Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Methods 53 (2011) 13–19

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Purification and characterization of HIV–human protein complexes

Stefanie Jäger [a], Natali Gulbahce [a], Peter Cimermancic [a], Joshua Kane [a], Nanhai He [b], Seemay Chou [b], Iván D'Orso [c], Jason Fernandes [c], Gwendolyn Jang [c], Alan D. Frankel [c], Tom Alber [b], Qiang Zhou [b], Nevan J. Krogan [a,*]

[a] Department of Cellular and Molecular Pharmacology, University of California-San Francisco, 1700 4th Street, San Francisco, CA 94158, United States
[b] Department of Biochemistry and Molecular Biology, University of California-Berkeley, 374 Stanley Hall, Berkeley, CA 94720-3220, United States
[c] Department of Biochemistry and Biophysics, University of California-San Francisco, 600 16th Street, San Francisco, CA 94158, United States

ARTICLE INFO

Article history:
Available online 12 August 2010

Keywords:
Virus-host interaction
Mass spectrometry
Protein interaction
HIV
Proteomics

ABSTRACT

To fully understand how pathogens infect their host and hijack key biological processes, systematic mapping of intra-pathogenic and pathogen–host protein–protein interactions (PPIs) is crucial. Due to the relatively small size of viral genomes (usually around 10–100 proteins), generation of comprehensive host–virus PPI maps using different experimental platforms, including affinity tag purification-mass spectrometry (AP-MS) and yeast two-hybrid (Y2H) approaches, can be achieved. Global maps such as these provide unbiased insight into the molecular mechanisms of viral entry, replication and assembly. However, to date, only two-hybrid methodology has been used in a systematic fashion to characterize viral–host protein–protein interactions, although a deluge of data exists in databases that manually curate from the literature individual host–pathogen PPIs. We will summarize this work and also describe an AP-MS platform that can be used to characterize viral-human protein complexes and discuss its application for the HIV genome.

1. Introduction

Protein–protein interaction networks have been generated using AP-MS and Y2H targeting various organisms, ranging from bacteria to humans. Yeast two-hybrid screenings consist of testing all pair wise combinations of proteins, which generates a collection of binary interactions. High-throughput Y2H maps have been generated for Saccharomyces cerevisiae [16,23,41,47], *Caenorhabditis elegans* [28,45], *Drosophila* [18] and humans [10,35,39]. Affinity tag purification-mass spectrometry approaches identify groups of proteins that participate in complexes and these have been used to study the cellular make-up of *Escherichia coli* [2,5], *S. cerevisiae* [17,22,27] and human cells [14]. While AP-MS assays have a higher propensity of detecting stable, stoichiometric complexes, Y2H screens tend to detect transient protein interactions [47]. Therefore the data from both approaches are complementary with respect to revealing physical connections between proteins, complexes and biological processes.

These unbiased approaches have also been used to study PPIs between proteins that are derived from a specific virus. For example, an intraviral Hepatitis C (HCV) Y2H interaction map was built using a limited set of predefined coding segments, which revealed the functional interactions between the proteins in the viral life cycle when a cell culture system is absent [15]. Also, using a Y2H approach, intraviral protein–protein interaction networks have been generated for two herpesviruses, Kaposi sarcoma-associated herpesvirus (KSHV) and Varicella-Zoster virus (VZV) [40]. The resulting PPI networks appear as a single highly connected module whereas cellular networks (e.g. yeast and human) have been observed to be organized in functional modules. Despite a broad range of pathogenicity, herpesviruses share a significant percentage of common conserved genes and the authors attempted to define a core set of interactions conserved among these viruses. Calderwood et al. proteomically interrogated another herpesvirus, Epstein-Barr virus (EBV), by classifying the genes into two evolutionary classes based on conservation and showed enrichment for interactions among proteins in the same evolutionary class [6]. Another example of an intraviral PPI network based yeast-two-hybrid matrix analysis was obtained for SARS coronavirus [43]. SARS-CoV has 14 ORFs, most of whose functions are unknown. Interestingly, one of the accessory proteins turned out to be highly connected, and the authors propose that although not essential for viral replication in cell culture systems, it could enhance the global stability of the SARS proteome network and pathogenicity.

These pair-wise interaction studies have also been extended into studying the interaction landscape between viral proteins

* Corresponding author.
E-mail address: krogan@cmp.ucsf.edu (N.J. Krogan).

1046-2023/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.ymeth.2010.08.007

and host factors. For example, Lotteau and colleagues published a proteome-wide, Y2H-based mapping of interactions among HCV and human proteins. They reported 314 interactions (in addition to 170 literature curated interactions) and discovered that HCV CORE protein was a major perturbator of the insulin, Jak/STAT and TGFß pathways [11]. More recently, another study targeted Vaccinia virus, a large double stranded DNA virus with more than 280 ORFs and a prototype of the Orthopoxvirus, which includes several pathogenic poxviruses such as variola virus, a lethal human-specific pathogen that causes smallpox [49]. The authors reported a comprehensive yeast-two hybrid screening with 109 protein–protein interactions between vaccinia proteins and human proteins and provided functional insight into a number of uncharacterized viral proteins. Finally, Shapira et al. introduced a multilayered approach to uncover pathways in H1N1 infection by combining yeast-two-hybrid analysis and genome-wide expression profiling [38]. They found human factors mediating virus–host interactions, which were further studied via depletion analysis in primary lung cells. These types of unbiased physical and regulatory models of virus–host interactions provide a promising direction for the unveiling of new virus biology and development of new viral drugs [31].

