

Elution Profile Analysis of SDS-induced Subcomplexes by Quantitative Mass Spectrometry*

Yves Texier‡, Grischä Toedt§, Matteo Gorza¶, Dorus A. Mans||**, Jeroen van Reeuwijk||**, Nicola Horn‡, Jason Willer§§, Nicholas Katsanis§§, Ronald Roepman||**, Toby J. Gibson§, Marius Ueffing‡¶|||, and Karsten Boldt‡¶|||

Analyzing the molecular architecture of native multiprotein complexes via biochemical methods has so far been difficult and error prone. Protein complex isolation by affinity purification can define the protein repertoire of a given complex, yet, it remains difficult to gain knowledge of its substructure or modular composition. Here, we introduce SDS concentration gradient induced decomposition of protein complexes coupled to quantitative mass spectrometry and *in silico* elution profile distance analysis. By applying this new method to a cellular transport module, the IFT/lebercilin complex, we demonstrate its ability to determine modular composition as well as sensitively detect known and novel complex components. We show that the IFT/lebercilin complex can be separated into at least five submodules, the IFT complex A, the IFT complex B, the 14-3-3 protein complex and the CTLH complex, as well as the dynein light chain complex. Furthermore, we identify the protein TULP3 as a potential new member of the IFT complex A and showed that several proteins, classified as IFT complex B-associated, are integral parts of this complex. To further demonstrate EPASIS general applicability, we analyzed the modular substructure of two additional complexes, that of B-RAF and of 14-3-3-ε. The results show, that EPASIS provides a robust as well as sensitive strategy to dissect the sub-

structure of large multiprotein complexes in a highly time- as well as cost-effective manner. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.O113.033233, 1382–1391, 2014.

Understanding the orchestration and dynamics of cellular function on the molecular level is one of the challenges in biology. It is now clear that cell regulatory decisions are made by molecular switching events in large but highly dynamic, often coalescing, protein complexes (1, 2). Protein complex isolation by affinity purification is a common technique, used for the identification of the protein composition of these molecular machines (3), contributing to the elucidation of spatial and temporal patterns of large protein networks and functional modules within these networks (4). Interaction data derived from different protein complex analyses are the basis for predictions of biological pathways or disease mechanisms concerning those proteins (5). Still, in most cases it is difficult or even impossible to determine how, or even if the co-purified proteins assemble as a single module in a cell, limiting the fine-grained description of the complex structure (6, 7). The possibility to integrate module and submodule information in higher order protein networks is extremely valuable for their understanding and opens the route to define pathways of information flow within and between discrete molecular machines. Zooming in on a protein mutated in early childhood blindness, lebercilin, we have previously identified proteins of the intraflagellar transport (IFT) machinery to interact with lebercilin (5). IFT appears as a physical entity driving vesicular trafficking through the connecting cilium that bridges the inner and the outer segment of vertebrate photoreceptors (8). IFT, like many other multiprotein complexes, is an example for a functionally fairly well described molecular machine with yet unknown molecular topology and mechanical properties.

To determine composition, as well as protein complex topology of lebercilin and its interaction with IFT components, we developed a novel workflow, which we termed “elution profile analysis of SDS-induced subcomplexes by quantitative

From the ‡Division of Experimental Ophthalmology and Medical Proteome Center, Centre for Ophthalmology, University of Tübingen, 72076 Tübingen, Germany; §Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany; ¶Department of Protein Science, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; ||Department of Human Genetics, **Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands, The Netherlands; §§Center for Human Disease Modeling, Department of Cell Biology, Duke University, Durham, NC 27710

Received August 9, 2013, and in revised form, January 20, 2014.

Published, MCP Papers in Press, February 21, 2014, DOI 10.1074/mcp.O113.033233

Author contributions: Y.T., G.T., M.G., T.J.G., M.U., and K.B. designed research; Y.T., M.G., D.A.M., J.v.R., N.H., and K.B. performed research; G.T., J.W., and N.K. contributed new reagents or analytic tools; Y.T., G.T., D.A.M., J.v.R., N.H., R.R., and K.B. analyzed data; Y.T., N.K., R.R., T.J.G., M.U., and K.B. wrote the paper.

mass spectrometry" (EPASIS)¹. The approach is a combination of affinity purification (AP) with mild destabilization by sodium dodecylsulfate (SDS), enabling gradual decomposition of protein complexes, with quantitative mass spectrometry (MS) and *in silico* elution profile distance analysis (EPD, Fig. 1).

EXPERIMENTAL PROCEDURES

Cell Culture—HEK293-T cells were grown in DMEM (PAA, Pasching, Austria) supplemented with 10% fetal bovine serum and 0.5% Penicillin/Streptomycin. Cells were seeded, grown overnight and then transfected with the corresponding SF-TAP-tagged (9) DNA constructs using PEI reagent (Polysciences, Warrington, PA) according to the manufacturer's instructions. 48 h later, cells were harvested in lysis buffer containing 0.5% Nonidet-P40 (N P-40), protease inhibitor mixture (Roche, Freiburg, Germany), and phosphatase inhibitor cocktails II and III (Sigma-Aldrich, Taufkirchen, Germany) in TBS (30 mM Tris-HCl, pH 7.4, and 150 mM NaCl) for 20 min at 4 °C. Cell debris and nuclei were removed by centrifugation at $10,000 \times g$ for 10 min.

Protein Complex Destabilization—For protein complex destabilization, the cleared lysates were transferred to anti-FLAG M2 agarose (Sigma-Aldrich). After one hour of incubation, the resin was washed three times using wash buffer (TBS containing 0.1% N P-40 and phosphatase inhibitor cocktails II and III, Sigma-Aldrich). For the SDS-destabilization of the protein complexes, the resin was then incubated 3 min with each concentration of SDS (0.00025%, 0.0025%, 0.005%, 0.01%, and 0.1% for lebercilin or 0.00025%, 0.002%, 0.004%, 0.008%, 0.016%, 0.05% for 14–3–3-ε and B-RAF) in SDS-elution buffer (TBS containing phosphatase inhibitor cocktails II and III) at 4 °C. The flow through was collected and precipitated by methanol-chloroform. After every elution step a single wash step was performed. Subsequent to the SDS gradient, the remaining proteins were eluted from the resin by incubation for 3 min with FLAG peptide (200 μg/ml; Sigma-Aldrich) in wash buffer. A schematic representation is given in Fig. 1.

