, 34.0, 32.7, 29.1, 29.0, 28.6, 28.1, 26.3, 25.9.

1-(4-((8-Azidooctyl)oxy)phenyl)ethan-1-one (3):



To a solution of compound **2** (0.50 g, 1.5 mmol) in DMF (5 mL) under Ar was added sodium azide (0.30 g, 4.6 mmol). The reaction mixture was stirred for 18 hours at 80 °C and then allowed to cool down. H_2O was added and the aqueous layer was extracted with Et_2O. The combined organic layers were washed with brine, dried with $MgSO_4$, filtered and concentrated under reduced pressure. Purification of the residue by column chromatography (pentane/DCM) afforded the title compound **3** (0.41 g , 1.4 mmol, 93%) as a colorless liquid. R_f(50% DCM in pentane) = 0.45. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.93 (d, *J* = 4.8 Hz, 2H), 6.92 (d, *J* = 4.8 Hz, 2H), 4.02 (t, *J* = 6.8 Hz, 2H), 3.26 (t, 6.8 Hz, 2H), 2.55 (s, 3H), 1.81 (m, 2H), 1.61 (m, 2H), 1.47 (m, 2H), 1.39 (m, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 196.8, 163.0, 130.5, 130.1, 114.1, 68.1, 51.4, 29.1, 29.0, 28.8, 26.6, 26.3, 25.8.

1-(4-((8-Azidooctyl)oxy)phenyl)-3-(dimethylamino)propan-1-one (STA-55):



To a solution of compound **3** (164 mg, 0.57 mmol) in EtOH (2 mL) under Ar was added dimethylamine hydrochloride salt (92 mg, 1.1 mmol), paraformaldehyde (34 mg, 1.1 mmol) and one drop of conc. HCl. The reaction mixture was stirred at 90 °C for 18 hours and then allowed to cool down. Purification of the reaction mixture by column chromatography (DCM/MeOH) recovered compound **3** (127 mg, 0.43 mmol, 75%) and afforded the title compound **STA-55** (42.2 mg, 0.12 mmol, 21%) as yellow solid. R_f(10% MeOH in DCM) = 0.5. ^{1}H (400 MHz, $CDCl_3$): δ 7.96 (d, *J* = 8.8 Hz, 2H), 6.92 (d, *J* = 8.8 Hz, 2H); 4.03 (t, *J* = 6.4 Hz, 2H), 3.68 (t, *J* = 6.4 Hz, 2H), 3.51 (t, *J* = 6.4 Hz, 2H), 3.27 (t, *J* = 7.2 Hz, 2H), 2.84 (s, 6H), 1.81 (m, 2H), 1.59 (m, 2H), 1.46 (m, 2H), 1.37 (m, 6H).

^{13}C NMR (100 MHz, $CDCl_3$): δ 194.1, 163.9, 130.6, 128.2, 114.4, 68.2, 52.8, 51.4, 43.3, 33.4, 29.1, 29.0, 28.9, 28.7, 26.6, 25.8. HRMS (ESI) m/z: $[M + H]^+$ calculated for $C_{19}H_{30}O_2$: 347.24415, found 347.24433.

In situ labeling procedure

ALDH plasmids. For the preparation of the different constructs, full length human cDNA was purchased from Source Bioscience and cloned into mammalian expression vector pcDNA3.1, containing genes for ampicillin and neomycin resistance. ALDH1A3 was cloned into pcDNA3.1. A FLAG-linker was cloned into the vector at the C-terminus of ALDH1A3. Two step PCR mutagenesis was performed to substitute the active site cysteine for an alanine in the hALDH1A3-FLAG to obtain hALDH1A3-C314A-FLAG. All plasmids were grown in XL-10 Z-competent cells and prepped (Maxi Prep, Qiagen). The sequences were confirmed by sequence analysis at the Leiden Genome Technology Centre.

Cell culture. U2OS cells were grown in DMEM with stable glutamine and phenol red with 10% New Born Calf serum, penicillin and streptomycin at 37 °C and 7% CO_2. A549 cells were grown in DMEM with stable glutamine and phenol red with 10% New Born Calf serum, penicillin and streptomycin at 37 °C and 5% CO_2. Medium was refreshed every 2-3 days and cells were passaged twice a week. Cell lines were purchased from ATCC and were regularly tested for mycoplasma contamination. Cultures were discarded after 2-3 months of use.