Collectively, the global network properties of human proteins targeted by pathogens, including bacteria and viruses, were recently studied [13]. It was observed that pathogenic proteins preferentially interact with human hub proteins and "bottleneck" factors in human pathways. Although 190 pathogens were analyzed in this study, 98.3% of the interactions were obtained from viruses and 77.9% of them were associated with HIV. Interactions of each of its 18 proteins have been individually studied in numerous labs, mostly using Y2H, but also AP/MS, *in vitro* binding and other methodologies. In an attempt to catalog these data, the National Institute of Allergy and Infectious Diseases Division of AIDS (NIAID) has initiated the development of an HIV-1 Human Protein Interaction database [33]. From HIV relevant publications, 2589 unique HIV-human PPIs among 1448 human proteins were curated (Fig. 1); 32% of these interactions are reported to be direct, physical interactions. Surprisingly, 37% of the human proteins on this list interact with more than one HIV-1 protein. For example, mitogen-activated protein kinase 1 (MAPK1), a signaling protein, has been described to interact with 10 HIV proteins.

Since the HIV-human interactions are mostly literature-curated [8], it is hard to know if the nature of the interactions are physiologically relevant or due to the apparent bias in the literature towards highly studied proteins [48]. The fact that the number of direct interactions reported for each protein vary considerably, ranging from only one for polymerase or reverse transcriptase and up to 219 for the 14 kDa protein Tat, suggests that this database includes false positives for some proteins while for others there still might be host interactors to be discovered. Therefore, although there have been a variety of excellent studies on HIV-1 human interactions, providing invaluable information about host

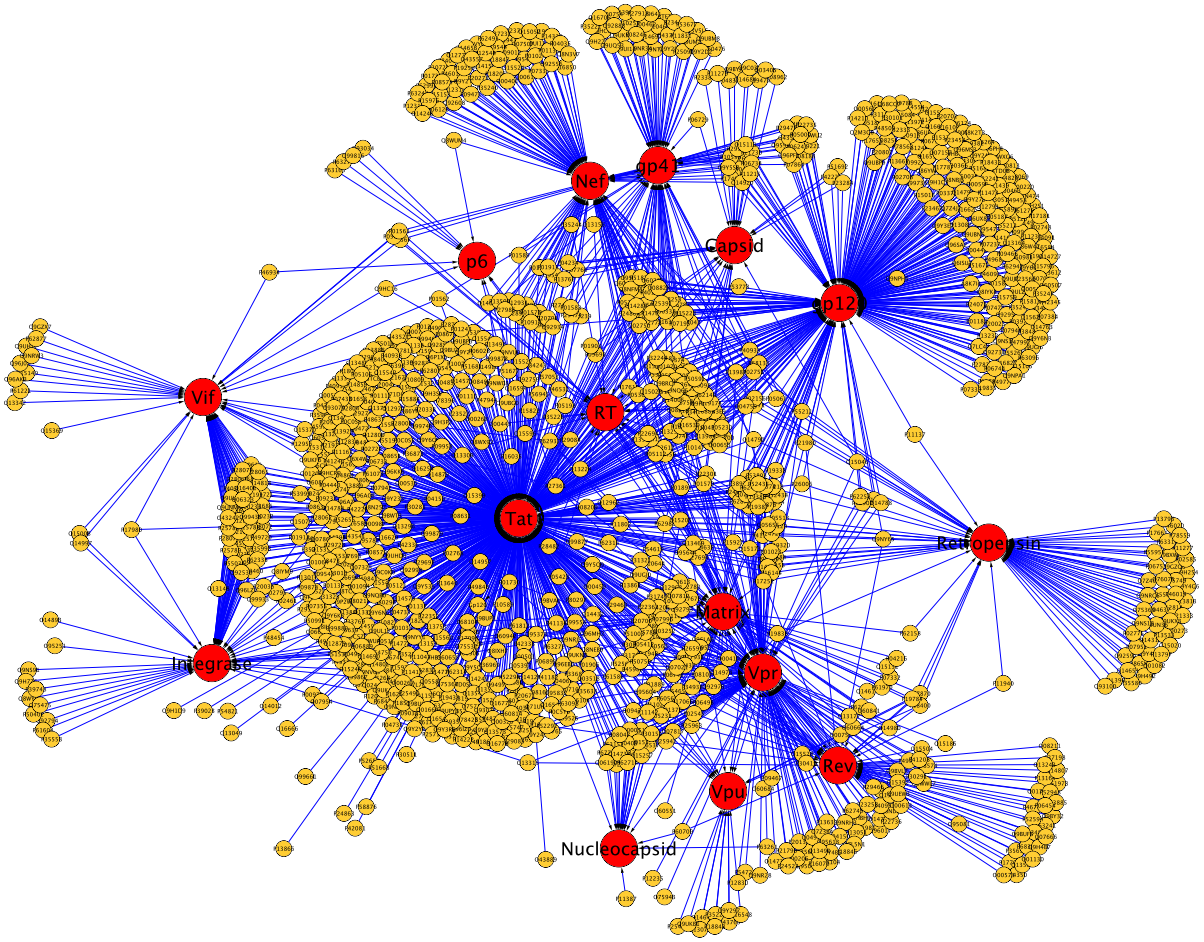

Fig. 1. Literature derived HIV–human protein–protein interaction map. A network displaying HIV-human interactions derived from the National Institute of Allergy and Infectious Diseases Division of AIDS (NIAID) HIV-1 Human Protein Interaction Database. HIV proteins correspond to red nodes whereas yellow nodes represent host factors. In total, 1785 unique HIV-human interactions among 1175 human and 15 HIV proteins are presented. Further work will be required to determine which interactions are direct or functionally relevant.