Tandem Affinity Purification (SF-TAP)—For SF-TAP, the cleared lysates were incubated for one hour at 4 °C with Strep-Tactin superflow (IBA, Göttingen, Germany). Subsequently, the resin was washed three times in wash buffer (TBS containing 0.1% N P-40 and phosphatase inhibitor cocktails II and III, Sigma-Aldrich). Proteins were eluted with desthiobiotin (2 mM in TBS). For the second purification step, the eluates were transferred to anti-FLAG M2 agarose (Sigma) and incubated for one hour at 4 °C. The beads were washed three times with wash buffer and proteins were eluted with FLAG peptide (200 μg/ml, Sigma-Aldrich) in wash buffer.

Immunoblot Analysis—Equal amounts of eluates were separated by SDS-PAGE and electrophoretically transferred to PVDF membranes. Membranes were blocked in 5% nonfat milk in TBS/0.1% Tween20 (Sigma-Aldrich) and incubated with anti-TRAF3IP1 (1:1000, mouse, Abnova, Taipei City, Taiwan), anti-14-3-3-ε (1:1000, rabbit, Santa Cruz, Santa Cruz), anti-C20orf11 (GID8, 1:1000, rabbit, Sigma-

Aldrich), and anti-FLAG-HRP (1:1000, mouse, Sigma-Aldrich). Secondary antibodies from Jackson ImmunoResearch (West Grove, PA) were applied (1:15,000) and protein bands were visualized using ECL plus (GE Healthcare, Freiburg, Germany). See Supplemental Fig. 9.

Mass Spectrometric Analysis—After precipitation of the proteins by methanol-chloroform, a tryptic in-solution digestion was performed as described previously (10). LC-MS/MS analysis was performed on a NanoRSLC3000 HPLC system (Dionex, Idstein, Germany) coupled to a LTQ OrbitrapXL, or to a LTQ OrbitrapVelos mass spectrometer (Thermo Fisher Scientific, Bonn, Germany) by a nano spray ion source. Tryptic peptide mixtures were automatically injected and loaded at a flow rate of 6 μl/min in 98% buffer C (0.1% trifluoroacetic acid in HPLC-grade water) and 2% buffer B (80% acetonitrile and 0.08% formic acid in HPLC-grade water) onto a nano trap column (75 μm i.d. $\times$ 2 cm, packed with Acclaim PepMap100 C18, 3 μm, 100 Å; Dionex). After 5 min, peptides were eluted and separated on the analytical column (75 μm i.d. $\times$ 25 cm, Acclaim PepMap RSLC C18, 2 μm, 100 Å; Dionex) by a linear gradient from 2% to 35% of buffer B in buffer A (2% acetonitrile and 0.1% formic acid in HPLC-grade water) at a flow rate of 300 nL/min over 33 min for EPASIS samples, respectively over 80 min for SF-TAP samples. Remaining peptides were eluted by a short gradient from 35% to 95% buffer B in 5 min. The eluted peptides were analyzed by using a LTQ Orbitrap XL, or a LTQ OrbitrapVelos mass spectrometer. From the high-resolution mass spectrometry prescan with a mass range of 300–1500, the ten most intense peptide ions were selected for fragment analysis in the linear ion trap if they exceeded an intensity of at least 200 counts and if they were at least doubly charged. The normalized collision energy for collision-induced dissociation was set to a value of 35, and the resulting fragments were detected with normal resolution in the linear ion trap. The lock mass option was activated and set to a background signal with a mass of 445.12002 (11). Every ion selected for fragmentation was excluded for 20 s by dynamic exclusion.

For qualitative results, the raw data were analyzed using Mascot (Matrix Science, Boston, USA; version 2.4.0) and Scaffold (version Scaffold_4.0.3, Proteome Software Inc., Portland, USA). Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by extract_msn.exe version 5.0. All MS/MS samples were analyzed using Mascot. Mascot was set up to search the Swiss-Prot_2012_05 database (selected for Homo sapiens, 2012_05, 20245 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 1.00 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine and glutamine and oxidation of methionine were specified in Mascot as variable modifications. Scaffold was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm (12) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (13). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Furthermore, proteins were only considered to be specific protein complex components if they were not detected in the control experiments.

For quantitative analysis, MS raw data were processed using the MaxQuant software (version 1.3.0.5 (14)). Trypsin/P was set as cleaving enzyme. Cysteine carbamidomethylation was selected as fixed modification, methionine oxidation, and protein acetylation were allowed as variable modifications. Two missed cleavages per peptide were allowed. The peptide and protein false discovery rates were set to 1%. The initial mass tolerance for precursor ions was set to 6 ppm

¹ The abbreviations used are: EPASIS, Elution profile analysis of SDS-induced subcomplexes by quantitative mass spectrometry; AP, Affinity purification; BN-PAGE, Blue Native Polyacrylamide gel electrophoresis; CMC, Critical micellar concentration; CTLH, C-terminal to Lissencephaly type-1-like homology motif; EPD, Elution profile distance; GPCR, G-protein coupled receptor; IFT, Intraflagellar transport; IFT-A, Intraflagellar transport complex A; IFT-B, Intraflagellar transport complex B; PCP, Protein correlation profiling; PPM, Parts Per Million; SF-TAP, Strep-FLAG-tandem affinity purification; FLAG-IP, FLAG immunoprecipitation.