Transient transfection of U2OS cells. One day prior to transfection $4*10^5$ U2OS cells were seeded in a 6-wells plate. Cells were transfected by addition of a 3:1 mixture of polyethyleneimine (6 µg) and plasmid DNA (2 µg) in 200 µL serum free medium per well. The medium was refreshed after 24 hours and after 48 hours the cells were used for subsequent assays.

***In situ* activity-based protein profiling.** Growth medium from cells grown in 6-wells plate was removed and 1 mL serum free medium containing probe **STA-55** (10 µM, 0.1% DMSO) was added. The cells were then incubated for 1 hour. For competitive ABPP cells were first incubated with vehicle or inhibitor (10 µM, 0.1% DMSO) for 30 minutes followed by **STA-55** (1 µM final concentration) for 1 hour. The medium was then removed, the cells were washed with 2 mL PBS and then harvested in 1 mL PBS using a cell scraper. The cells were moved to an eppendorf tube and the suspension was centrifuged for 5 minutes at 1200 rpm. The PBS was removed and the samples snap frozen and stored at -80 °C until use.

CuAAC reaction and in-gel fluorescence analysis. Cell pellets were thawed on ice, lysed by addition of ice-cold lysis buffer (MilliQ, 1x protease inhibitor cocktail (Roche c*O*mplete EDTA free)) and incubated on ice (15-30 min). The protein concentration was determined by a Quick Start™ Bradford Protein assay (Bio-Rad). The protein fractions were diluted to a total protein concentration of 1 mg/mL. From each sample 40 µL was taken and treated with 5 µL from a freshly prepared "click" mixture containing 9 mM $CuSO_4$ (2.5 µL/sample, 18 mM in H_2O), 45 mM NaAsc (1.5 µL/sample, 150 mM in H_2O), 1.8 mM THPTA (0.5 µL/sample, 18 mM in DMSO) and 9 µM Cy5-alkyne (0.5 µL/sample, 90 µM in DMSO from Thermo Fischer Scientific). The samples were incubated for 1 hour at 37°C and then 15 µL 4x SDS page sample buffer was added. The samples were denatured at 100 °C for 5 minutes. 8 µg per sample was resolved on a SDS-PAGE gel (10% acrylamide, 180V, 75 min). Gels were visualized with a ChemiDoc XRS (Bio-Rad) using Cy3 and Cy5 multichannel settings (605/50 and 695/55, filters respectively) and stained with Coomassie or transferred to a 0.2 µm polyvinylidene difluoride membranes by Trans-Blot Turbo™ Transfer system (Bio-Rad) after scanning. Fluorescence was normalized to Coomassie staining or to FLAG-tag signal and quantified with Image Lab (Bio-Rad).

Western Blotting. Proteins were transferred to a 0.2 µm polyvinylidene difluoride membranes by Trans-Blot Turbo™ Transfer system (Bio-Rad). Membranes were washed with TBS (50 mM Tris, 150 mM NaCl), washed with TBS-T (50 mM Tris, 150 mM NaCl, 0.05% Tween 20) and then blocked with 5% w/v milk powder in TBS-T for 1 hour at room temperature. Membranes were then incubated with primary antibody in 5% milk TBS-T (α-FLAG: 1 h, RT), washed three times with TBS-T, incubated with matching secondary antibody in 5% milk TBS-T (1 h, RT) and washed with TBS-T and TBS. The blot was developed in the dark using an imaging solution (10 mL Luminol, 100 µL ECL enhancer and 3 µL 30% H_2O_2) and chemiluminescence was visualized using a ChemiDoc XRS (Bio-Rad). The signal was normalized to Coomassie staining and quantified with Image Lab (Bio-Rad). Primary antibody: monoclonal mouse anti-FLAG (1:5000, Sigma-Aldrich, F3165). Secondary antibody: HRP-coupled goat-anti-mouse (1:5000, Santa Cruz, sc2005).

