Affinity Purification Strategies for Proteomic Analysis of Transcription Factor Complexes

Roberto Giambruno,[#] Florian Grebien,[#] Alexey Stukalov, Christian Knoll, Melanie Planyavsky, Elena L. Rudashevskaya,[⊥] Jacques Colinge, Giulio Superti-Furga,* and Keiryn L. Bennett*

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, AKH BT 25.3, 1090 Vienna, Austria

S *Supporting Information*

ABSTRACT: Affinity purification (AP) coupled to mass spectrometry (MS) has been successful in elucidating protein molecular networks of mammalian cells. These approaches have dramatically increased the knowledge of the interconnectivity present among proteins and highlighted biological functions within different protein complexes. Despite significant technical improvements reached in the past years, it is still challenging to identify the interaction networks and the subsequent associated functions of nuclear proteins such as transcription factors (TFs). A straightforward and robust methodology is therefore required to obtain unbiased and reproducible interaction data. Here we present a new approach for TF AP-MS, exemplified with the CCAAT/enhancer binding protein alpha (C/EBPalpha). Utilizing the advantages of a double tag and three different MS strategies, we conducted a total of six independent AP-MS strategies to analyze the protein−protein interactions of C/EBPalpha. The resultant data were combined to produce a cohesive C/EBPalpha interactome. Our study describes a new methodology that robustly identifies specific molecular complexes associated with transcription factors. Moreover, it emphasizes the existence of TFs as protein complexes essential for cellular biological functions and not as single, static entities.

KEYWORDS: *affinity purifications, transcription factors, mass spectrometry*
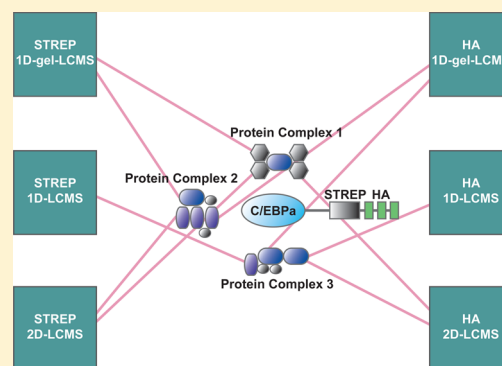
INTRODUCTION

Affinity purification of proteins coupled to mass spectrometry (AP-MS) has markedly improved our knowledge of protein interactions and functions, and has become the hallmark for diverse biological discoveries.[1,2] In addition, AP-MS has allowed the characterization of several molecular networks in a variety of organisms, including mammals.[3−5] Although this approach has proven very robust and highly reproducible for certain classes of proteins,[6] the same cannot be assumed for nuclear proteins. This is particularly the case for transcription factors (TFs) that strongly associate with chromatin. Limiting factors for efficient AP-MS studies are the general low abundance of these proteins and the technical difficulties associated with the purification of TFs from mammalian cells.[7] As the elucidation of the function of TFs is highly relevant to human diseases, a more detailed understanding of the biological role of TFs is required.

TFs have often been found to operate in large, dynamic multiprotein complexes involved in several different cellular processes.[8] TFs have also been shown to associate with different interactors that can modulate transcriptional output.[9] In addition, mutations affecting the protein interaction profile of specific TFs result in aberrant phenotypes as previously reported for the protein C/EBPalpha.[10]

In the hematopoietic system, the TF C/EBPalpha is a master regulator of the myeloid compartment, and the role of this protein is specifically required for functional neutrophilic differentiation.[11] Deletion of the *C/EBPalpha* gene results in the complete absence of mature granulocytes.[12] In addition, mutations in the C/EBPalpha protein have been identified in patients with acute myeloid leukemia (AML).[11] Thus, changes in C/EBPalpha function are closely linked to leukemogenesis.[13] It has been proposed that the functions of C/EBPalpha are dependent on the formation of different protein complexes to promote cell cycle arrest and terminal differentiation.[14−16] A global and comprehensive characterization of the TF C/EBPalpha-containing protein complexes is therefore critical in understanding the role of this protein. C/EBPalpha was chosen as an example to demonstrate our new AP-MS methodology in the identification of protein complexes associated with a selected transcription factor.

EXPERIMENTAL PROCEDURES

Cells

The cDNA for rat C/EBPalpha was stably expressed in the myeloid progenitor cell line FDCP-1 from a modified bicistronic pMSCV-IRES-GFP retroviral vector (Clontech, Mountain View, CA, USA) with a carboxyl-terminal streptavidin (STREP)−hemagglutinin (HA) tag. Transduced cells were isolated by FACS-sorting for the GFP-positive population. Cells were cultivated in RPMI (PAA Laboratories, Cölbe, Germany) supplemented with 10% FCS (Invitrogen, Carlsbad, CA, USA), penicillin (100 U/mL) (PAA Laboratories), and streptomycin (100 μg/mL) (PAA Laboratories) and 5 ng/mL murine IL-3 (PeproTech, Rocky Hill, NJ, USA).

Affinity Purification − Preparation

All steps described in the protocol were at 4 °C. Purifications were performed from 1 × 10^9 freshly harvested cells. After being washed with PBS, cells were incubated in buffer N (300 mM sucrose, 10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.1 mM DTT, 0.75 mM spermidine, 0.15 mM spermine, 0.1% Nonidet P-40, 50 mM NaF, 1 mM Na$_3$VO$_4$, protease inhibitors) for 5 min on ice to lyse the cytoplasm. Nuclei were collected by centrifugation (500g for 5 min), and the supernatant (containing the cytoplasmic fraction) was removed. The nuclear pellet was washed (1×) with buffer N. For the extraction of nuclear proteins, the nuclei were resuspended in buffer C420 (20 mM HEPES pH 7.9, 420 mM NaCl, 25% glycerol, 1 mM EDTA, 1 mM EGTA, 0.1 mM DTT, 50 mM NaF, 1 mM Na$_3$VO$_4$, protease inhibitors), vortexed briefly, and shaken vigorously for 30 min. After centrifugation for 1 h at 100000g, the supernatant contained the soluble nuclear proteins, while nuclear membranes and insoluble chromatin remained in the pellet.

The protein concentration of the soluble nuclear fraction was measured by Bradford assay (using γ-globin as the standard). Fifteen milligrams of nuclear extract was subjected to either one-step or two-step affinity purifications. Prior to purification, all nuclear extracts were adjusted to 2 mg/mL and 150 mM NaCl with HEPES buffer (20 mM HEPES, 50 mM NaF, 1 mM Na$_3$VO$_4$, protease inhibitors). Samples were incubated for 20 min at 4 °C on a rotating wheel with RNase A, benzonase, and avidin to remove RNA and nucleic acids (DNA) and to saturate endogenously biotinylated proteins, respectively. An aliquot of the sample was retained for immunoblot (IB) analysis (Supporting Information, Supplementary Figure 1b, lysate).