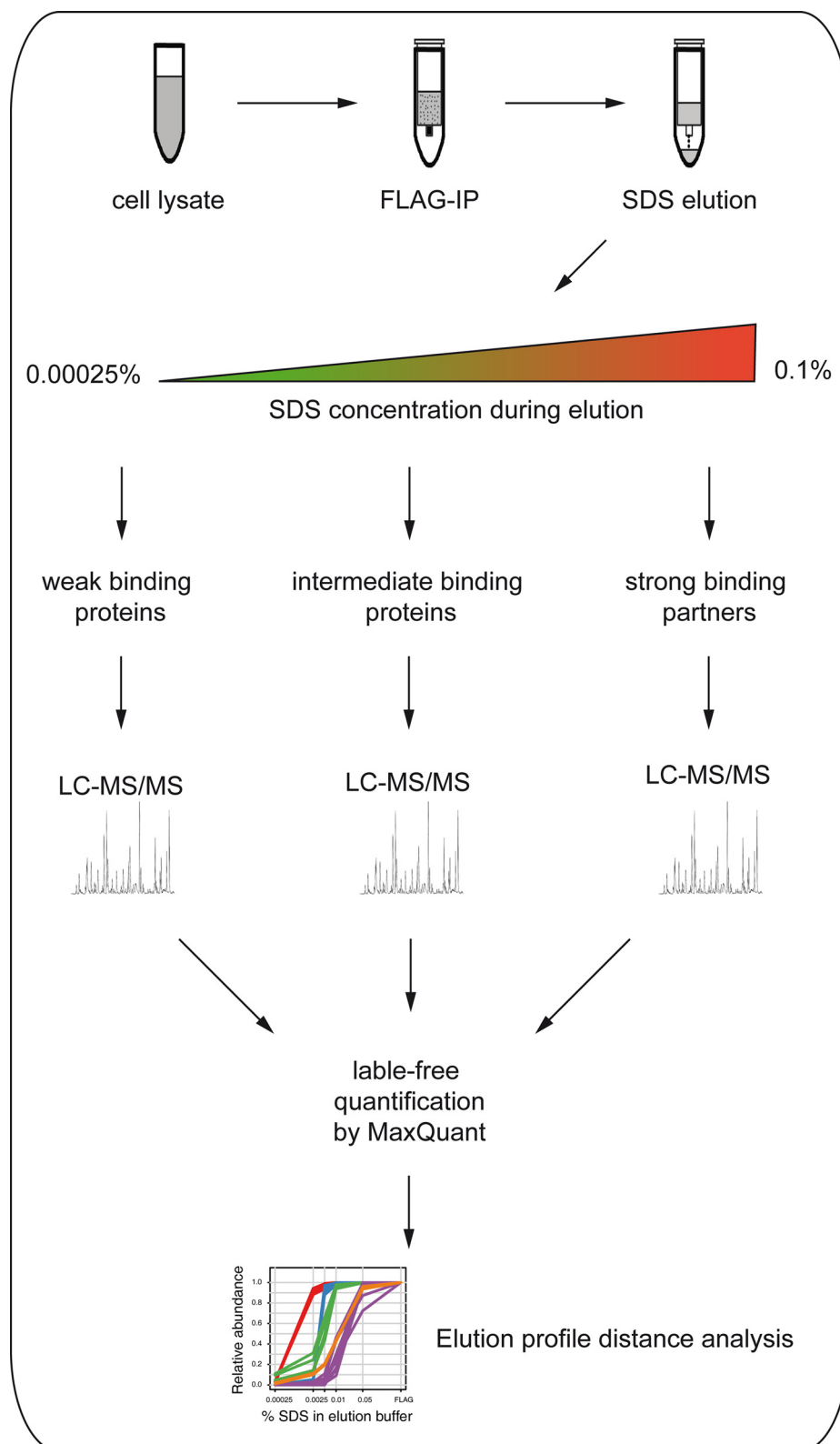


FIG. 1. **Workflow of EPASIS.** The protein complex is purified by FLAG immunoprecipitation (FLAG-IP) and eluted by a SDS concentration gradient. The eluates are analyzed by mass spectrometry, quantified by label-free quantification (MaxQuant) followed by elution profile distance analysis to detect the submodule composition of complexes.

and the first search option was enabled with 10 ppm precursor mass tolerance. The fragment ion mass tolerance was set to 0.5 Da. The human subset of the human proteome reference set provided by SwissProt (Release 2012_01 534,242 entries) was used for peptide and protein identification. Contaminants like keratins were automatically detected by enabling the MaxQuant contaminant database search. A minimum number of two unique peptides with a minimum length of seven amino acids needed to be detected to perform protein quantification. Only unique peptides were selected for quantification. For label-free quantification the minimum LFQ count was set to three, the re-quantify option was chosen. The option match between runs was enabled with a time window of 2 min, fast LFQ was disabled (see also parameters.txt in [supplementary File S1](#)).

Statistical Data Analysis—Statistical analysis of the data was carried out using R (15). For the eight IFT/lebercilin vector experiments (48 measurements), in total 175471 unique peptides with minimum peptide length of seven amino acids were identified by searching against the forward version of the database and only 239 unique peptides were identified by searching against a reversed version of the database which indicates a peptide false positive identification rate of 0.14% (239/175471). Without filtering, 1081 proteins were detected for the forward search and ten for the reverse search leading to an indicated protein false positive identification rate of 0.93% (10/1081). To reduce the number of false positive protein identifications, proteins were considered as detected, if they were identified by at least two unique peptides, had a minimal MS/MS spectra count of three (760/1081) and were not flagged as contaminant by MaxQuant (731/760). Additionally, three repeated experiments (18 measurements) using an empty vector as control were performed and the same filter criteria were applied. The Euclidian distance between proteins both detected in the control and the IFT/lebercilin experiments was calculated and the proteins were excluded from further considerations if they showed a distance less than 0.1 in two or more out of 24 comparisons between both experiments (205/731). Finally, proteins had to be present in at least 5/8 (62.5%) repeated experiments, resulting in a total of 290 Proteins that were further analyzed.

For the protein complexes of 14-3-3- ϵ and B-RAF, five experiments were performed. Additionally, five repeated experiments using the SF-TAP vector as a control were performed with the same SDS-gradient. The statistical analysis was performed as described above, leading to a total number of 135 proteins for 14-3-3- ϵ and 32 proteins for B-RAF, which were further analyzed.

Reproducibility of Elution Profiles—Protein intensities for all SDS concentrations of an experiment were combined and the values log2-transformed. To investigate the linear relationship between data points, regression lines determined by minimizing the sum of squares of the Euclidean distance of points to the fitted line (“orthogonal regression”) are shown in Fig. 2 ([supplemental Fig. S1 and S2](#) for 14-3-3- ϵ and B-RAF). Correlations between repeated experiments were estimated using the Pearson correlation coefficient together with its 95% confidence interval. To investigate the safe isolation of elution profiles for different SDS concentrations, Spearman’s correlation scores were calculated and plotted in [supplemental Fig. S3](#) ([supplemental Fig. S4 and S5](#) for 14-3-3- ϵ and B-RAF).