factors crucial for HIV pathogenicity, a systematic approach of building the HIV-1–host protein–protein interaction network would help to get a clearer picture of the interconnection of the different virus components with the host cell. In this paper, we describe such an approach, based on AP-MS methodology, and describe how it can be used to proteomically interrogate HIV as well as other viruses.

2. Methods

In the following section, we describe: (1) different strategies that can be employed to affinity tag HIV proteins, (2) purification protocols, and (3) ways in which the resulting data can be analyzed and integrated with other types of information.

2.1. Tagging strategies

2.1.1. Affinity tagging in the context of the viral genome

Characterization of HIV–human protein–protein interactions during infection would arguably result in a dataset that would be most physiologically relevant. In order to accomplish this, however, one would have to tag the proteins in the context of the viral genome, infect the appropriate cells with these genetically altered viruses and then purify and identify the complexes. So far Integrase, Vif and Vpr have been successfully tagged within the provirus while maintaining infectivity [32,36,48]. Affinity purification of these proteins from infected cells led to the discovery of the Vif/Cul5 complex and the finding that DDB1 is an additional component of the Vpr/DCAF1 complex. While this strategy is a very elegant way to identify virus–host interactions, it is not applicable to most of the HIV ORFs due to the overlapping organization of the HIV genome. Tagging these factors on either the N- or C-termini would disrupt the functions of the essential, overlapping proteins and would have detrimental effects on the function of the virus. Similarly, placing tags on the individual factors that comprise the polyproteins (Gag, Pol and Env) would result in similar expression problems for downstream proteins and/or adversely affect processing events by Protease, which would result in a disabled or comprised virus.

2.1.2. Transient transfection of tagged proteins

A more amenable AP-MS approach for comprehensively characterizing HIV–human protein–protein interactions is to individually clone each of the factors into an appropriate tagging construct, separately express these factors in a human cell line, and then purify and characterize the resulting complexes. Although simpler, one could argue that the resulting data may be less relevant since some of the viral proteins need other HIV factors for proper function or localization (e.g. MA, IN, Vpr and RT as part of the preintegration complex) and the tagged proteins may be significantly overexpressed when compared to levels during infection.

It is also worth pointing out that the late genes of HIV-1 are expressed from intron-containing mRNAs and depend on the Rev protein for nuclear export and translation. Removing inhibitory sequences and adapting the codon usage for mammals can achieve efficient expression of these ORFs in the absence of Rev. Such codon-optimized versions have been generated of all HIV-1 late genes by different labs [3,7,9,21,30,37,44], and can nowadays be conveniently made by gene synthesis.

2.1.3. Creation of Jurkat cell lines stably expressing tagged proteins

A variation on the approach described in Section 2.1.2 is to use the tagged constructs to create stable cell lines in a more relevant cell line (e.g. T cells). Presumably, this strategy would also provide lower (and therefore more natural) expression levels. Furthermore, many more known HIV-related host proteins are expressed in Jurkat cell lines. For example, APOBEC3G and Tetherin, characterized substrates of the Vif/CUL5 and Vpu/Cul1 ubiquitin ligase complexes, respectively, are more highly expressed in Jurkat cells when compared to HEK293 cells. However, one could carry out these experiments in the presence of interferon to induce the expression of factors involved in innate immunity.

A further step would be to carry out the purifications in the presence of HIV infection. Even though in this scenario, two copies of the viral proteins essential for virus production and infection would be present (i.e. tagged and untagged), this approach would potentially identify intraviral interactions and those that are dependent on other HIV proteins or the viral RNA.

2.1.4. Double pull-down approach

Some HIV proteins are likely to interact with several host protein complexes in order to perform multiple functions during the viral replication cycle. Information about the composition of different HIV–host complexes can be gained using the "split-tag" approach [34]. In this strategy, an affinity-tagged HIV protein is co-expressed with an interaction partner carrying a different tag and a tandem affinity purification (TAP) is performed (Fig. 2). The first step involves purification of the tagged HIV factor, which would enrich for all the different complexes that are associated with this viral protein. Next, purifying a tagged host factor from this viral-enriched population identifies complexes that would only include that particular human protein.