Two-Step Affinity Purification

Two-step STREP-HA purifications were performed essentially as previously described.[17] Precleared extracts were incubated with 200 μL of StrepTactin sepharose beads (IBA, Göttingen, Germany) and incubated for 2 h at 4 °C on a rotating wheel. The sample was transferred to a Biospin column (BioRad, Hercules, CA, USA), and the flow-through was removed by gravity flow (an aliquot of the flow-through was retained for IB analysis, Supplementary Figure 1b, STREP-FT). Beads were washed 3× with TNN-HS buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 50 mM NaF, 1 mM Na$_3$VO$_4$, protease inhibitors). Proteins bound to the beads were eluted with 3 × 300 μL 2.5 mM biotin in TNN-HS buffer (an aliquot of the biotin eluate was retained for IB analysis, Supplementary Figure 1b, STREP-Eluate). The biotin eluate was subsequently incubated with anti-HA-agarose beads (Sigma, St.Louis, MO, USA) for 2 h at 4 °C on a rotating

wheel. Samples were centrifuged for 3 min at 300g (an aliquot of the flow-through was retained for IB analysis, Supplementary Figure 1b, HA-FT), and the beads were washed 3× with TNN-HS buffer. Another two washing steps with TNN-HS buffer without detergent and inhibitors were performed to remove traces of detergent that are detrimental to LC−MS analysis. For one-dimensional gel-based liquid chromatography mass spectrometry (1D-gel-LC−MS) analysis, bound proteins were eluted in 100 μL of 4× Laemmli buffer, boiled for 5 min and centrifuged for 3 min at 300g. The supernatant was transferred to a new eppendorf tube (100 μL) and 1/10 was retained for IB analysis (Supplementary Figure 1b, HA-Eluate). For one-dimensional gel-free liquid chromatography mass spectrometry (1D-LC−MS), bound material was eluted with 500 μL of 100 mM formic acid and immediately neutralized with 125 μL of 1 M triethylammonium bicarbonate (TEAB). From the final eluate, 200 μL was retained for IB analysis (Supplementary Figure 1c, HA-Eluate).

One-Step Affinity Purification: STREP-tag

For one-step STREP purifications, precleared nuclear extracts were incubated with 200 μL of StreptTactin sepharose beads and at 4 °C incubated for 2 h on a rotating wheel. Samples were centrifuged for 3 min at 300g, the flow-through was transferred to a new eppendorf tube (for IB analysis, Supporting Information, Supplementary Figure 2a, STREP-FT), and the beads were washed 3× with TNN-HS buffer, and 2× using TNN-HS buffer without detergent and inhibitors (all washes in batch). For 1D-gel-LC−MS analysis, bound proteins were incubated with 100 μL of 2.5 mM biotin in TNN-HS buffer 10 min at 4 °C, followed by centrifugation for 3 min at 300g. The supernatant was transferred to a new eppendorf tube and 1/10 was retained for IB analysis (Supplementary Figure 2a, STREP-Eluate). Samples were then alkylated with iodoacetamide and separated by 1D SDS-PAGE on a 4−12% bis-Tris gel (NuPAGE, Invitrogen). Proteins were visualized by silver staining, and the entire gel lane was excised and the 20 slices digested in situ with modified porcine trypsin (Promega Corp., Madison, WI, USA) as previously described.[18] The obtained peptides were pooled into 10 samples and analyzed by online LC−MS/MS (1D-gel-MS).

The elution for 1D-LC−MS was performed using 312.5 μL of biotin in TNN-HS buffer for 10 min at 4 °C, followed by centrifugation for 3 min at 300g. From the eluted material, 100 μL was retained for IB analysis (Supplementary Figure 2b, STREP-Eluate). The remainder of the samples was digested with trypsin, and 2 × 5% of the resultant peptides were analyzed as technical duplicates by LC−MS/MS (1D-LC−MS). The material not used for 1D-LC−MS was separated by off-line reversed-phase liquid chromatography, and the resultant 10 fractions were analyzed by two-dimensional gel-free LC−MS (2D-LC−MS).

One-Step Affinity Purification: HA-tag

For the one-step HA purification, precleared nuclear extracts were incubated with 200 μL of anti-HA sepharose beads for 2 h at 4 °C on a rotating wheel. Samples were centrifuged for 3 min at 300g, the flow-through was transferred to a new eppendorf tube (for IB analysis, Sup. 2a, HA-FT), and beads were washed 3× with TNN-HS and 2× with TNN-HS buffer without detergent and inhibitors. For the 1D-gel-LC−MS approach, bound proteins were eluted in 100 μL of 4× Laemmli buffer, boiled for 5 min, and centrifuged for 3 min at 300g. The supernatant was then transferred to a new eppendorf tube (100

μL) and 1/10 was retained for IB analysis (Sup. 2a, HA-Eluate). The eluted proteins were separated by 1D-SDS-PAGE, and entire lanes were excised in 20 slices, digested in situ with trypsin, and treated as reported before (1D-gel-MS). For the 1D-LC−MS, proteins were eluted with 500 μL of 100 mM formic acid and immediately neutralized with 125 μL of 1 M TEAB. As described above, 200 μL of the final eluate was retained for IB analysis (Sup. 2b, HA-Eluate).

Reversed-Phase Reversed-Phase (RPRP) Separation[19,20]

Tryptic digests were concentrated and purified by solid phase extraction (SPE) (UltraMicroSpin columns 3−30 μg capacity, Nest Group Inc., Southboro, MA, USA) prior to injection onto a Phenomenex column (150 × 2.0 mm Gemini-NX 3 μmC18 110 Å, Phenomenex, Torrance, CA, USA) on an Agilent 1200 series HPLC (Agilent Biotechnologies, Palo Alto, CA) with UV detection at 214 nm. HPLC solvent A consisted of 20 mM NH_4OH pH 10.5 in 5% acetonitrile and solvent B consisted of 20 mM NH_4OH pH 10.5 in 90% acetonitrile. Ten time-based fractions were collected and acidified. The sample volume was reduced to approximately 2 μL in a vacuum centrifuge and reconstituted to 10 μL with 5% formic acid.

Liquid Chromatography−Mass Spectrometry

All affinity purifications were analyzed on a hybrid linear trap quadrupole (LTQ) Orbitrap Velos mass spectrometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) coupled to a 1200 series high-performance liquid chromatography system (Agilent Technologies, Palo Alto, CA, USA) via a nanoelectrospray ion source using liquid junction (Proxeon, Odense, Denmark). Solvents for HPLC separation of peptides were as follows: solvent A consisted of 0.4% formic acid in water, and solvent B consisted of 0.4% formic acid in 70% methanol and 20% isopropanol. From a thermostatted microautosampler, 8 μL of the tryptic peptide mixture was automatically loaded onto a trap column (Zorbax 300SB-C18 5 μm, 5 × 0.3 mm, Agilent Biotechnologies, Palo Alto, CA, USA) with a binary pump at a flow rate of 45 μL/min. 0.1% trifluoroacetic acid was used for loading and washing the precolumn. After washing, the peptides were eluted by back-flushing onto a 16-cm fused silica analytical column with an inner diameter of 50 μm packed with C18 reversed phase material (ReproSil-Pur 120 C18-AQ, 3 μm, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The peptides were eluted from the analytical column with a 27 min gradient ranging from 3 to 30% solvent B, followed by a 25 min gradient from 30 to 70% solvent B and, finally, a 7 min gradient from 70 to 100% solvent B at a constant flow rate of 100 nL/min.[21] The analyses were performed in a data-dependent acquisition mode and dynamic exclusion for selected ions was 60 s. A top 15 collision-induced dissociation (CID) method were used, and a single lock mass at m/z 445.120024 $(Si(CH_3)_2O)_6)$[21] was employed. Maximal ion accumulation time allowed in CID mode was 50 ms for MS^n in the LTQ and 500 ms in the C-trap. Automatic gain control was used to prevent overfilling of the ion traps and was set to 5000 in MS^n mode for the LTQ and 10^6 ions for a full FTMS scan. Intact peptides were detected in the Orbitrap Velos at 60 000 resolution at m/z 400.