Protein Consensus Profiles—Consensus profiles of known marker protein groups ([supplemental Table S1](#) for lebercilin and [supplemental Table S2 and S3](#) for 14-3-3- ϵ and B-RAF) were calculated by averaging the normalized cumulative intensities of the protein group per concentration step for all experiments, similar to Andersen *et al.* (16). The elution profile distance between a protein and a consensus profile was calculated as:

$$epd(x, c) = \sqrt{\frac{\sum_{i=1}^n (x_i - c_i)^2}{n - 1}}$$

with x being the cumulative intensity of a protein, c the average cumulative intensity of the consensus profile and the number of the fraction, i . Dividing the Euclidean distance by the maximum possible cumulative elution profile distance ($\sqrt{n - 1}$), allows to compare EPASIS experiments with different numbers of SDS-concentrations. Elution profile distances (EPD) to consensus profiles were calculated for all detected proteins (lebercilin $n = 290$, 14-3-3- ϵ $n = 135$, B-RAF $n = 32$). A stepwise ($n = 1000$) parameter search was performed to estimate the optimal EPD threshold to maximize the specificity and sensitivity to assign known subcomplex members to the consensus profile (lebercilin [supplemental Fig. S6](#), 14-3-3- ϵ [supplemental Fig. S7](#), B-RAF [supplemental Fig. S8](#)). For lebercilin, 60 new candidate proteins for the reference subcomplexes were identified (27 for 14-3-3- ϵ and six for B-RAF), by using the identified EPD threshold of 0.11 (0.064 for 14-3-3- ϵ and 0.155 for B-RAF). To perform nonmetric multidimensional scaling, the elution profiles were averaged across the experiments ($n = 8$ for lebercilin and $n = 5$ for 14-3-3- ϵ and B-RAF) and Euclidean distances between them were calculated. A stable solution was estimated by using random starts (17) and the best ordination (stress: 0.03 for lebercilin and 0.04 for 14-3-3- ϵ and B-RAF) was selected (Fig. 3B for lebercilin, Fig. 4B and 4D for 14-3-3- ϵ and B-RAF).

RESULTS

Following FLAG-based AP of Strep/FLAG tandem affinity purification tag (SF-TAP)-fused (9) lebercilin from HEK293T cells, we destabilized the purified protein complexes by treatment with very low concentrations of SDS. This approach was previously described in combination with Blue Native polyacrylamide gel electrophoresis (BN-PAGE) (18). The underlying mechanism is based on hydrophobic interaction of monomeric SDS with the proteins, starting way below critical micellar concentration (CMC) (19), and resulting in a partial destabilization of the tertiary structure (20). The destabilization leads to the sequential elution of proteins, depending on the sensitivity of their interaction with IFT/lebercilin to SDS and employs the fact that low concentrations of SDS can be used to destabilize noncovalent binding of proteins. Assuming that the binding affinity of proteins within a single submodule is higher than their affinity to proteins outside a module, a stepwise increase of the SDS concentration will lead to early decomposition of labile interactions at low concentrations of SDS, whereas binding within a submodule, stabilized by affinity, avidity and possible binding partners as well as docking motifs, decomposes at higher concentrations. It is important to mention here, that the stability of interactions within a submodule does not need to be equal. To discriminate a submodule of a larger protein complex, its resistance to dissociation only needs to be higher than the stability of its interaction with the bait protein. To increase the sensitivity, robustness and feasibility of the approach we applied highly sensitive MS in combination with label-free based quantification (21) and a refined *in silico* protein correlation profiling (PCP) approach. The latter procedure is based on calculating the similarity of elution profiles to a consensus profile of known complex members (16).

The destabilization of the IFT/lebercilin complex resulted in at least five different subcomplexes coeluting with distin-

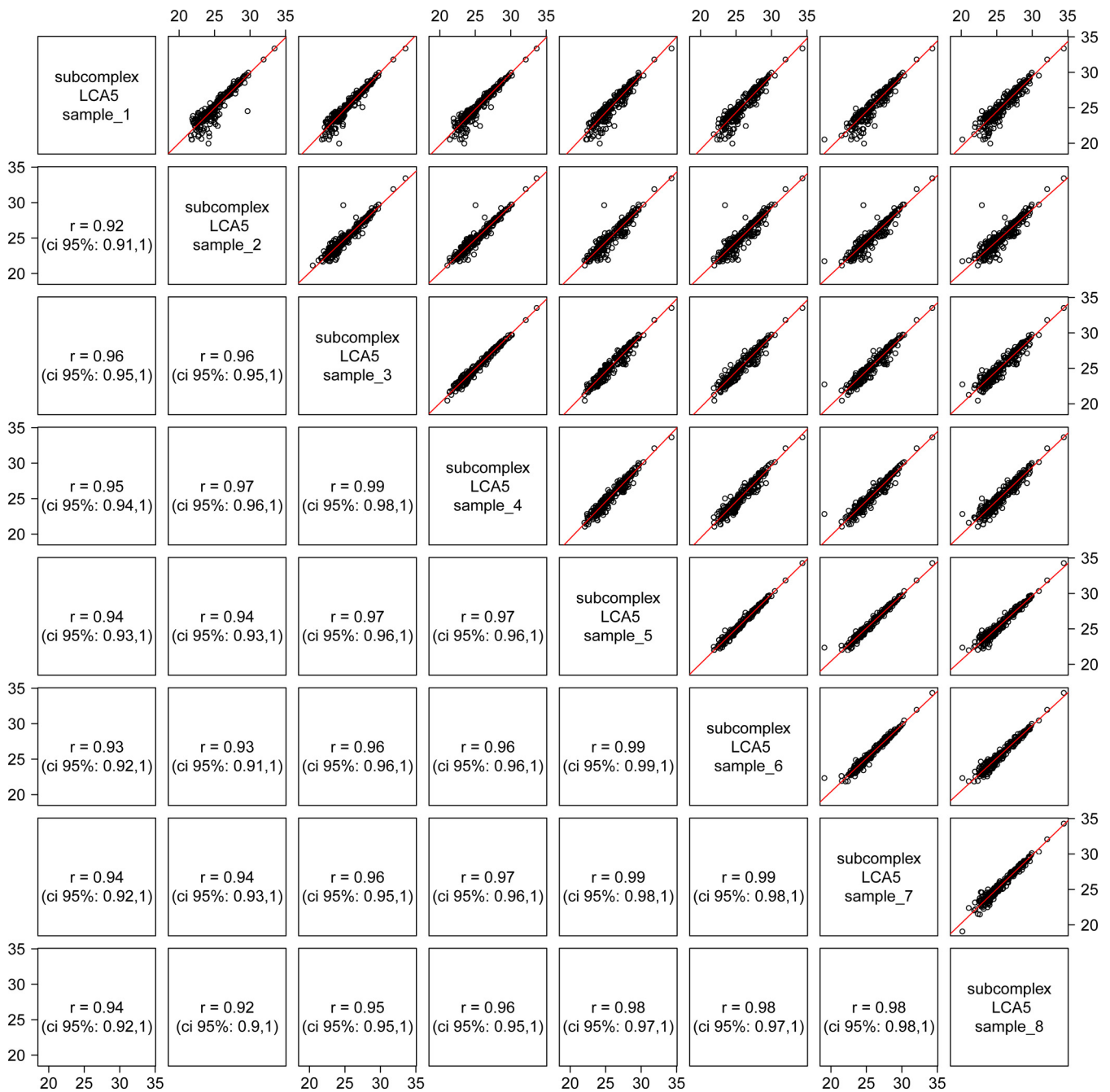


FIG. 2. **Reproducibility.** Scatter plots of log2-transformed protein ($n = 290$) intensities from replicated experiments (experiment 1–8). Orthogonal regression lines are shown in red; Pearson correlation coefficients (r) and their 95% confidence intervals (ci) are shown.

guishable profiles (Fig. 3, [supplemental Fig. S9](#) and [supplemental Table S4](#)), confirming already postulated submodules (5). Consensus profiles for known modules of the IFT/lebercilin complex were calculated and the EPD to these consensus profiles was determined. Candidate proteins were selected based on a short EPD (≤ 0.11 , [supplemental Fig. S6](#)) to a consensus profile. The robustness of the approach is demonstrated by the high reproducibility of the results obtained from eight independent experiments (Fig. 2, [supplemental Fig. S3](#)).