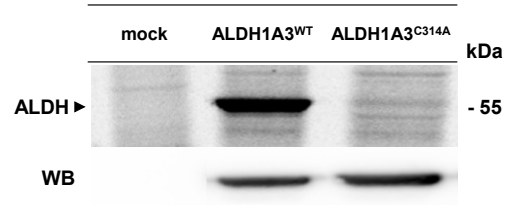
In situ activity-based proteomics

Sample preparation. Protocol adapted from previously described procedure.[26] Cells were treated *in situ*, harvested, lysed and adjusted to 1 mg/mL protein concentration as described above. 250 µL was taken from each sample and to this 25 µL freshly prepared "click" mixture containing 1 mM $CuSO_4$ (2.5 µL/sample, 100 mM in H_2O), 5 mM NaAsc (1.25 µL/sample, 1 M in H_2O), 0.4 mM THPTA (1 µL/sample, 100 mM in DMSO), 40 µM biotin-alkyne (2.5 µL/sample, 4 mM in DMSO) and MilliQ (17.75 µL/sample) was added. Samples were incubated for 1 hour at 37 °C while shaking (300 rpm). Excess click reagents were then removed by chloroform/methanol precipitation as described before. Precipitated proteomes were then suspended in urea buffer (250 µL, 6 M urea and 25 mM ammonium bicarbonate), DTT (2.5 µL, 1 M) was added and the mixture was then incubated for 15 min at 65 °C while shaking (600 rpm). The samples were then allowed to cool down to RT and then alkylated by addition of iodoacetamide (20 µL, 0.5 M) for 30 minutes at RT in the dark. Addition of SDS (70 µL, 10% (v/v)) was followed by heating at 65 °C for 5 minutes. For each sample 50 µL 50% slurry of Avidin-Agarose from egg white (Sigma-Aldrich) was washed three times with PBS by centrifugation and transferred in PBS (1 mL) to a 15 mL tube. To this another 2 mL of PBS was added followed by the corresponding proteome sample. The beads were incubated with the proteome for 2 hours at room temperature using an overhead shaker. The beads were then isolated by centrifugation (2 min, 2500 g), washed with SDS in PBS (0.5% (w/v)) and washed three times with PBS. The beads were then transferred to low-binding Eppendorf tubes and proteins were digested overnight at 37 °C and 950 rpm shaking in 250 µL digestion buffer (100 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 2% acetonitrile and 0.5 µg sequencing grade trypsin (Promega)). Digestion was stopped by addition of formic acid (12.5 µL) and the beads filtered off by centrifugation (2 min, 600 g) using a Bio-Spin column (Bio-Rad). Samples were then desalted using stage tips, collected in low-binding Eppendorf tubes, concentrated using a SpeedVac (Eppendorf) and stored at -20 °C until reconstitution before measurement.[27] All samples were prepared in at least three biological replicates.

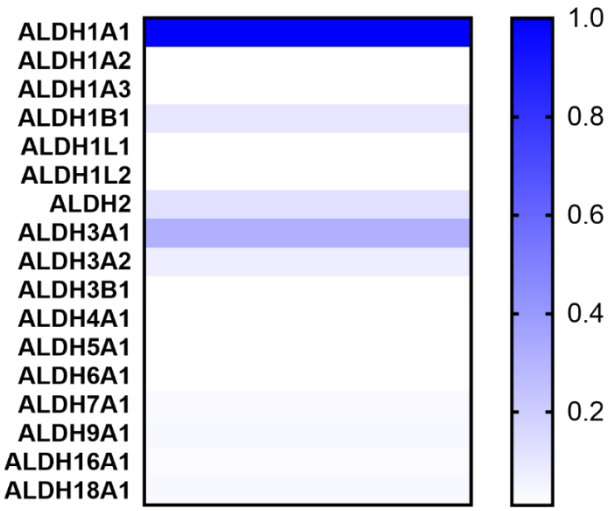
LC-MS/MS measurement and analysis. Samples were reconstituted in LC-MS sample solution (50 µL, MilliQ, 3% acetonitrile/0.1% formic acid/20 fmol/µL enolase). Samples were then analysed using a NanoACQUITY UPLC System (Waters) coupled to a SYNAPT G2-Si high-definition mass spectrometer (Waters) as previously described.[26,28] Of each sample 5 µL was loaded on a nanoEASE™ M/Z Symmetry C18 trap column (particles 5 µm, 100 Å , 180 µm x 20 mm, Waters) with 0.1% formic acid and separated on an nanoEASE™ M/Z HSS C18 T3 analytical column (particles 1.8 µm, 75 µm x 250 mm, Waters) heated at 80 °C. A multistep gradient running from 5-40% acetonitrile containing 0.1% formic acid during a 70 minute method at 300 nL/min was used to achieve peptide separation. Survey scans (m/z 50-2000 Da) were acquired in the Synapt with a scan time of 0.6 seconds in positive resolution mode. The collision energy was set to 4 V in the trap cell for low-energy MS mode. For the elevated energy scan, the transfer cell collision energy was ramped using drift-time specific collision energies. The lock mass was sampled every 30 seconds. MS raw files were analysed with ProteinLynx Global SERVER (PLGS, v3.0.3, Waters). The MS^E identification was also performed with PLGS using the human proteome from Uniprot (uniprot-homo-sapiens-trypsin-reviewed-2016-08-29.fasta). The following parameter settings were used: low energy threshold 150 counts, elevated energy threshold 30, peptide and protein FDR 1%, enzyme specificity trypsin, max missed cleavages max 2, variable modification methionine oxidation, fixed modification carbamidomethylation cysteine, at least: fragments/peptide 2, fragments/protein 5, peptides/protein 1 and number of peptides to measure per protein 3. For label-free quantification ISOQuant (v1.5) was used.[29,30] Data were filtered to retain only proteins with two or more reported unique peptides and quantified in at least 3 replicates of the positive control (probe-treated). Proteins were designated as significantly enriched by the probe when they showed 2-fold enrichment in quantification value when comparing negative control (vehicle-treated) with positive control (probe-treated) samples and probability as determined by a Student's *t* test (<0.05).