Protein Identification

Peak list information was extracted from the RAW MS files and converted into an MGF format with the msconvert tool (ProteoWizard Library v2.1.2708). The MGF files were searched against the mouse component of the UniProtKB/

SwissProt database (www.uniprot.org; releases 2010-09, 2011-12, and 2012-05), including all protein isoforms plus the rat C/EBPalpha and known contaminant sequences (e.g., human keratin). An initial search was performed with the protein search engine Mascot (www.matrixscience.com, version 2.3.02). Mass error tolerances on the precursor and fragment ions were ±10 ppm and ±0.6 Da, respectively. Only fully tryptic peptides were considered with a maximum of one missed cleavage site, and carbamidomethyl cysteine and methionine oxidation set as fixed and variable modifications, respectively. The Mascot peptide ion score threshold was equal to 30, and at least three peptide identifications per protein were required.

For both the precursor and fragment ion data, linear recalibration transformations that minimize the mean square deviation of the measured from theoretical values were deduced from the initial identifications. Recalibrated MGF files were searched against the same protein database with Mascot and Phenyx (GeneBio, SA, version 2.5.14)[22] using narrower mass tolerances (±4 ppm and ±0.3 Da). All other search parameters were identical to the initial first pass search. Mascot and Phenyx output files were processed by internally developed parsers to filter and integrate protein identifications. For Mascot and Phenyx identifications, the following peptide score thresholds were used: $T_1 = 14$, $T_2 = 40$, and $T_3 = 10$; and $T_1 = 4.2$, $T_2 = 4.75$, and $T_3 = 3.5$, respectively (P-value $<10^{-3}$). Proteins with at least two unique peptides above score T_1 or with a single peptide above T_2 were selected as unambiguous identifications. Additional peptides from these validated proteins with a score $>T_3$ were appended to the final result. The validated identifications from both algorithms were merged, spectral conflicts were discarded, and protein groups were defined according to shared peptides. A false discovery rate (FDR) of <1% for protein identifications and <0.1% for peptides (including peptides exported with lower scores) was estimated by applying the same filtering procedure against a database of reversed protein sequences.

Bioinformatic Analyses

All affinity purifications were analyzed based on protein spectral counts. For each purification strategy, proteins identified in the control cells were subtracted from the proteins identified from the corresponding C/EBPalpha affinity purifications. Moreover, contaminants such as keratin, spectrin, plectin, and anti-HA antibody chains were removed from the list as known nonspecific binders frequently observed in AP-MS.

■ RESULTS

To determine the protein interactome of C/EBPalpha, an affinity-tagged variant of the TF was expressed in the myeloid progenitor cell line FDCP-1. A well-established, double-affinity purification tag (STREP−HA) was used throughout.[17,23] The tag consists of two STREP motifs and three HA epitopes. A double-affinity tag enables the enrichment of a specific protein and associated interactors in two sequential purification steps. In parallel, nonspecifically binding proteins are removed from the sample.[23] The first purification step takes advantage of the strong affinity that the STREP-peptide has for a variant of streptavidin (StrepTactin). Bound protein complexes are specifically eluted with biotin. The second purification consists of an immunopurification step using an anti-HA-antibody. Bound proteins are eluted with formic acid or sodium dodecyl sulfate (SDS) buffer for analysis by 1D-LC−MS or 1D-gel-LC−MS, respectively.
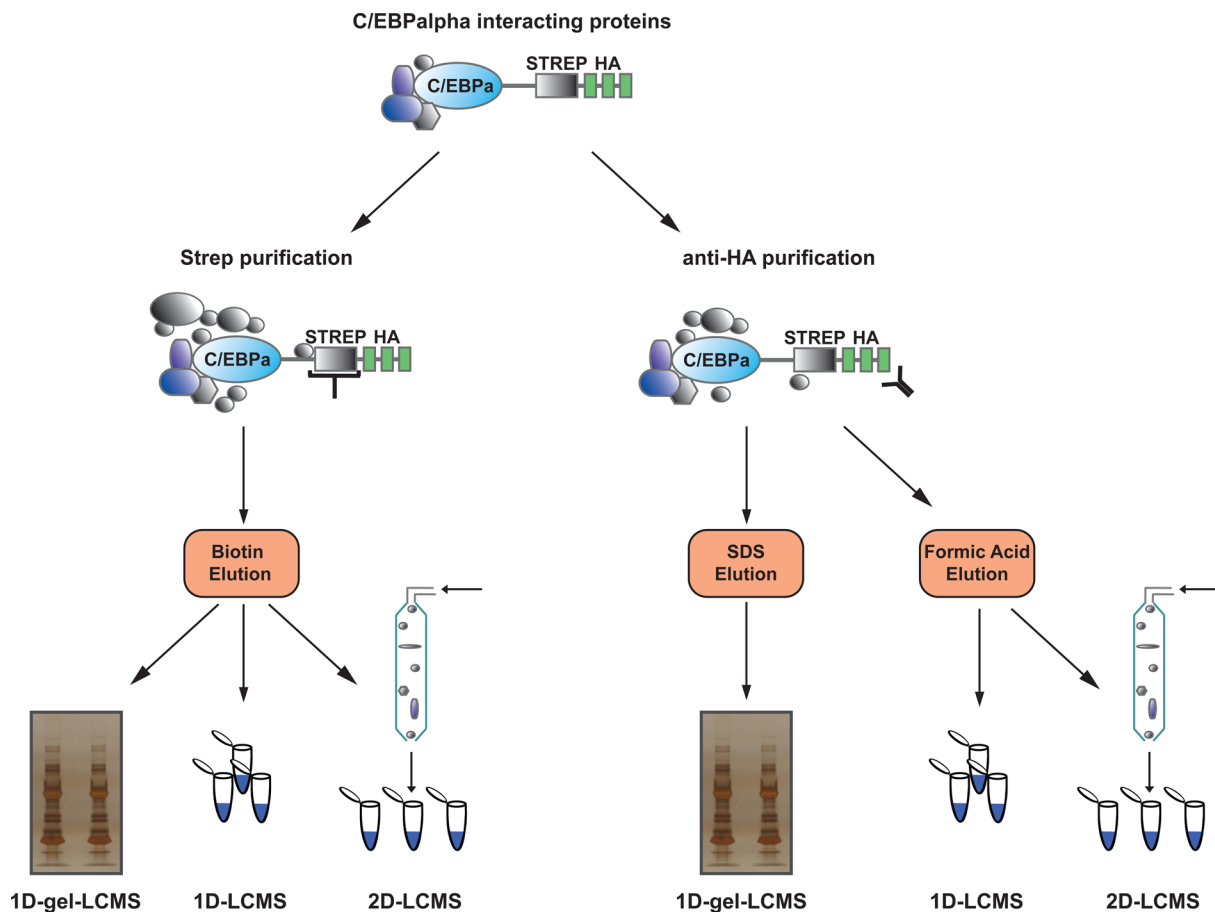
Figure 1. Summary and flowchart of the different purification procedures used in this study. The two single step STREP and HA purifications are displayed on the left and right, respectively. The different elution buffers are displayed in the orange boxes. Finally, each purification step is visualized by a schematic of the alternative LC−MS analyses.