By using consensus profiles for IFT complex A (IFT-A) and IFT complex B (IFT-B), we can verify that IFT-A and IFT-B exist as two discrete submodules, eluting from the IFT/lebercilin complex at different SDS concentrations with clearly separated profile values. The IFT-A elutes at a very low SDS concentration (0.0025%) with an EPD-value of less than or equal to 0.02 for all known IFT-A proteins ($n = 6$) to its consensus profile (EPD_{IFT-A}), while all IFT-B proteins had values greater than or equal to 0.382 for the IFT-A consensus profile.

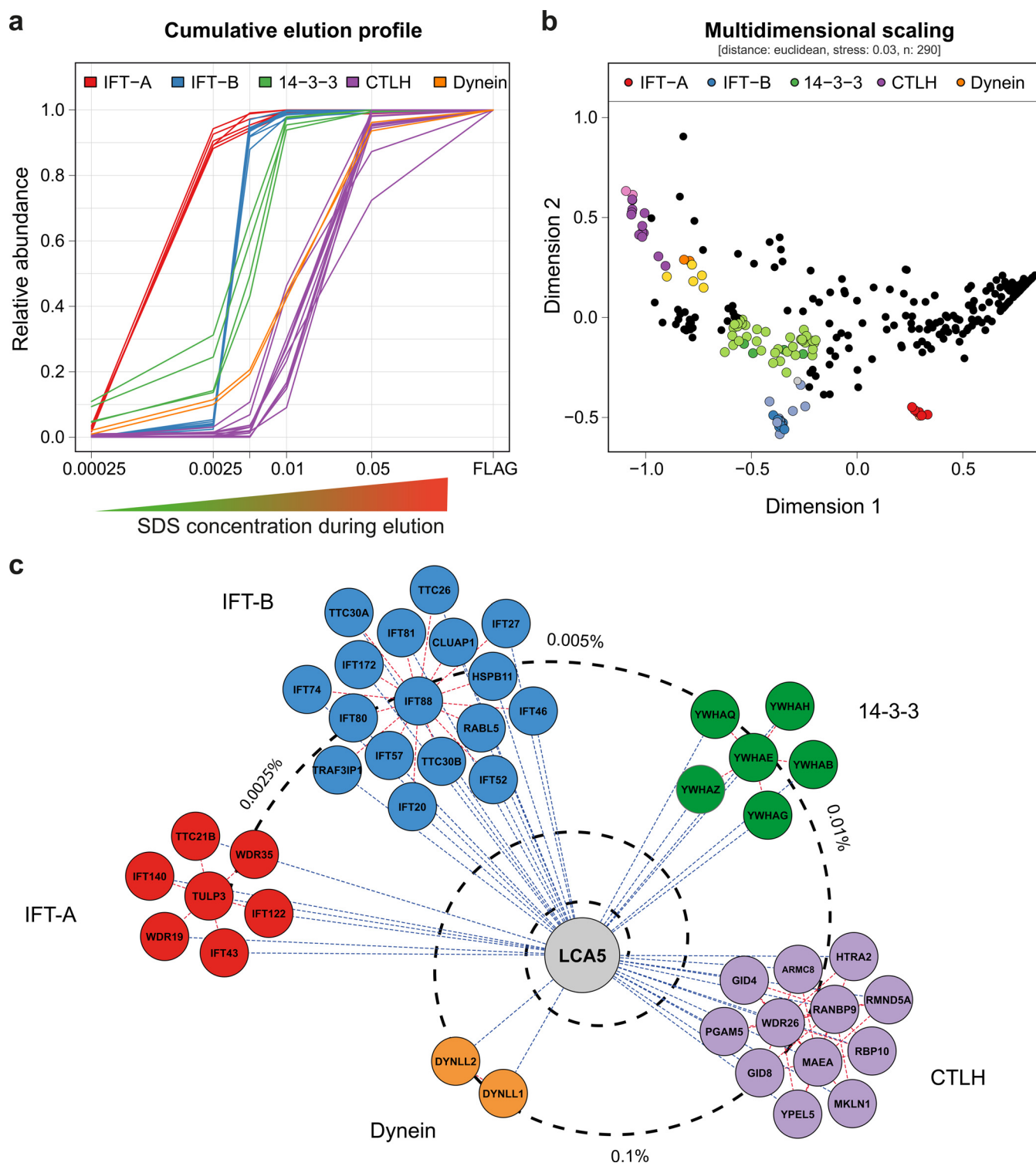


FIG. 3. EPASIS of the IFT/lebercilin protein complex. *A*, Visualization of the elution profiles after analyzing the eluted fractions by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) and label-free quantification. The cumulative relative abundance (y axis) is plotted against increasing SDS concentrations (x axis). For each submodule, consensus profiles were generated from known members. Color lines represent known members of the corresponding subcomplex. *B*, Nonmetric multidimensional scaling ordination plot based on Euclidean distances of elution profiles (stress 0.03). Data points ($n = 290$) present the average of replicated data ($n = 8$). *C*, The predicted network of the interactome of lebercilin. The modules are grouped according to their elution from the IFT/lebercilin complex during the EPASIS approach. The closer they are to lebercilin the higher the SDS-concentration that is needed for their complete dissociation. The dashed spiral line symbolizes the different SDS concentrations. The interactions shown by blue, dashed lines were determined by EPASIS, the red, dashed lines by SF-TAP.