In this way, one could systematically create lines dually expressing each viral-host protein pair that was derived from the

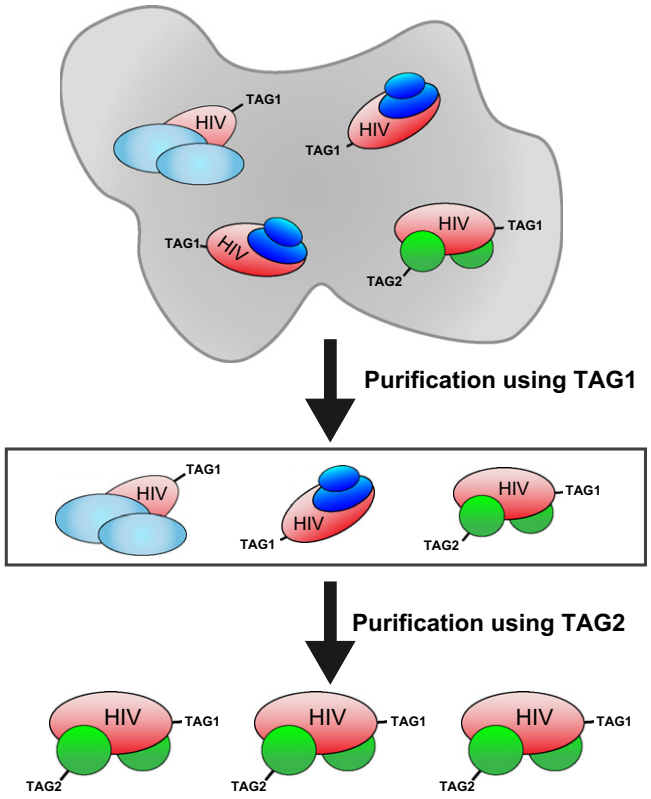

Fig. 2. The double pull-down approach to characterize HIV–human protein complexes. In this strategy, cell lines are expressing two proteins, one viral and one host, each with a different affinity tag. The first purification step enriches for the viral protein, and presumably all the complexes it is associated with whereas the second purification step targets a host protein and enriches for a specific and stoichiometric viral-host protein complex.

single HIV purification experiments and subject the extract to this "double pull-down" strategy. This approach would: (1) verify the relevance of individual interactions derived from single purification experiments, (2) place host factors into their respective complexes via the co-enrichment patterns and (3) identify more physiological, stoichiometric HIV-human protein complexes, which would more likely be used for subsequent functional assays or even structural studies. Of course, an additional, complementary strategy would be to carry out single, reciprocal purifications of the tagged human proteins, alone and also in the presence of the appropriate viral protein, which would help functionally verify the host-pathogen protein–protein interactions and expand the network to include more host proteins and complexes.

2.2. Purification of affinity-tagged HIV proteins

2.2.1. Cloning

The HIV-1 open reading frames encoding precursor proteins (GagPol, Gag, Pol, gp160), subunits (MA, CA, NC, p6, PR, RT, IN, gp120, gp41) and accessory proteins (Vif, Vpr, Vpu, Nef, Tat, Rev) are PCR amplified from either a proviral vector or codon-optimized templates and ligated into the vector pcDNA4/TO (Invitrogen) carrying either a 5' 3xFlag2xStrep (FS) or a 3' 2xStrep3xFlag (SF) tag [19]. In protease containing constructs, the catalytic site needs to be mutated to avoid cytotoxicity of the active enzyme. Expression of full-length GagPol can be achieved by insertion of an extra T nucleotide into the ribosomal frame shift site. A signal peptide sequence added to the gp41 ORF mediates membrane association.

2.2.2. Transient transfections

A convenient cell line for large-scale AP-MS experiments is HEK293 since these cells are easy to culture and transfect. HEK293 cells are maintained in DMEM high glucose, 10% FBS and antibiotics at 37 °C in 5% CO2. For transient transfections, 2.5×106 cells are seeded per 150 cm^2 dish and transfected the following day with 5 μg plasmid using standard calcium phosphate precipitation [24]. Cells are harvested 36–48 h after transfection using detachment with 10 mM EDTA. Expression of antiviral restriction factors like Tetherin can be induced by treatment with 10,000 U/ml IFNα2a (PBL InterferonSource) [42].

2.2.3. Generation of stable Jurkat T cell clones

Since T cells are the natural target cells for HIV, expression and purification of HIV proteins from a T cell line is desirable. Since large-scale T cell transfections are difficult, stably transfected cell lines should be generated. A regulatable expression system should be used since expression of several HIV proteins affect cell viability. Jurkat TRex cells (Invitrogen), for example, stably express the tetracylin repressor protein so that genes from Tet operon containing plasmids like pcDNA4/TO are only expressed upon tetracyclin induction. Jurkat TRex cells are maintained in RPMI, 10% FCS, 10 μg/ml blasticidin, 1% PenStrep. For generation of stable clones, 1×10^6 cells are transfected with 2 ug linearized pcDNA4/TO plasmid by electroporation according to the manufacturer's instructions (Amaxa). Stably transfected cells are selected in 300 μg/ml Zeocin for several weeks and single clones are isolated by limited dilution. Expression of HIV proteins is induced by 1 μg/ml Doxycylin for 12–24 h. Expression levels of proteins with a short half-life (e.g. Vif) can be increased by addition of 500 nM protease inhibitor MG132 (Calbiochem) during induction.