Because of the predominantly nuclear localization of C/EBPalpha, the standard whole cell extract lysis buffers were not recommended. Starting from whole cell extracts, STREP-HA purifications of C/EBPalpha were poorly enriched, and only the selected transcription factor and a few interactors were identified by MS (data not shown). After several protocols for extracting nuclear proteins (data not shown) were assessed, a high-salt extraction approach using 420 mM NaCl was chosen.[24] As nuclear extracts are rich in nucleic acids, abundant DNA- and RNA-binding proteins can be preferentially copurified and identified by LCMS. This occurs at the expense of the identification of the low-abundance proteins that typically constitute transcription factor complexes. Thus, the nuclear extracts were treated with RNase A and benzonase to remove nucleic acids that are not protected by bound transcription factors. In parallel, the samples were also treated with free avidin to remove the majority of the naturally biotinylated proteins (e.g., carboxylases) prior to affinity purification using StrepTactin. As avidin does not display any affinity for STREP, this preclearing step increases the specificity of the purification.

Standardized two-step STREP-HA purifications[17] from 15 mg of pretreated nuclear extracts were conducted on C/EBPalpha-overexpressing FDCP-1 cells and control mock-infected cells that do not express any STREP-HA-tagged protein. Eluates were analyzed by 1D-gel-LC−MS and 1D-LC−MS, and the identified proteins were filtered by subtracting all proteins detected in the purifications from the control cell line.

Despite high quantities of purified C/EBPalpha, only a few protein interactors were observed (Supporting Information, Supplementary Figure 1). Mainly ribosomal proteins, chaperones, and secreted proteins were identified. The exception was Ddit3, a known C/EBPalpha interactor[25] (Supplementary Figure 1a). Although the results showed that the two-step AP-MS approach correctly purified C/EBPalpha, the method appeared to be too harsh to retain and subsequently identify the different C/EBPalpha interactors. This is most likely due to the transient and/or unstable nature of transcription factor complexes.[26] Following treatment of the STREP-eluate with four different cross-linking reagents, the presence of several high molecular weight protein complexes associated with the purified C/EBPalpha was confirmed by anti-HA immunoblot (Supplementary Figure 2a,b). On the basis of the immunoblot of the one-step procedure combined with the two-step data from AP-MS, we surmised that the second purification step of the STREP-HA purification is indeed limiting for the detection of transient C/EBPalpha protein complexes. In consideration of these findings and to augment the sensitivity of the AP-MS analyses, our next efforts were focused on one-step affinity purifications. The caveat with less stringent purification conditions inherent with a one-step procedure is that the frequency of nondirect interactors in the purified material increases.[2] To enhance the dynamic range of the LCMS analyses and boost the number of proteins detected, an additional 2D-LCMS approach was introduced.[20]
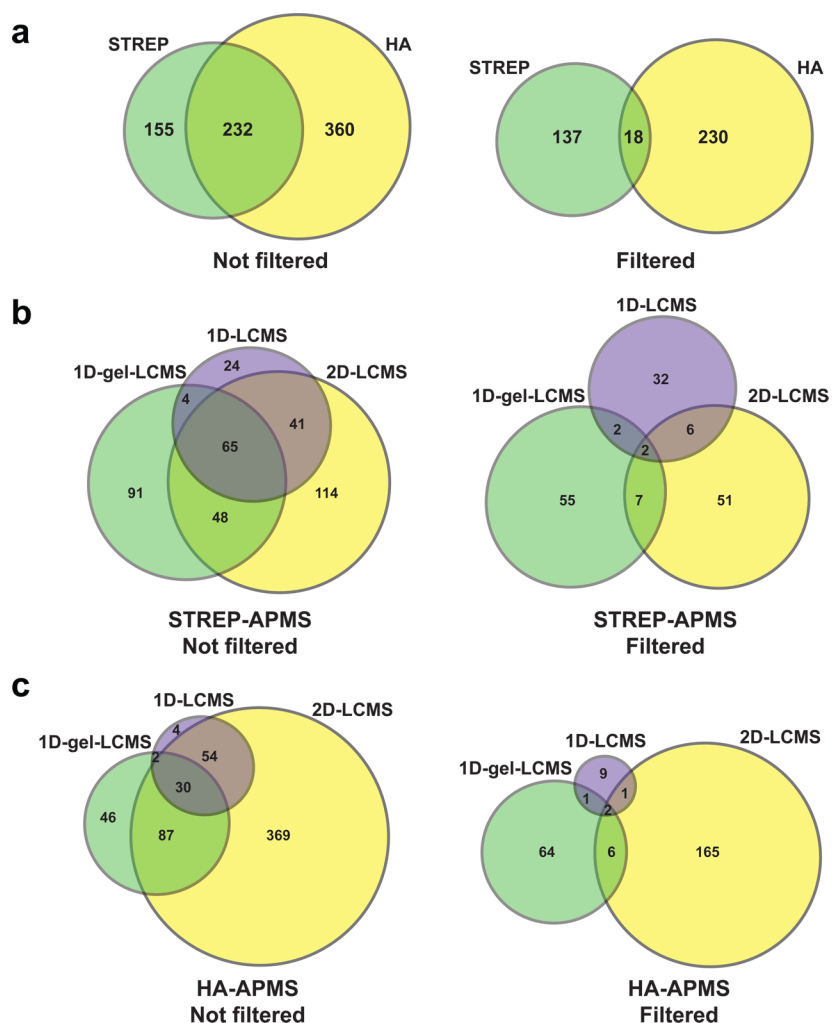
Figure 2. Venn diagrams of proteins identified in the STREP- and HA-purifications (adapted from BioVenn, Copyright T. Hulsen). (a) Comparison of all proteins identified by the six AP-MS strategies, in the nonfiltered and the filtered networks. (b) and (c) Representation of the proteins identified with the different LCMS analyses for both STREP- and HA-purifications. Comparison of the nonfiltered and the filtered networks.

Taking advantage of the STREP and HA motifs in the double tag, one-step purifications were combined with three LCMS approaches. A total of six independent AP-MS strategies were thus designed to characterize the C/EBPalpha interactome (Figure 1). The STREP motif allows specific elution of the purified TF complex with biotin. The eluate is compatible with 1D-, 2D-LC−MS and 1D-gel-LC−MS. Conversely, the HA-bound material was nonspecifically eluted with either SDS for 1D-gel-LC−MS or with formic acid for 1D- and 2D-LC−MS. Although a specific elution with HA-peptides is possible,[27] this approach is not compatible with gel-free LCMS as a large excess of free HA peptide drastically reduces the dynamic range of the MS analysis (data not shown).