Heatmap Competitive ABPP analysis. Only significantly enriched ALDH enzymes were selected for analysis. The mean raw LFQ intensities from quadruplicate measurements were normalized to DMSO (= 0) and maximal LFQ **STA-55** (= 1) for each protein individually. The heatmap was prepared using Graphpad Prism® 7 (Graphpad Software Inc.).

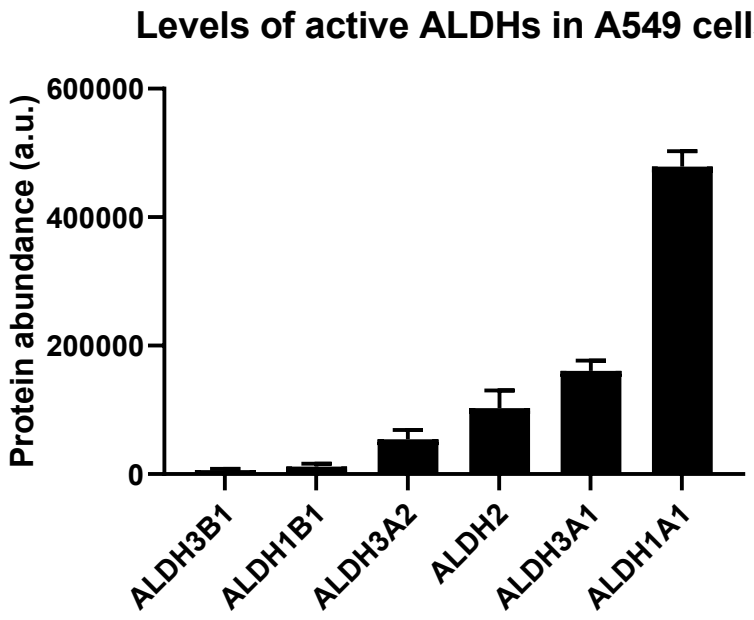
Supplementary Data



Supplementary Fig. 8.1 | *In situ* labeling of ALDH1A3^{WT} and mutant ALDH1A3^{C314A} with STA-55 (1 µM) for 1 h at 37 °C and α-FLAG Western blot.



Supplementary Fig. 8.2 | Heatmap showing relative protein levels of ALDH enzymes in A549 cells using deep proteome data from Bekker-Jensen *et al.* analysed using the Expression Atlas.[20,21]



Supplementary Fig. 8.3 | Data represent mean LFQ levels ± SD of the ALDHs significantly enriched in the A549 lung cancer cell line. *N* = 3 experiments (biological replicates).