2.2.4. Cell lysis and preclearing

In order to obtain two independent datasets for each protein and cell line, both Flag and Strep affinity purifications can be per-

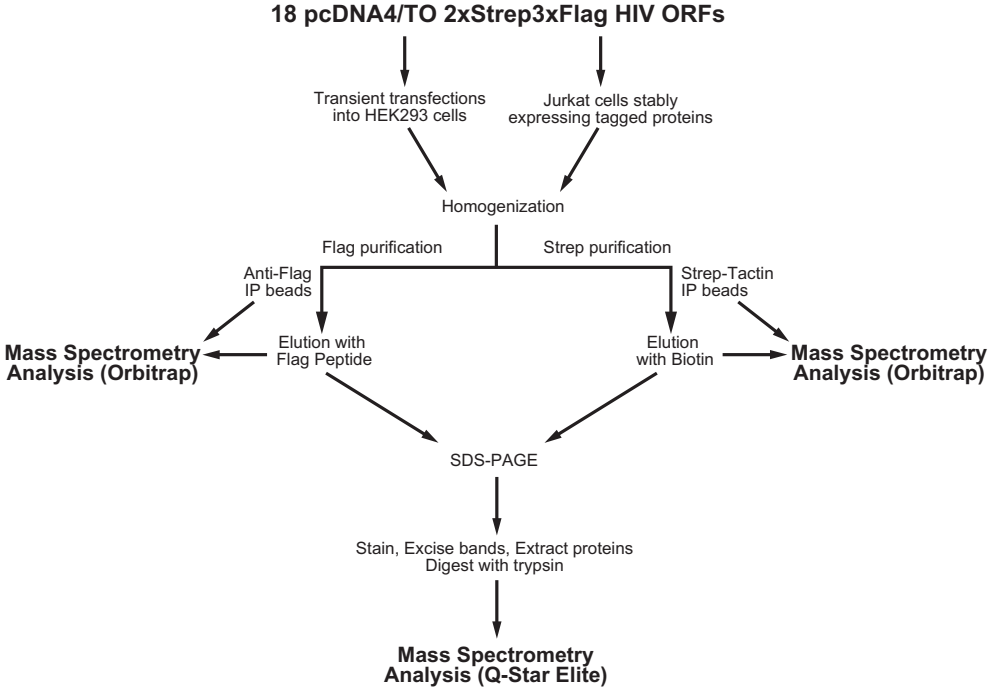

Fig. 3. The purification-mass spectrometry strategy to characterize HIV–human protein complexes. Cloned viral genes are inserted into a construct that fuses a 2XStrep3XFlag dual affinity tag on the C-terminus of each factor. These constructs can be used for transient transfection in HEK293 cells or for generating stably expressing Jurkat cells. After lysis, the extract is subjected to either Anti-Flag or Strep-Tactin IP beads where an aliquot of the beads, as well as the elution, is subjected to trypsin digestion and the material is analyzed using the OrbiTrap mass spectrometer. In both cases, a portion of the eluate is also subjected to SDS–PAGE analysis, where the gel is stained, bands excised, proteins extracted and the protein is digested with trypsin and analyzed using a Q-Star Elite mass spectrometer. The data obtained from both sets of purifications and multiple points during the isolation are then integrated together and subjected to an algorithm to derive quantitative viral-host protein–protein interactions. See text for a more detailed description.

formed (Fig. 3). Two plates of (3×10^7) HEK293 cells or 5×10^8 Jurkat cells (500 ml culture), respectively, are washed with PBS, pelleted and resuspended in 2 ml cold lysis buffer (0.5% NP40, 50 mM Tris–HCl pH 7.4; 150 mM NaCl, 1 mM EDTA, protease and phosphatase inhibitors). The amount of detergent in the lysis buffer is always a compromise between protein solubilization, maintenance of relevant interactions and elimination of unspecific binding. Usually a concentration between 0.2% NP40 and 1% NP40 + 0.25% CHAPS is used. In addition, the cells can be mechanically disrupted by douncing, sonication or homogenization to enhance protein extraction.

Insoluble material is then pelleted for 20 min at 2800*g*. Preclearing of the supernatant with unspecific beads significantly reduces background binding. Depending on the beads used for the subsequent AP, 60–100 μl of either mouse IgG Agarose (Sigma) or Sepharose 4FF (GE Healthcare) is added to the supernatant and incubated for 1 h on an overhead shaker at 4 °C. Mass spectrometry (MS) analysis of these preclearing beads yields a background binding dataset.

2.2.5. Affinity purification

For purification of individual proteins, 30–50 ul IP beads (anti-Flag M2 Affinity Gel, SIGMA or Strep-Tactin Sepharose, IBA) is added to the precleared lysate and the immunprecipitation is performed in batch on an overhead shaker for at least 1 h at 4 °C. Beads are then washed extensively either in columns (Poly-Prep, BioRad) or in batch (2 ml dolphin tubes, BLD Science) with cold 0.1% NP40, 50 mM Tris–HCl pH 7.4, 150 mM NaCl, 1 mM EDTA. In case of purification of proteins with RNA binding motifs (Gag, NC, Tat, Rev), unspecific association of RNA associated proteins like splicing factors, RNA helicases, etc. can be reduced by incubation of the beads with 1500 U RNase A (Fermentas) for 30 min on ice, followed by washing. The last washing step is performed with detergent-free buffer to avoid interference with MS analysis. 10 μl of the beads are directly analyzed by MS using on-bead trypsin digest, while the rest of the beads is eluted with 30–50 μl of either 100 μg/ml 3xFLAG peptide (Elim Biopharmaceuticals) or D-desthiobiotin solution (IBA), respectively (Fig. 3). Five microliters of the eluate is analyzed by SDS–PAGE and silver staining. The rest is analyzed directly by MS as well as after fractionation on a SDS polyacrylamide gradient gel (4–20%, BioRad) and staining with GelCode Blue Safe Protein Stain (Thermo Scientific).