Following the affinity purification protocols described above, all one-step purifications were performed as biological replicates, and the eluted material was analyzed on an LTQ Orbitrap Velos mass spectrometer. The quality of the affinity purifications was monitored by immunoblotting for exogenous C/EBPalpha protein (Supporting Information, Supplementary Figure 3) and confirmed by the number of spectral counts specific for the TF. In all C/EBPalpha samples, high numbers of nuclear proteins were identified associated with the tagged TF (data not shown). In addition, residual biotinylated carboxylases and anti-HA antibody chains originating from

the STREP- and the HA-purifications, respectively, were also apparent (see Supplementary Table 1).

For all proteins identified in the STREP- versus the HA-purifications, a 45% overlap was evident. Subtraction of proteins detected in the negative controls, however, dramatically reduced the number of common proteins (Figure 2a). The same trend was observed when the proteins identified by all three different LCMS strategies were compared. After filtering, the only shared protein was exogenous and endogenous C/EBPalpha (Figure 2b,c). Interestingly, both the qualitative and semiquantitative profile of the proteins identified with each of the two tags was substantially different. This confirmed the high variability that has previously been reported in the literature for different tag-based affinity purifications[28] (Figure 2a). Except for the 1D-gel-free LC−MS analyses of the anti-HA-tag purifications, all strategies identified a high number of proteins that specifically associated with C/EBPalpha. Among these, 21 interactions have been previously reported[29] (Figure 3). In addition, both purification strategies were characterized by a different set of background proteins that associate with the tag and/or the affinity resin. As summarized in Table S1, each approach can be preferentially selected based on a particular experimental design.
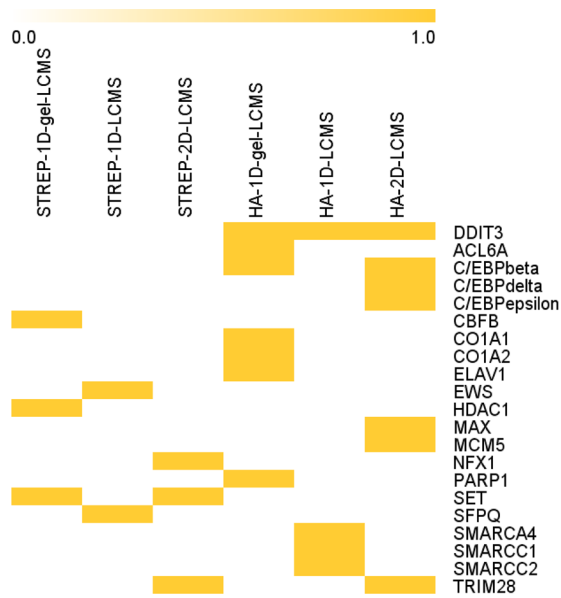

Figure 3. Representation of previously described C/EBPalpha-interactors that were identified by STREP- and HA-purifications.

associated with different tag-based purification, various elution protocols, and alternative MS analyses may all preferably identify certain classes of proteins. Therefore, a common network featuring the proteins identified from all six strategies was created. According to the interaction data from the literature, all proteins identified with our approaches were clustered into previously described multiprotein complexes (Figure 4). This approach markedly increased the connections between the six different strategies. C/EBPalpha was found to interact with different protein complexes, and nine of these were reproducibly retrieved by the different affinity purification methods (Figures 4 and 5). For instance, the MLL-SWI/SNF complex was identified by five out of six strategies, confirming the previously reported interaction of C/EBPalpha with this complex.[10,30,31] On the basis of these data, a total of nine functional nuclear protein complexes were identified from the intersection of more than two independent affinity purification approaches (Figure 5).

To confirm the robustness of the complexes, the interactome of a previously described C/EBPalpha mutant was analyzed in the same cellular system. This mutant harbors a duplication of lysine 313 (C/EBPalpha KK) in the basic region-leucine zipper domain.[32] Mutations affecting this domain frequently occur in AML patients[10,32] and were shown to abolish DNA binding of C/EBPalpha. Hence, a 1D-gel-LC−MS AP-MS experiment was conducted using both the STREP- and HA-tags. Following the same approach described for C/EBPalpha, a common protein−protein interaction network was generated for the KK mutant
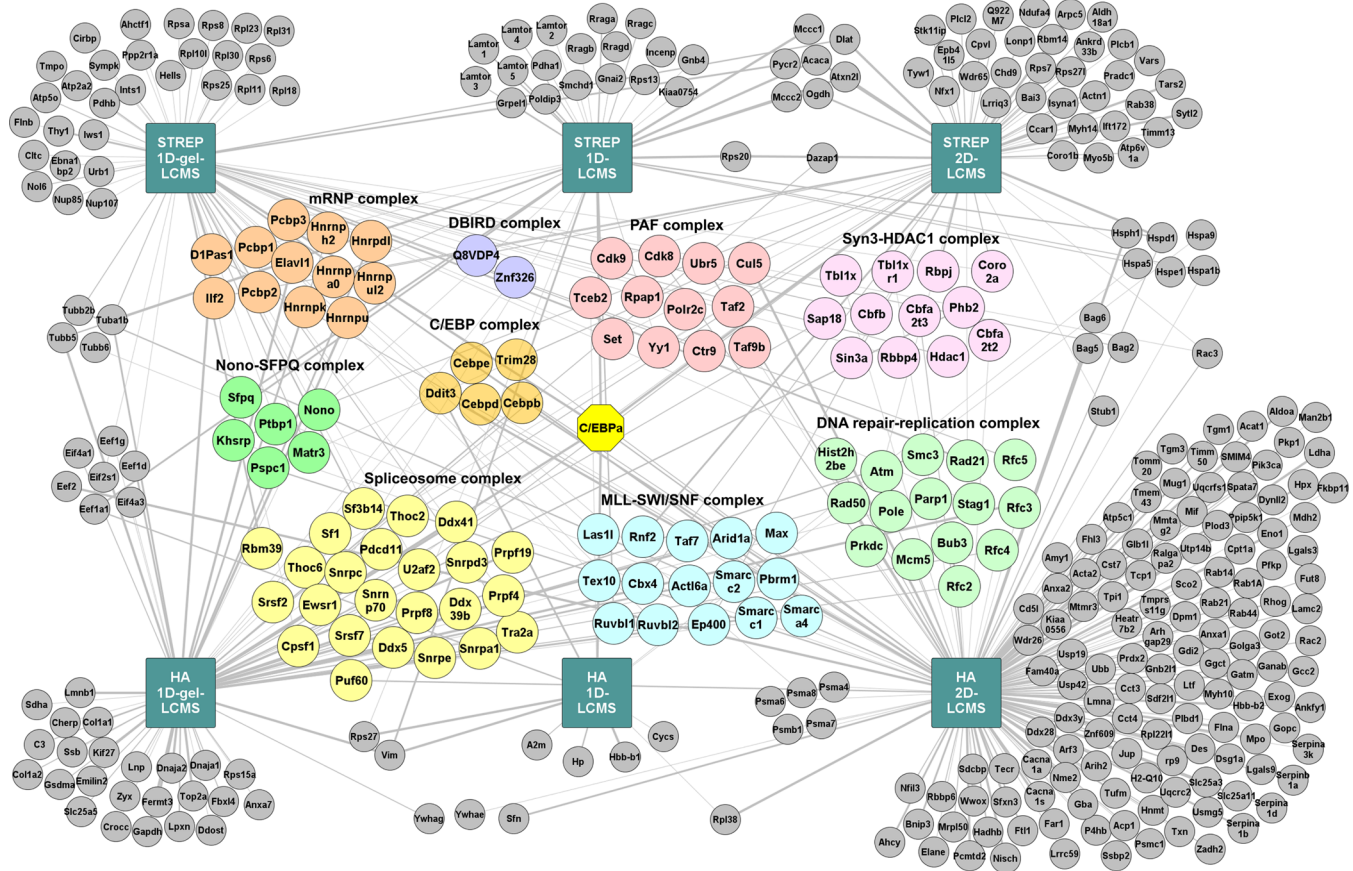
The low number of overlapping C/EBPalpha-interacting proteins from the affinity purification strategies could be due to inherently different biophysical properties of proteins bound to C/EBPalpha. In particular, the biochemical procedures