IFT-B components elute mainly at a SDS concentration of 0.005%, showing tightly clustered elution profiles for all the known IFT-B members ($EPD_{IFT-B} \leq 0.025$, $n = 10$, EPD_{IFT-B} for all IFT-A proteins ≥ 0.378). In addition to the proteins already previously described as components of the IFT-B, proteins which we termed “IFT-B associated proteins” also perfectly co-elute with the IFT complex B. Their association to the IFT-B was further validated and reproduced by analyzing the interactome of SF-TAP-tagged HSPB11 and TRAF3IP1 (supplemental Table S5). This provides further evidence that these proteins are integral parts of the IFT complex B.

The phosphopeptide-binding 14-3-3 regulatory protein isoforms mainly elute at concentrations of 0.005% and 0.01% SDS from the IFT/lebercilin complex. Previous studies have shown that the 14-3-3 proteins can form a stable complex (9, 22). Relative to the other subcomplexes, here the 14-3-3 proteins form a less homogeneous group of elution profiles ($EPD_{14-3-3} \leq 0.071$, $n = 4$). Nevertheless, it is possible to clearly distinguish them from the other submodules ($EPD_{IFT-A} \geq 0.349$, $EPD_{IFT-B} \geq 0.173$, $EPD_{CTLH} \geq 0.309$, $EPD_{DYN} \geq 0.283$). It needs to be noted that the group of proteins with an $EPD_{14-3-3} \leq 0.11$ is quite large and does not only include 14-3-3 isoforms. It also contains a number of ribosomal proteins, for example. In this case, the resolution of the gradient chosen is not high enough to clearly isolate the 14-3-3 subcomplex from other potential subcomplexes with similar elution profiles.

The two isoforms of dynein light chain (DYNLL1, DYNLL2) are among the strongest binders to the IFT/lebercilin complex and can also be defined as a submodule by our method ($EPD_{DYN} \leq 0.009$, $n = 2$, $EPD_{CTLH} \geq 0.113$). The dynein light chains elute at the same SDS concentration as the lebercilin bait protein.

Several proteins that were described as miscellaneous in our previous study (5) also seem to co-elute with similar profiles to the dynein light chains. However, the determination of the EPD-value showed that these proteins can be separated from dynein and lebercilin and therefore can be considered to be a different submodule ($EPD_{CTLH} \leq 0.101$, $n = 12$, $EPD_{DYN} \geq 0.111$, except for the protein C20orf11 (GID8) $EPD_{DYN} = 0.058$ and the protein ARMC8 $EPD_{DYN} = 0.080$). Some proteins of this miscellaneous set have been reported to be part of a complex termed CTLH that is thought to be involved in microtubular dynamics (23).

In addition to identifying the affinity-based substructure of the previously described complexes, the EPASIS approach revealed novel, potential protein components of known complexes. The tubby-related protein 3 (TULP3), for instance, was identified using EPASIS with an EPD to IFT-A of 0.010. TULP3 was previously described as being associated with the IFT complex A (24). To confirm this, and to validate our method, we analyzed the interactome of SF-TAP-tagged TULP3. This resulted in the detection of all IFT-A proteins within the com-

plex retrieved with the tagged TULP3 (supplemental Table S7, Fig. 3).

The EPASIS method can be applied to many different protein complexes. To show its applicability to protein complexes of different composition and function, we applied it to two additional complex-forming proteins, 14-3-3- ϵ and B-RAF. Protein complexes for both SF-TAP-tagged baits were purified by one-step FLAG-affinity purification.

The reference proteins for possible submodules within the 14-3-3- ϵ complex (supplemental Table S2) were selected from the literature (9, 25) and the String interaction database (www.string-db.org). The destabilization of the protein complex of 14-3-3- ϵ showed several submodules eluting from the complex (Fig. 4A and 4B). Five 14-3-3 proteins dissociated as a stable submodule from the complex ($EPD_{14-3-3} \leq 0.064$). Additionally, a complex of three MARK (microtubule affinity-regulating kinase) proteins ($EPD_{MARK} \leq 0.038$) and two submodules of kinesin proteins could be detected. The latter two modules represent the kinesin light chain proteins ($EPD_{KIN-LC} \leq 0.042$, $EPD_{KIN-HC} \geq 0.079$) and the kinesin heavy chain proteins ($EPD_{KIN-HC} \leq 0.035$, $EPD_{KIN-LC} \geq 0.040$) that can be separated by their EPD-values. The complete list of the detected interactors and potential candidates of the corresponding submodules is shown in supplemental Table S8.

The possible submodules for the B-RAF protein complex (supplemental Table S3) were also chosen from literature. The detected protein complex of B-RAF consists of clearly less complex members compared with the other two baits (Fig. 4C and 4D). After statistical data analysis, a list of only 32 specific interaction partners (supplemental Table S9) remained for further evaluation. Within those, the 14-3-3 protein complex could be determined as a submodule ($EPD_{14-3-3} \leq 0.055$). Also a second, very specific interacting submodule of B-RAF, the mitogen activated pathway kinases (MAPK1 and MAPK2, $EPD_{MAPK} \leq 0.012$) eluted from the complex. Furthermore, a module consisting of the two HSP90-subunits could be observed ($EPD_{HSP90} \leq 0.011$).

DISCUSSION

We here describe a novel approach to decipher the topology and submodule composition of protein complexes from affinity purified protein complex mixtures, which we termed EPASIS. Using the EPASIS approach, we were able to comprehensively describe the substructure of the IFT/lebercilin complex by showing that it consists of several, clearly separable submodules. We were able to recapitulate both, IFT-A and IFT-B using a tagged interactor that is not integral to either submodule. In addition, we were able to strengthen the hypothesis that proteins classified as associated interactors may be integral parts of particular subcomplexes.

The detection of a novel member of the IFT-A, TULP3, shows the power of this method in combining highly sensitive quantitative MS and in silico EPD analysis. A previous study showed that the protein TULP3 is associated to the IFT-A but