2.2.6. Subcellular fractionations

Subcellular compartments may have to be enriched prior to purification to identify functional HIV–human protein–protein interactions. Nuclear proteins like Integrase, Rev, Tat, and Vpr can be extracted from the nuclear fraction using high salt buffer according to the standard protocol [1]. The remaining insoluble material can further be treated with 1.5 U/μl benzonase nuclease (Merck) to solubilize chromatin-associated proteins like Integrase.

Membrane-associated proteins like Gag, Vpu, Nef and Env can be affinity purified from membrane fractions enriched by flotation in a discontinuous iodixanol gradient. To this end, cells are hypotonically lysed and disrupted by dounce homogenization. The lysate is adjusted to 40% OptiPrep (SIGMA), overlaid with 28% Optiprep and TNE buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 5 mM EDTA) and centrifuged at 165,000*g* for 3 h at 4 °C. The light membranes and associated proteins will float to the top and appear as an opaque band at the OptiPrep-buffer interface, where it can be collected and applied to AP-MS.

2.2.7. Double pull-down experiments

For the double pull-down experiment, $10 \times 150\ cm^2$ plates HEK293 cells are co-transfected with vectors coding for the strep-tagged viral protein and one or more Flag-tagged host proteins. The affinity purifications are performed as described above, with the first step being scaled up accordingly. The eluates after both steps are compared by SDS–PAGE and mass spectrometry.

2.3. Specificity and validation

An inherent problem of AP-MS experiments is the high number of unspecific interactions that can be detected. Usually the data obtained from the purification of the affinity tagged protein of interest is compared to a negative control using untagged protein, untransfected cell lysate, tagged GFP or preimmune serum. While this helps to identify a limited set of unspecific binders, proteins often have distinct background interactions depending on their localization and nature, e.g. nuclear proteins have a different set of background interactors than membrane proteins. In general, the more unrelated proteins with similar characteristics are analyzed, the easier it is to identify specific, and therefore physiologically relevant interactions.

Since AP-MS studies reveal little information about whether the association is direct or indirect, interactions should be confirmed using different methodologies, including *in vitro* binding studies. For a functional validation, the data can be compared to the published studies using genome-wide RNAi screens to identify host factors needed for HIV replication [4,25,46,50]. However, since the HIV accessory genes Vif, Vpu, Vpr, and Nef are often dispensable for replication in cell culture systems [29], specific assays need to be applied in these cases.

2.3.1. Quantitative HIV–human protein–protein interaction scoring system

Once pull-down samples are acquired and analyzed by mass spectrometry, proteomic information concerning host–pathogen interactions can be ascertained. MS identification of a host protein as a putative interactor allows for further investigation into the host–pathogen interaction, employing methods such as yeast two-hybrid, biochemical assays, or viral infectivity assays, to verify and establish the biological relevance of the viral-host interaction. In this regard, proper quantitative analysis of mass spectrometry data derived from affinity purified viral or host proteins is essential in identifying biologically meaningful PPIs in order to minimize time and resources expended on false-positive MS identified interactors. However, one major caveat of reported PPIs obtained from AP-MS experiments is often little information relating to protein abundance or specificity is directly revealed. Also, the interactor abundance may not even be the best indicator of the interaction reliability, especially since some protein abundance strongly depend on interaction affinity as well as its concentration in the cell or in the final experimental sample. If the purified material from a single affinity purification of one tagged protein bait is analyzed, even when contrasted with data from a non-tagged control, it remains incredibly difficult to ascertain specificity and reproducibility with respect to putative interactors. Furthermore, shotgun sequencing approaches to protein identification suffer from poor sampling of IP proteins independent of precision of sample purification replicates. For example, shotgun sequencing of the same sample may only result in 30–40% overlap with respect to proteins identified and may require 5–10 separate runs to obtain 95% coverage [12,20,26].

To identify physiologically relevant PPIs, it helps to have information pertaining to protein abundance in the sample, some metric of bait-interactor specificity, and some metric of interactor reproducibility. As an example concerning specificity, RNA binding proteins (e.g. Tat and Rev) will often be identified with ribosomal protein subunits due to binding to an RNA molecule and not a direct PPI occurring *in situ*. Likewise, any highly abundant cellular proteins (i.e. cytoskeleton proteins) tend to be identified in high

abundance irrespective of the employed bait, indicating a problematic disconnect between protein abundance and interaction specificity. On the other hand, such non-specific interactions may not be very reproducible, which could in principle allow us to recognize them and filter them out.