Figure 4. C/EBPalpha network generated from the intersection of the proteins identified in the six one-step protein purification approaches. From this analysis, is possible to identify nine protein complexes associated with C/EBPalpha. All proteins shown in the figure were identified in our AP-MS analyses.
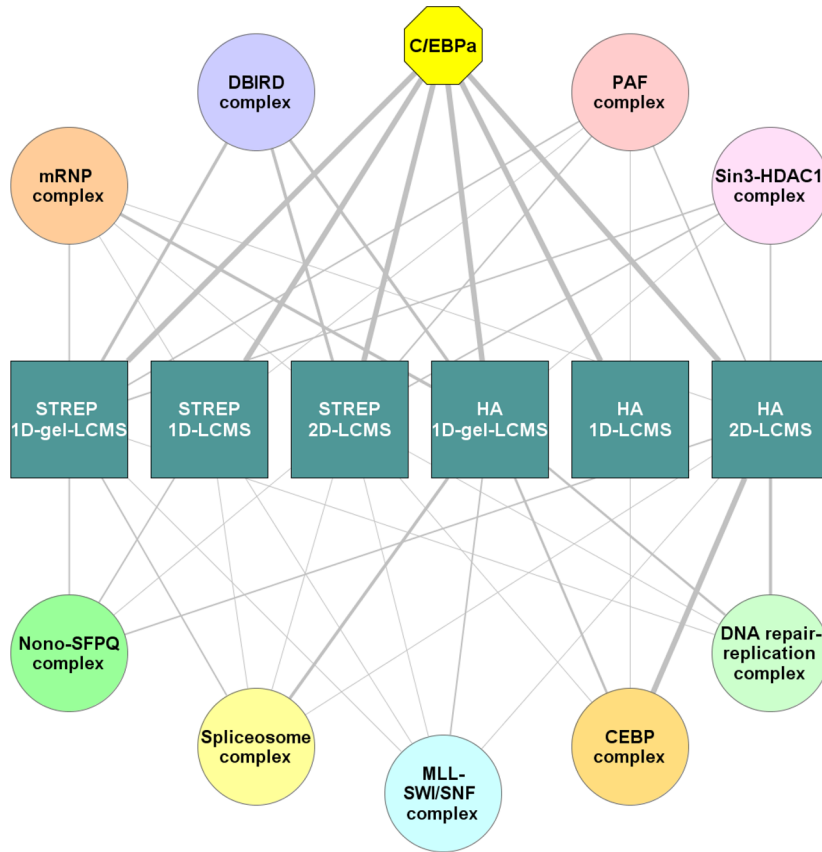
Figure 5. Interaction network of protein complexes associated with C/EBPalpha identified by the six LCMS analyses for both STREP- and HA-purifications. The network edges are proportional to the number of protein identified by each strategy.
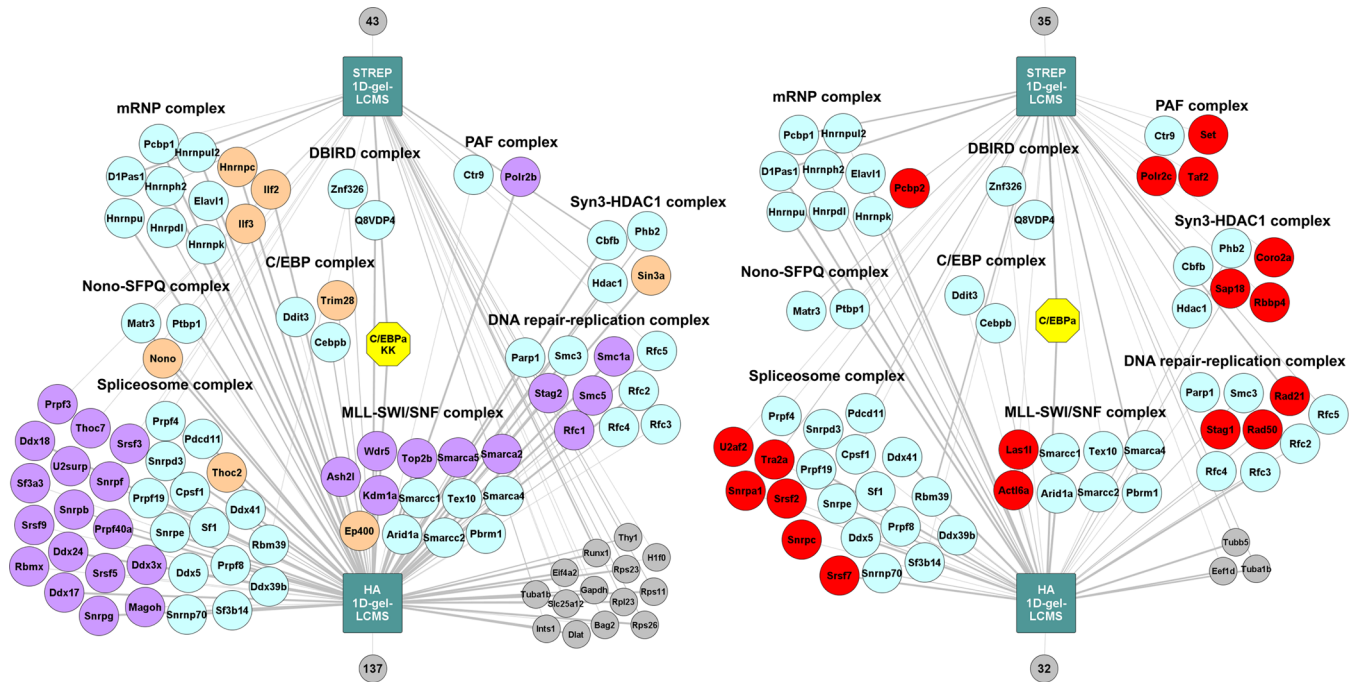

Figure 6. Interaction network of protein complexes associated with C/EBPalpha and C/EBPalpha KK identified by STREP 1D-gel-LC−MS and HA 1D-gel-LC−MS analyses. Light blue, shared interactors of C/EBPalpha and C/EBPalpha KK; light orange, interactors identified in the C/EBPalpha interactome of Figure 4; red, interactors only associated with C/EBPalpha; violet, interactors that solely associate with C/EBPalpha KK; gray, detected proteins that are not part of the C/EBPalpha complexes. For graphical reasons, only the numbers of the proteins exclusively identified by each 1D-gel-LC−MS analysis are shown. All proteins displayed were identified in our AP-MS analysis, and the entire list of the proteins is available in the Supporting Information.