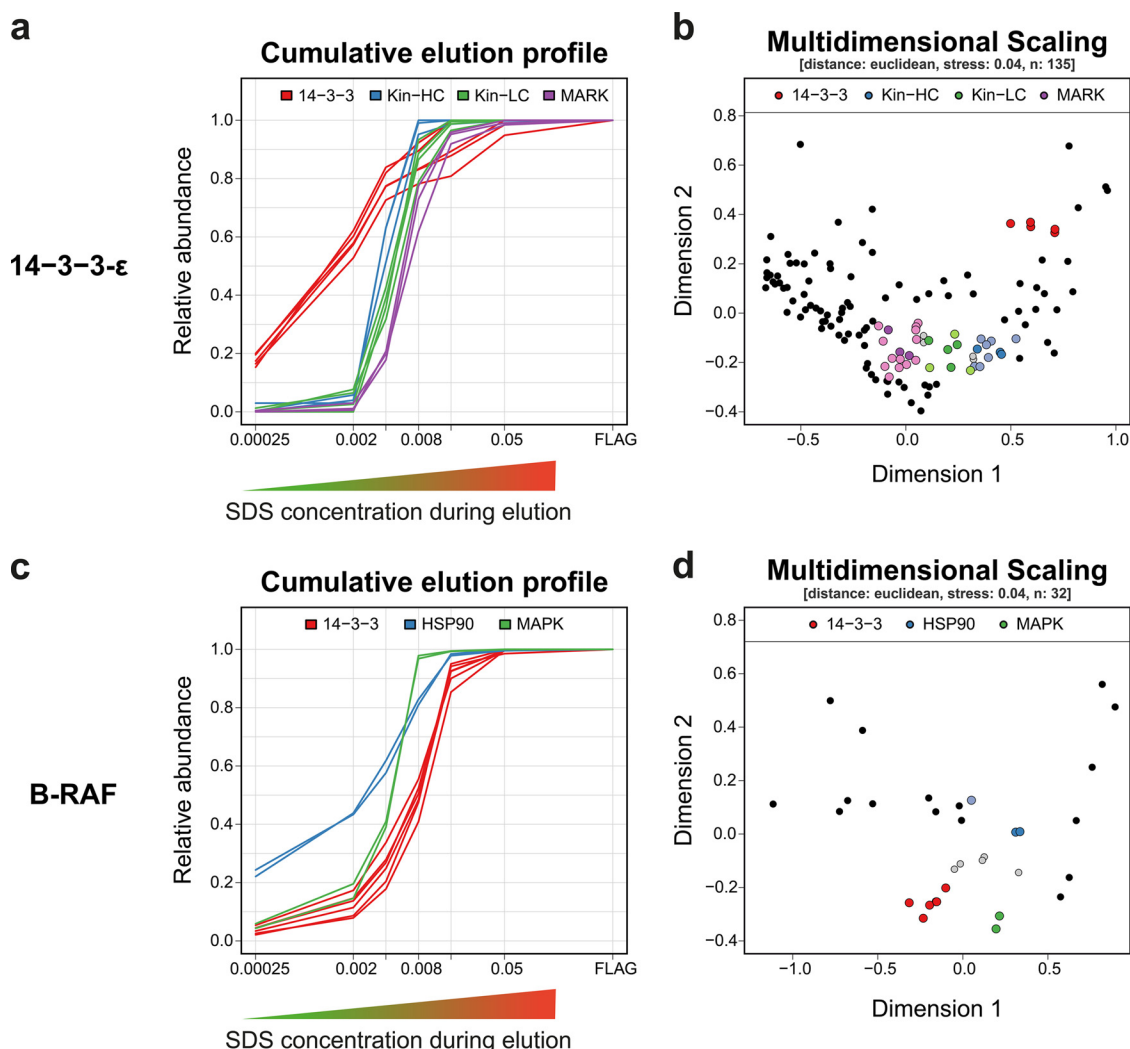


FIG. 4. EPASIS of the protein complexes of 14-3-3- ϵ and B-RAF. A, Cumulative elution profiles of the reference proteins of the 14-3-3- ϵ complex by increasing the SDS-concentration from 0.00025% to 0.05%. The consensus group of the 14-3-3 proteins is shown in red and is the first group eluting from the complex. In blue the group of kinesin heavy chain proteins (Kin-HC) is shown and elutes slightly before the kinesin light chain proteins (Kin-LC, green). The consensus group eluting at the highest SDS concentration is a group of microtubule affinity-regulating kinases (MARK, violet). B, Nonmetric multidimensional scaling ordination plot of the proteins eluting from the 14-3-3- ϵ complex, based on Euclidean distances of elution profiles (stress 0.04). Data points ($n = 135$) present the average of replicated data ($n = 5$). C, Cumulative elution profiles of three consensus protein groups of the B-RAF complex induced by increasing concentrations of SDS (0.00025%-0.05%). Two HSP90 proteins elute first from the complex (blue). The second eluting consensus group consists of the two mitogen activated kinases MAPK1 and MAPK2 (green). The strongest binding consensus proteins are the 14-3-3 proteins (red). D, Nonmetric multidimensional scaling ordination plot of the proteins eluting from the B-RAF complex, based on Euclidean distances of elution profiles (stress 0.04). Data points ($n = 32$) present the average of replicated data ($n = 5$).

not that it should be considered integral. At least under our experimental conditions, IFT-A is the primary interactor for TULP3. In concert with IFT-A proteins, TULP3 is by its phosphoinositide-binding properties essential for ciliary GPCR trafficking and negatively regulates Hedgehog signaling (24).

The more stable binding of dynein light chain, as well as of the 14-3-3 protein isoforms, compared with IFT proteins, is in-line with the fact that these were among the few proteins that could previously be identified by lebercilin SF-TAP purification (9).

The detection of the complete CTLH complex as a linked submodule demonstrates the sensitivity of the EPASIS approach. In contrast to SF-TAP, where only RANBP9 was detected, and to the Strep-SILAC approach where we could only detect some of the components (RANBP9, C20ORF11 (GID8), YPEL5, WDR26), with EPASIS we were able to identify the CTLH complex as a distinct submodule of the IFT/lebercilin protein complex. In combination with SF-TAP analysis of several complex members (Supplemental Table 5, Fig. 3), we here present a comprehensive description of this complex

that will help to elucidate its as yet only vaguely described function (26). Further effort will be needed to determine the function of this complex, possibly in the context of dynein-driven IFT as suggested by EPASIS.

To demonstrate that the EPASIS method is also applicable to protein complexes with different functions and compositions, we applied the strategy to the complex of the adapter protein 14-3-3- ϵ as well as to the one of the serine/threonine kinase B-RAF. For both protein complexes, we could detect several different submodules of various functions. The results clearly showed that this method allows the detection of stable submodules being released from the protein complexes as discrete submodules. Nevertheless, the destabilization of these two additional protein complexes showed, that not all previously described interactors can be detected within such submodules. This might be explained by the fact that many of these proteins are only weakly associated to complexes and act as regulators or linkers, which bind to complexes as single entities. Furthermore, it is possible that the low avidity within submodules leads to their disassembly because of the application of SDS instead of detaching from the complex as submodules. To further study those, additional experiments with baits within these suspected submodules would be necessary. Nevertheless, the two examples of complexes with highly regulated functions in the cellular signaling machinery demonstrate that EPASIS is also applicable to these types of proteins and not only to scaffold proteins like lebercilin.