Each pull-down experiment can be represented as a vector of abundance scores (defined below) for all of the unique interactors found in our approach. When the specific interactor is not pulled down with particular bait, its abundance score is set to zero. The vectors of different affinity purification experiments can then be organized into a two-dimensional matrix, and the vectors of the replicated experiments extending further into the third dimension. In order to quantify abundance of identified proteins within a sample and to allow normalization of this data across different samples such that values between multiple datasets can be accurately compared, one can use the label-free SIn normalization method described previously [20]. The SIn formula takes into account the spectral intensities from each interactor, the total spectral intensities observed in the MS run, and the length of the protein identified, and has been shown to outperform other methods of quantifying abundance such as spectral counts or number of interactors identified [20]. Assuming constant binding capacities for the same baits in different experimental runs, we can define a new SIn score (pSIn) for a particular interactor from a single pull-down experiment as a proportion of the real SIn scores in that particular experiment. The pSIn is now normalized and therefore equivalent across all of the experiments, and is used to fill in the matrix cells. The first of the three proposed metrics, abundance, is then an average of interactor pSIn scores in the third (replicates) dimension in our data matrix. The second metric, reproducibility, which is derived from the values in the third dimension as well, should be a measure of how the interactor amounts (or pSIn scores) are reproducible among replicated experiments. The more reproducible numbers would be hence uniformly distributed, but less reproducible numbers would form peak(s), which suggests using the entropy as a measure of system organization. Before applying a standard entropy equation to the data, one has to normalize the scores in this third dimension so that they add up to one. To avoid biases due to different number of replicates, we also normalize the entropy, dividing it with the maximal entropy possible. The third metric, specificity, is defined similarly to abundance, except that instead of using the proportion of pSIn scores in the vector (or in first dimension), we use the proportions of abundances in the second dimension of the matrix. To avoid the inconvenience of dealing with interactions described by three scores, one can compress the three metrics into a single score using Principal Component Analysis (PCA), a statistical technique most commonly used to linearly transform the original set of data into the set of uncorrelated variables with the goal to reduce the dimensionality of this original data set.

3. Summary/outlook

Viruses, like HIV, are incredibly complicated, resourceful organisms that are involved in many diverse functions during infection. However, their genomes are surprisingly small considering the tasks they must carry out, and therefore they rely very heavily on the cellular machinery in the host cells they infect. Based on this, one might expect that one viral protein would be involved in multiple processes and therefore would hijack several host complexes during infection. Using an approach like AP-MS, therefore, would be a powerful way to identify these relationships, especially when it is conducted in an unbiased and systematic way. Overlaying the PPI network with genetic information derived from global RNAi screens [4,25,46,50] would help identify which of these physical interactions have functional consequences. Finally, comparing host-pathogen protein–protein interaction data using a variety of different viruses will help identify commonalities with respect to infection and should help develop new therapeutic strategies. Since vaccine and drug development targeting HIV proteins has been problematic, identifying novel host targets that aid in infection may represent the next step in combating HIV infection.

Acknowledgments

We thank members of the Krogan lab for helpful comments. This work was funded by NIH. NJK is a Searle Fellow and Keck Young Investigator Fellow.

References

[1] S.M. Abmayr, T. Yao, T. Parmely, J.L. Workman, Curr. Protoc. Mol. Biol. (2006). Chapter 12: Unit 12.1.
[2] M. Arifuzzaman, M. Maeda, A. Itoh, K. Nishikata, C. Takita, R. Saito, T. Ara, K. Nakahigashi, H.C. Huang, A. Hirai, et al., Genome Res. 16 (2006) 686–691.
[3] D.L. Bolton, M.J. Lenardo, J. Virol. 81 (2007) 8878–8890.
[4] A.L. Brass, D.M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R.J. Xavier, J. Lieberman, S.J. Elledge, Science 319 (2008) 921–926.
[5] G. Butland, J.M. Peregrin-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, et al., Nature 433 (2005) 531–537.
[6] M.A. Calderwood, K. Venkatesan, L. Xing, M.R. Chase, A. Vazquez, A.M. Holthaus, A.E. Ewence, N. Li, T. Hirozane-Kishikawa, D.E. Hill, et al., Proc. Natl. Acad. Sci. USA 104 (2007) 7606–7611.
[7] B.K. Chakrabarti, W.P. Kong, B.Y. Wu, Z.Y. Yang, J. Friborg, X. Ling, S.R. King, D.C. Montefiori, G.J. Nabel, J. Virol. 76 (2002) 5357–5368.
[8] A. Chatr-aryamontri, A. Ceol, D. Peluso, A. Nardozza, S. Panni, F. Sacco, M. Tinti, A. Smolyar, L. Castagnoli, M. Vidal, et al., Nucleic Acids Res. 37 (2009) D669–673.
[9] P. Cherepanov, W. Pluymers, A. Claeys, P. Proost, E. De Clercq, Z. Debyser, FASEB J. 14 (2000) 1389–1399.
[10] F. Colland, X. Jacq, V. Trouplin, C. Mougin, C. Groizeleau, A. Hamburger, A. Meil, J. Wojcik, P. Legrain, J.M. Gauthier, Genome Res. 14 (2004) 1324–1332.
[11] B. de Chassey, V. Navratil, L. Tafforeau, M.S. Hiet, A. Aublin-Gex, S. Agaugue, G. Meiffren, F. Pradezynski, B.F. Faria, T. Chantier, et al., Mol. Syst. Biol. 4 (2008) 230.
[12] E. Durr, J. Yu, K.M. Krasinska, L.A. Carver, J.R. Yates, J.E. Testa, P. Oh, J.E. Schnitzer, Nat. Biotechnol. 22 (2004) 985–992.
[13] M.D. Dyer, T.M. Murali, B.W. Sobral, PLoS Pathog. 4 (2008) e32.
[14] R.M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M.D. Robinson, L. O'Connor, M. Li, et al., Mol. Syst. Biol. 3 (2007) 89–90.
[15] M. Flajolet, G. Rotondo, L. Daviet, F. Bergametti, G. Inchauspe, P. Tiollais, C. Transy, P. Legrain, Gene 242 (2000) 369–379.
[16] M. Fromont-Racine, J.C. Rain, P. Legrain, Nat. Genet. 16 (1997) 277–282.
[17] A.C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B. Dumpelfeld, et al., Nature 440 (2006) 631–636.
[18] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols, et al., Science 302 (2003) 1727–1736.
[19] C.J. Gloeckner, K. Boldt, A. Schumacher, R. Roepman, M. Ueffing, Proteomics 7 (2007) 4228–4234.
[20] N.M. Griffin, J. Yu, F. Long, P. Oh, S. Shore, Y. Li, J.A. Koziol, J.E. Schnitzer, Nat. Biotechnol. 28 (2010) 83–89.
[21] L. Hermida-Matsumoto, M.D. Resh, J. Virol. 74 (2000) 8670–8679.
[22] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, et al., Nature 415 (2002) 180–183.
[23] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, Proc. Natl. Acad. Sci. USA 98 (2001) 4569–4574.
[24] R.E. Kingston, C.A. Chen, H. Okayama, Curr. Protoc. Cell Biol. (2003). Chapter 20: Unit 20.3.
[25] R. Konig, Y. Zhou, D. Elleder, T.L. Diamond, G.M. Bonamy, J.T. Irelan, C.Y. Chiang, B.P. Tu, P.D. De Jesus, C.E. Lilley, et al., Cell 135 (2008) 49–60.
[26] J.A. Koziol, A.C. Feng, J.E. Schnitzer, Anal. Chem. 78 (2006) 3203–3207.
[27] N.J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A.P. Tikuisis, et al., Nature 440 (2006) 637–643.
[28] S. Li, C.M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P.O. Vidalain, J.D. Han, A. Chesneau, T. Hao, et al., Science 303 (2004) 540–543.
[29] M.H. Malim, M. Emerman, Cell Host Microbe 3 (2008) 388–398.
[30] K.L. Nguyen, M. llano, H. Akari, E. Miyagi, E.M. Poeschla, K. Strebel, S. Bour, Virology 319 (2004) 163–175.
[31] X. Peng, E.Y. Chan, Y. Li, D.L. Diamond, M.J. Korth, M.G. Katze, Curr. Opin. Microbiol. 12 (2009) 432–438.
[32] C. Petit, O. Schwartz, F. Mammano, J. Virol. 73 (1999) 5079–5088.
[33] R.G. Ptak, W. Fu, B.E. Sanders-Beer, J.E. Dickerson, J.W. Pinney, D.L. Robertson, M.N. Rozanov, K.S. Katz, D.R. Maglott, K.D. Pruitt, et al., AIDS Res. Hum. Retroviruses 24 (2008) 1497–1502.
[34] O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, B. Seraphin, Methods 24 (2001) 218–229.