(Figure 6). Despite the different number of proteins identified, both C/EBPalpha and the C/EBPalpha KK mutant were found to associate with the same protein complexes. The composition of some macromolecular complexes, however, was altered. For instance, more proteins described as part of the MLL-SWI/SNF complex were connected to the C/EBPalpha KK mutant. On the other hand, the components of both the Sin3a-Hdac1 and PAF complexes were less represented (Figure 6). In the context of the C/EBPalpha KK mutant, these data suggest an alternative biological function for these protein complexes.

■ DISCUSSION AND CONCLUSION

TFs are known to dynamically associate with several proteins and, depending on the protein partners, TFs can execute different transcriptional programs.[9] In fact, TFs have been reported to act as macromolecular machinery complexes.[33,34] Using C/EBPalpha as model, we developed a novel methodology to identify different molecular complexes associated with TFs in mammalian cells. The use of a double tag for parallel, independent single-step AP-MS procedures from nuclear extracts markedly boost the number of proteins detected as compared to sequential two-step AP-MS approaches. Ultimately, the intersection of the data generated by the different AP-MS strategies strongly emphasized the specific molecular complexes associated with the selected TF to the detriment of so-called nonspecifically binding proteins.

A low overlap in the proteins identified by each independent AP-MS strategy was apparent, even within protocols adopting the same tag affinity purification approach (Figure 2b,c). In contrast, however, when each protein was considered as a component of previously described macromolecular complexes, the overlap among the different AP-MS procedures noticeably increased. Interestingly, each AP-MS strategy adopted favored the identification of alternative members that belong to the same protein complex (Figure 4). Although determining direct interactors of a TF can be quite complex, the use of the methodology proposed here can elucidate the interactome of the TF of interest. In our study, a total of nine independent protein complexes were identified associating with C/EBPalpha (Figure 5). The same macromolecular complexes were found to associate with a previously reported C/EBPalpha DNA binding mutant (C/EBPalpha K313KK). Thus, the specificity of these interactions is highlighted. More interestingly, however, was that subtle differences were observed within the composition of some of the nine protein complexes. This was particularly evident for the complexes that strongly associate with DNA, as observed for the PAF and the Sin3a−Hdac1 complexes. Thus, the data are indicative of alternative biological functions for C/EBPalpha and the C/EBPalpha KK mutant (Figure 6).

Hence, the combined analysis of data sets generated through different experimental strategies enabled the robust and reproducible identification of explicit molecular complexes associated with the C/EBPalpha protein. Our studies suggest that the adoption of different AP-MS strategies for a TF is more efficient in detecting specific protein complexes associated to the protein of interest than increasing the number of biological replicates for each single AP-MS approach. Indeed, different elution buffers and diverse MS strategies alter the reproducibility of the proteins complexes identified by the different independent AP-MS analysis.[28] Conversely, the combined analysis of data generated by at least two different AP-MS approaches tremendously enhances the robustness of TF molecular networks. Taking into account affordability and

depth of analysis required, the six strategies described here provide the framework for optimization of the purification procedure (see Supporting Information, Supplementary Table 1). The combination of experimental strategies for the AP-MS-enabled characterization of interactors of the transcription factor C/EBPalpha represents a dramatic improvement compared to standard AP-MS protocols.

In summary, our systematic analysis showed that one-step purifications from nuclear extracts of cells are necessary to analyze the protein interaction network of a transcription factor. Depending on the purification strategy chosen, the MS analysis must be adjusted accordingly. For STREP-tag purifications, analysis by 1D- and 2D-gel-free LC−MS is highly recommended. For anti-HA antibody-based purifications, a fractionation step at the protein (1D-gel-LC−MS) or peptide level (2D-LC−MS) is essential. Finally, to increase the robustness of the analysis of transcription factor complexes, we recommend the use of a combination of at least two independent tag purifications.

Overall, this study describes a new methodology to characterize the C/EBPalpha protein interaction network. We believe this approach can be robustly and generically extended to other transcription factors and other common nuclear proteins.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

A complete and descriptive list of the peptides identified by each purification is available. In the same section, the network of the STREP-HA purifications and the IB analysis of the different purifications are available. Finally, a description of advantages and disadvantaged of each purification strategy is provided. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*(K.L.B.) E-mail: kbennett@cemm.oeaw.ac.at. Tel: +43-1-40160-70010. Fax: +43-1-40160-970000.
*(G.S.-F.) E-mail: gsuperti@cemm.oeaw.ac.at. Tel: +43-1-40160-70001. Fax: +43-1-40160-970000.

Present Address

⊥Institute of Medical Chemistry, Medical University of Vienna, Waehringer Strasse 10, A-1090 Vienna, Austria CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences Lazarettgasse 14, AKH BT 25.3 1090 Vienna, Austria.

Author Contributions

#These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS USED

AP, affinity purification; MS, mass spectrometry; TF, transcription factor; C/EBPalpha, CCAAT/enhancer binding protein alpha; AML, acute myeloid leukemia; IB, immunoblot; FT, flow-through; STREP, streptavidin-binding peptide motifs; HA, hemeagglutinin; SDS, sodium dodecyl sulfate; 1D-LC–MS, one-dimensional gel-free liquid chromatography mass spectrometry; 1D-gel-LC–MS, one-dimensional gel-based liquid chromatography mass spectrometry; 2D-LCMS, two-dimensional gel-free LC–MS; MLL, myeloid/lymphoid or mixed-lineage leukemia; SWI/SNF, SWItch/Sucrose Non-Fermentable; CID, collision-induced dissociation; LTQ, linear trap quadrupole; FDCP-1, factor dependent cell progenitor-1

■ REFERENCES

(1) Pichlmair, A.; Kandasamy, K.; Alvisi, G.; Mulhern, O.; Sacco, R.; Habjan, M.; Binder, M.; Stefanovic, A.; Eberle, C. A.; Goncalves, A.; Burckstummer, T.; Muller, A. C.; Fauster, A.; Holze, C.; Lindsten, K.; Goodbourn, S.; Kochs, G.; Weber, F.; Bartenschlager, R.; Bowie, A. G.; Bennett, K. L.; Colinge, J.; Superti-Furga, G. Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* **2012**, *487* (7408), 486–490.

(2) Kocher, T.; Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **2007**, *4* (10), 807–815.

(3) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006**, *440* (7084), 631–616.

(4) Bouwmeester, T.; Bauch, A.; Ruffner, H.; Angrand, P. O.; Bergamini, G.; Croughton, K.; Cruciat, C.; Eberhard, D.; Gagneur, J.; Ghidelli, S.; Hopf, C.; Huhse, B.; Mangano, R.; Michon, A. M.; Schirle, M.; Schlegl, J.; Schwab, M.; Stein, M. A.; Bauer, A.; Casari, G.; Drewes, G.; Gavin, A. C.; Jackson, D. B.; Joberty, G.; Neubauer, G.; Rick, J.; Kuster, B.; Superti-Furga, G. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat. Cell Biol.* **2004**, *6* (2), 97–105.