Taken together, the EPASIS strategy is a powerful and sensitive tool to determine the molecular topology and composition of protein complexes in a highly time- as well as cost-effective and robust manner. In this study, we used a tag-fusion protein, transiently transfected to HEK293-T cells. However, EPASIS will be equally applicable to tagged proteins being expressed in stably engineered cell lines, which are becoming increasingly popular in the study of cell regulation (27, 28). Furthermore, it can be applied for analysis of immuno-precipitated and affinity-purified protein complexes from various origins. EPASIS can easily be adapted to different complexes by varying the concentrations of SDS or the profile of the gradient. Moreover, as shown here, the application of EPASIS not only allows the validation of previously known, or postulated submodules but also the identification of new protein interactions and their annotation to a specific complex or subcomplex. This contributes essential information toward experimental description of the molecular architecture and topology of large protein networks and allows validation of *in silico* predictions of cellular machineries (6, 7) as well as evaluation of their malfunction because of gene mutations (5). Protein complexes are the reaction centers that relay, propagate and transmit cell regulatory decisions. When combined with pre-existing knowledge or prediction tools such as socio-affinity index (6, 7) and analysis tools like cytoscape (www.cytoscape.org) or Reactome (www.reactome.org), just to mention a few, EPASIS contributes to test pre-

dicted modularities as well as dynamic changes within signaling pathways as a response to cellular or genetic cues.

Acknowledgments—We thank Gabrielle Wheway for critical proof-reading on the manuscript.

* This work was supported by the European Community's Seventh Framework Programme FP7 under grant agreement no. 241955; SYSCILIA (to R. Roepman, M. Ueffing, N. Katsanis and T. J. Gibson), FP7 grant agreement no. 278568, PRIMES (to M. Ueffing and K. Boldt), FP7 grant agreement no. 241481, AFFINOMICS (to M. Ueffing) and the Kerstan Foundation (to M. Ueffing), and the Netherlands Organization for Scientific research (NWO Vici-016.130.664 to R. Roepman).

§ This article contains supplemental Figs. S1 to S9, Tables S1 to S9, and Files S1 and S2.

¶¶ To whom correspondence should be addressed: Medical Proteome Center, University of Tuebingen, Naegelstrasse 5, Tuebingen, 72074 Germany. Tel.: 0049 7071 29 84950; Fax: 0049 7071 29 5777; E-mail: karsten.boldt@uni-tuebingen.de.

||| These authors contributed equally to this work.

REFERENCES

1. Scott, J. D., and Pawson, T. (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224
2. Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J., and Davey, N. E. (2013) The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.* **6**, rs7
3. Kocher, T., and Superti-Furga, G. (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **4**, 807–815
4. Kiel, C., Vogt, A., Campagna, A., Chatr-aryamontri, A., Swiatek-de Lange, M., Beer, M., Bolz, S., Mack, A. F., Kinkl, N., Cesareni, G., Serrano, L., and Ueffing, M. (2011) Structural and functional protein network analyses predict novel signaling functions for rhodopsin. *Mol. Syst. Biol.* **7**, 551
5. Boldt, K., Mans, D. A., Won, J., van Reeuwijk, J., Vogt, A., Kinkl, N., Letteboer, S. J., Hicks, W. L., Hurd, R. E., Naggert, J. K., Texier, Y., den Hollander, A. I., Koenekoop, R. K., Bennett, J., Cremers, F. P., Gloeckner, C. J., Nishina, P. M., Roepman, R., and Ueffing, M. (2011) Disruption of intraflagellar protein transport in photoreceptor cilia causes Leber congenital amaurosis in humans and mice. *J. Clin. Invest.* **121**, 2169–2180
6. Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8**, 645–654
7. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurlier, 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., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636
8. Fliegauf, M., Benzing, T., and Omran, H. (2007) When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell Biol.* **8**, 880–893
9. Gloeckner, C. J., Boldt, K., Schumacher, A., Roepman, R., and Ueffing, M. (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics* **7**, 4228–4234
10. Gloeckner, C. J., Boldt, K., and Ueffing, M. (2009) Strep/FLAG tandem affinity purification (SF-TAP) to study protein interactions. *Curr. Protoc. Protein Sci.* Chapter 19, Unit19 20
11. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Hornig, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021
12. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made

- by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
13. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
 14. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
 15. R-Core-Team (2012) R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, <http://www.R-project.org/>.
 16. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574
 17. Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2013) vegan: Community Ecology Package. R package version 2.1–27/r2470. .
 18. Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L., and Braun, H. P. (2010) Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. *Plant Cell* **22**, 797–810
 19. Garavito, R. M., and Ferguson-Miller, S. (2001) Detergents as tools in membrane biochemistry. *J. Biol. Chem.* **276**, 32403–32406
 20. Bhuyan, A. K. (2010) On the mechanism of SDS-induced protein denaturation. *Biopolymers* **93**, 186–199
 21. Lubner, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keefe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32**, 279–289
 22. Obsil, T., and Obsilova, V. (2011) Structural basis of 14–3-3 protein functions. *Semin. Cell Dev. Biol.* **22**, 663–672
 23. Tomastikova, E., Ceniklova, V., Kohoutova, L., Petrovska, B., Vachova, L., Halada, P., Kocarova, G., and Binarova, P. (2012) Interactions of an *Arabidopsis* RanBPM homologue with LisH-CTLH domain proteins revealed high conservation of CTLH complexes in eukaryotes. *BMC Plant Biol.* **12**, 83
 24. Mukhopadhyay, S., Wen, X., Chih, B., Nelson, C. D., Lane, W. S., Scales, S. J., and Jackson, P. K. (2010) TULP3 bridges the IFT-A complex and membrane phosphoinositides to promote trafficking of G protein-coupled receptors into primary cilia. *Genes Dev.* **24**, 2180–2193
 25. Gloeckner, C. J., Boldt, K., Schumacher, A., and Ueffing, M. (2009) Tandem affinity purification of protein complexes from mammalian cells by the Strep/FLAG (SF)-TAP tag. *Methods Mol. Biol.* **564**, 359–372
 26. Kobayashi, N., Yang, J., Ueda, A., Suzuki, T., Tomaru, K., Takeno, M., Okuda, K., and Ishigatsubo, Y. (2007) RanBPM, Muskelein, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* **396**, 236–247
 27. Poser, I., Sarov, M., Hutchins, J. R., Heriche, J. K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechler, K., Durbin, R., Stewart, A. F., Peters, J. M., Buchholz, F., and Hyman, A. A. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409–415
 28. Gibson, T. J., Seiler, M., and Veitia, R. A. (2013) The transience of transient overexpression. *Nat. Methods* **10**, 715–721