[35] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al., Nature 437 (2005) 1173–1178.
[36] B. Schrofelbauer, Y. Hakata, N.R. Landau, Proc. Natl. Acad. Sci. USA 104 (2007) 4130–4135.
[37] S. Schwartz, M. Campbell, G. Nasioulas, J. Harrison, B.K. Felber, G.N. Pavlakis, J. Virol. 66 (1992) 7176–7182.
[38] S.D. Shapira, I. Gat-Viks, B.O. Shum, A. Dricot, M.M. de Grace, L. Wu, P.B. Gupta, T. Hao, S.J. Silver, D.E. Root, et al., Cell 139 (2009) 1255–1267.
[39] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, et al., Cell 122 (2005) 957–968.
[40] P. Uetz, Y.A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S.V. Rajagopala, M. Roupelieva, D. Rose, E. Fossum, et al., Science 311 (2006) 239–242.
[41] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, et al., Nature 403 (2000) 623–627.
[42] N. Van Damme, D. Goff, C. Katsura, R.L. Jorgenson, R. Mitchell, M.C. Johnson, E.B. Stephens, J. Guatelli, Cell Host Microbe 3 (2008) 245–252.
[43] A. von Brunn, C. Teepe, J.C. Simpson, R. Pepperkok, C.C. Friedel, R. Zimmer, R. Roberts, R. Baric, J. Haas, PLoS One 2 (2007) e459.
[44] R. Wagner, M. Graf, K. Bieler, H. Wolf, T. Grunwald, P. Foley, K. Uberla, Hum. Gene Ther. 11 (2000) 2403–2413.
[45] A.J. Walhout, R. Sordella, X. Lu, J.L. Hartley, G.F. Temple, M.A. Brasch, N. Thierry-Mieg, M. Vidal, Science 287 (2000) 116–122.
[46] M.L. Yeung, L. Houzet, V.S. Yedavalli, K.T. Jeang, J. Biol. Chem. 284 (2009) 19463–19473.
[47] H. Yu, P. Braun, M.A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, et al., Science 322 (2008) 104–110.
[48] X. Yu, Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, X.F. Yu, Science 302 (2003) 1056–1060.
[49] L. Zhang, N.Y. Villa, M.M. Rahman, S. Smallwood, D. Shattuck, C. Neff, M. Dufford, J.S. Lanchbury, J. Labaer, G. McFadden, J. Proteome Res. 8 (2009) 4311–4318.
[50] H. Zhou, M. Xu, Q. Huang, A.T. Gates, X.D. Zhang, J.C. Castle, E. Stec, M. Ferrer, B. Strulovici, D.J. Hazuda, et al., Cell Host Microbe 4 (2008) 495–504.