(5) Guruharsha, K. G.; Rual, J. F.; Zhai, B.; Mintseris, J.; Vaidya, P.; Vaidya, N.; Beekman, C.; Wong, C.; Rhee, D. Y.; Cenaj, O.; McKillip, E.; Shah, S.; Stapleton, M.; Wan, K. H.; Yu, C.; Parsa, B.; Carlson, J. W.; Chen, X.; Kapadia, B.; VijayRaghavan, K.; Gygi, S. P.; Celniker, S. E.; Obar, R. A.; Artavanis-Tsakonas, S. A protein complex network of Drosophila melanogaster. *Cell* **2011**, *147* (3), 690–703.

(6) Varjosalo, M.; Sacco, R.; Stukalov, A.; van Drogen, A.; Planyavsky, M.; Hauri, S.; Aebersold, R.; Bennett, K. L.; Colinge, J.; Gstaiger, M.; Superti-Furga, G. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat. Methods* **2013**, *10*, 307–314.

(7) Jiang, D.; Jarrett, H. W.; Haskins, W. E. Methods for proteomic analysis of transcription factors. *J. Chromatogr., A* **2009**, *1216* (41), 6881–6889.

(8) Grosveld, F.; Rodriguez, P.; Meier, N.; Krpic, S.; Pourfarzad, F.; Papadopoulos, P.; Kolodziej, K.; Patrinos, G. P.; Hostert, A.; Strouboulis, J. Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry. *Ann. NY Acad. Sci.* **2005**, *1054*, 55–67.

(9) Rodriguez, P.; Bonte, E.; Krijgsveld, J.; Kolodziej, K. E.; Guyot, B.; Heck, A. J.; Vyas, P.; de Boer, E.; Grosveld, F.; Strouboulis, J. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* **2005**, *24* (13), 2354–2366.

(10) Nerlov, C. C/EBPalpha mutations in acute myeloid leukaemias. *Nat. Rev. Cancer* **2004**, *4* (5), 394–400.

(11) Pabst, T.; Mueller, B. U.; Zhang, P.; Radomska, H. S.; Narravula, S.; Schnittger, S.; Behre, G.; Hiddemann, W.; Tenen, D. G. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat. Genet.* **2001**, *27* (3), 263–270.

(12) Zhang, D. E.; Zhang, P.; Wang, N. D.; Hetherington, C. J.; Darlington, G. J.; Tenen, D. G. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (2), 569–574.

(13) Paz-Priel, I.; Friedman, A. C/EBPalpha dysregulation in AML and ALL. *Crit. Rev. Oncog.* **2011**, *16* (1–2), 93–102.

(14) Iida, S.; Watanabe-Fukunaga, R.; Nagata, S.; Fukunaga, R. Essential role of C/EBPalpha in G-CSF-induced transcriptional activation and chromatin modification of myeloid-specific genes. *Genes Cells* **2008**, *13* (4), 313–327.

(15) Bararia, D.; Trivedi, A. K.; Zada, A. A.; Greif, P. A.; Mulaw, M. A.; Christopeit, M.; Hiddemann, W.; Bohlander, S. K.; Behre, G. Proteomic identification of the MYST domain histone acetyltransferase TIP60 (HTATIP) as a co-activator of the myeloid transcription factor C/EBPalpha. *Leukemia* **2008**, *22* (4), 800–807.

(16) Zada, A. A.; Pulikkan, J. A.; Bararia, D.; Geletu, M.; Trivedi, A. K.; Balkhi, M. Y.; Hiddemann, W. D.; Tenen, D. G.; Behre, H. M.; Behre, G. Proteomic discovery of Max as a novel interacting partner of C/EBPalpha: a Myc/Max/Mad link. *Leukemia* **2006**, *20* (12), 2137–2146.

(17) Rudashevskaya, E. L.; Sacco, R.; Kratochwill, K.; Huber, M. L.; Gstaiger, M.; Superti-Furga, G.; Bennett, K. L. A method to resolve the composition of heterogeneous affinity-purified protein complexes assembled around a common protein by chemical cross-linking, gel electrophoresis and mass spectrometry. *Nat. Protoc.* **2013**, *8* (1), 75–97.

(18) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **1996**, *68* (5), 850–858.

(19) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* **2005**, *28* (14), 1694–1703.

(20) Bennett, K. L.; Funk, M.; Tschernutter, M.; Breitwieser, F. P.; Planyavsky, M.; Ubaida Mohien, C.; Muller, A.; Trajanoski, Z.; Colinge, J.; Superti-Furga, G.; Schmidt-Erfurth, U. Proteomic analysis of human cataract aqueous humour: Comparison of one-dimensional gel LCMS with two-dimensional LCMS of unlabelled and iTRAQ(R)-labelled specimens. *J. Proteomics* **2011**, *74* (2), 151–166.

(21) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics: MCP* **2005**, *4* (12), 2010–2021.

(22) Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T.; Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **2003**, *3* (8), 1454–1463.

(23) Glatter, T.; Wepf, A.; Aebersold, R.; Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol. Syst. Biol.* **2009**, *5*, 237.

(24) Abmayr, S. M.; Yao, T.; Parmely, T.; Workman, J. L. Preparation of nuclear and cytoplasmic extracts from mammalian cells. *Curr. Protoc. Pharmacol.* **2006**, Chapter 12, Unit 12.3.

(25) Ubeda, M.; Wang, X. Z.; Zinszner, H.; Wu, I.; Habener, J. F.; Ron, D. Stress-induced binding of the transcriptional factor CHOP to a novel DNA control element. *Mol. Cell. Biol.* **1996**, *16* (4), 1479–1489.

(26) Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **2006**, *16* (8), 962–972.

(27) Nakatani, Y.; Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **2003**, *370*, 430–444.

(28) Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **2005**, *41* (1), 98−105.

(29) Koleva, R. I.; Ficarro, S. B.; Radomska, H. S.; Carrasco-Alfonso, M. J.; Alberta, J. A.; Webber, J. T.; Luckey, C. J.; Marcucci, G.; Tenen, D. G.; Marto, J. A. C/EBPalpha and DEK coordinately regulate myeloid differentiation. *Blood* **2012**, *119* (21), 4878−4888.

(30) Pedersen, T. A.; Kowenz-Leutz, E.; Leutz, A.; Nerlov, C. Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev.* **2001**, *15* (23), 3208−3216.

(31) Muller, C.; Calkhoven, C. F.; Sha, X.; Leutz, A. The CCAAT enhancer-binding protein alpha (C/EBPalpha) requires a SWI/SNF complex for proliferation arrest. *J. Biol. Chem.* **2004**, *279* (8), 7353−7358.

(32) Bereshchenko, O.; Mancini, E.; Moore, S.; Bilbao, D.; Mansson, R.; Luc, S.; Grover, A.; Jacobsen, S. E.; Bryder, D.; Nerlov, C. Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. *Cancer Cell* **2009**, *16* (5), 390−400.

(33) Collins, A.; Littman, D. R.; Taniuchi, I. RUNX proteins in transcription factor networks that regulate T-cell lineage choice. *Nat. Rev. Immunol.* **2009**, *9* (2), 106−115.

(34) Hurlin, P. J.; Ayer, D. E.; Grandori, C.; Eisenman, R. N. The Max transcription factor network: involvement of Mad in differentiation and an approach to identification of target genes. *Cold Spring Harbor Symp. Quant. Biol.* **1994**, *59*, 109−116.