References

1. Vasiliou, V., Pappa, A. & Petersen, D. R. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem. Biol. Interact.* **129**, 1–19 (2000).
2. Kutzbach, C. & Stokstad, E. L. R. Mammalian methylenetetrahydrofolate reductase Partial purification, properties, and inhibition by S-adenosylmethionine. *BBA - Enzymol.* **250**, 459–477 (1971).
3. Chern, M. K. & Pietruszko, R. Human aldehyde dehydrogenase E3 isozyme is a betaine aldehyde dehydrogenase. *Biochem. Biophys. Res. Commun.* **213**, 561–568 (1995).
4. Vasiliou, V. & Pappa, A. Polymorphisms of Human Aldehyde Dehydrogenases. *Pharmacology* **61**, 192–198 (2000).
5. Laurenzi, V. De *et al.* Sjögren–Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nat. Genet.* **12**, 52–57 (1996).
6. Geraghty, M. T. *et al.* Mutations in the Δ1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. *Hum. Mol. Genet.* **7**, 1411–1415 (1998).
7. Chambliss, K. L. *et al.* Two Exon-Skipping Mutations as the Molecular Basis of Succinic Semialdehyde Dehydrogenase Deficiency (4-Hydroxybutyric Aciduria). *Am. J. Hum. Genet.* **63**, 399–408 (1998).
8. Moreb, J. S., Muhoczy, D., Ostmark, B. & Zucali, J. R. RNAi-mediated knockdown of aldehyde dehydrogenase class-1A1 and class-3A1 is specific and reveals that each contributes equally to the resistance against 4-hydroperoxycyclophosphamide. *Cancer Chemother. Pharmacol.* **59**, 127–136 (2007).
9. Emadi, A., Jones, R. J. & Brodsky, R. A. Cyclophosphamide and cancer: Golden anniversary. *Nature Reviews Clinical Oncology* **6**, 638–647 (2009).
10. Yasgar, A. *et al.* A High-Content assay enables the automated screening and identification of small molecules with specific ALDH1A1-Inhibitory activity. *PLoS One* **12**, 1–19 (2017).
11. Serwa, R. & Tate, E. W. Activity-based profiling for drug discovery. *Chemistry and Biology* **18**, 407–409 (2011).
12. Koenders, S. T. A. *et al.* Development of a Retinal-Based Probe for the Profiling of Retinaldehyde Dehydrogenases in Cancer Cells. *ACS Cent. Sci.* **5**, 1965–1974 (2019).
13. Morgan, C. A., Parajuli, B., Buchman, C. D., Dria, K. & Hurley, T. D. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. *Chem. Biol. Interact.* **234**, 18–28 (2015).
14. Yang, S. M. *et al.* Discovery of Orally Bioavailable, Quinoline-Based Aldehyde Dehydrogenase 1A1 (ALDH1A1) Inhibitors with Potent Cellular Activity. *J. Med. Chem.* **61**, 4883–4903 (2018).
15. Parajuli, B., Fishel, M. L. & Hurley, T. D. Selective ALDH3A1 inhibition by benzimidazole analogues increase mafosfamide sensitivity in cancer cells. *J. Med. Chem.* **57**, 449–461 (2014).
16. Khanna, M. *et al.* Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. *J. Biol. Chem.* **286**, 43486–43494 (2011).
17. Moretti, A. *et al.* Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD+ and retinoic acid. *Sci. Rep.* **6**, 35710 (2016).
18. Moreb, J. S., Zucali, J. R., Ostmark, B. & Benson, N. A. Heterogeneity of aldehyde dehydrogenase expression in lung cancer cell lines is revealed by Aldefluor flow cytometry-based assay. *Cytom. Part B Clin. Cytom.* **72B**, 281–289 (2007).
19. Kang, J. H. *et al.* Aldehyde dehydrogenase is used by cancer cells for energy metabolism. *Exp. Mol. Med.* **48**, 1–13 (2016).
20. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **4**, 587–599 (2017).
21. Petryszak, R. *et al.* Expression Atlas update - An integrated database of gene and protein expression in humans,

animals and plants. *Nucleic Acids Res.* **44**, 746–752 (2016).

22. Ogata, H., Goto, S., Fujibuchi, W. & Kanehisa, M. Computation with the KEGG pathway database. *Biosystems* **47**, 119–128 (1998).
23. Thomas, P. D. *et al.* PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research* **31**, 334–341 (2003).
24. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, 506–515 (2019).
25. Van Esbroeck, A. C. M. *et al.* Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor BIA 10-2474. *Science* **356**, 1084–1087 (2017).
26. Van Rooden, E. J. *et al.* Mapping in vivo target interaction profiles of covalent inhibitors using chemical proteomics with label-free quantification. *Nat. Protoc.* **13**, 752–767 (2018).
27. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
28. Distler, U., Kuharev, J., Navarro, P. & Tenzer, S. Label-free quantification in ion mobility–enhanced data-independent acquisition proteomics. *Nat. Protoc.* **11**, 795–812 (2016).
29. Distler, U. *et al.* Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **11**, 167–170 (2014).
30. Kuharev, J., Navarro, P., Distler, U., Jahn, O. & Tenzer, S. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* **15**, 3140–3151 (2015).