Bioorganic & Medicinal Chemistry Letters 25 (2015) 3079–3086

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

BMCL Digest

Hide and seek: Identification and confirmation of small molecule protein targets

Andrei Ursu, Herbert Waldmann *

[a] Department of Chemical Biology, Max-Planck-Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany
[b] Chemical Biology, Faculty of Chemistry and Chemical Biology, TU Dortmund, Otto-Hahn-Strasse 6, 44221 Dortmund, Germany

ARTICLE INFO

Article history:
Received 27 April 2015
Revised 1 June 2015
Accepted 4 June 2015
Available online 12 June 2015

Keywords:
Target identification
Target confirmation
Affinity chromatography
Linker
Chemical probes

ABSTRACT

Target identification and confirmation for small molecules is often the rate limiting step in drug discovery. A robust method to identify proteins addressed by small molecules is affinity chromatography using chemical probes. These usually consist of the compound of interest equipped with a linker molecule and a proper tag. Recently, methods emerged that allow the identification of protein targets without prior functionalization of the small molecule of interest. The digest offers an update on the newest developments in the area of target identification with special focus on confirmation techniques.

Phenotypic screens of compound libraries provide novel opportunities to explore biological pathways that can be targeted with small molecules. Technological advances in areas such as cell and molecular biology, imaging, software development and miniaturization have enabled unprecedented assaying capabilities of large compound collections in a fully automated, multiplexed or paralleled fashion and with higher content than ever before. Although screening campaigns generate large datasets, these typically do not directly provide information about the biomacromolecules responsible for the observed phenotype. Therefore, the identification of small molecules targets is an essential step to illuminate their mode of action and explore novel biological mechanisms. Although drugs can be approved without a clear known target or mode of action, the full characterization of the protein binding profile of a small molecule is an important prerequisite for a complete picture of the biology behind it.

Target identification and confirmation for small molecules is a stepwise and complex process, time consuming and in many cases represents the rate-limiting step in drug discovery campaigns. A widely used method to find small molecule and target protein pairs relies on affinity chromatography employing chemical probes. Thus, the compound of interest typically is attached to a solid support (beads/resin) by means of a linker moiety equipped with a proper tag (biotin, desthiobiotin) or functional group (amine, alkyne, azide, alcohol etc). The main purpose of the linker appendage is to present the small molecule at some distance on the surface of the solid support. It usually consists of hydrophilic oligoethyleneglycol units of different length which generally lead to less non-specific binding of background contaminants than observed for hydrophobic linkers.[1] Novel linker designs employ L-proline units that fold into a rod-like stable helix allowing a better projection of the small molecule away from the surface of the solid support.[2] However, the polyproline rod has been employed in only a few cases and further applications are required to judge its performance relative to the established oligoethylene glycol linker. Thus, the chemical probe acts as bait to enrich binding partners when exposed to cellular lysates. After washing and elution procedures, the protein targets are digested and the resulting peptides are identified by means of liquid-chromatography mass spectrometry (LC–MS). On the one hand, this procedure is robust in respect to highly abundant protein targets with high binding affinity. On the other hand, poorly expressed targets or targets with low affinity for the probe may either be masked by contaminants rendering it difficult to distinguish between true and unspecific binders or will be removed from the bead surface due to stringent washing conditions. To overcome some of these drawbacks, chemical probes were additionally equipped with photoreactive groups (azide, diazirine, benzophenone) that allow the covalent capture of the protein of interest upon irradiation.[3] However, photolysis yields may be poor and unspecific labeling may be encountered.

* Corresponding author. Tel.: +49 (231) 133 2400; fax: +49 (231) 133 2499.
E-mail address: herbert.waldmann@mpi-dortmund.mpg.de (H. Waldmann).

http://dx.doi.org/10.1016/j.bmcl.2015.06.023
0960-894X/

A robust method to discriminate between specific and non-specific binding events relies on quantitative proteomics by means of stable isotope labeling by amino acids in cell culture (SILAC).[4] The principle of SILAC rests on the incorporation of light and heavy amino acids within the proteome of a living cell. Compared to the light amino acids, which are based on the usual isotopes of C (^{12}C), N (^{14}N) and H (^{1}H), the heavy lysate contains the corresponding stable heavy isotopes such as ^{13}C, ^{15}N or ^{2}H. Upon multiple cycles of cell division, cells will integrate the heavy amino acids into the newly synthesized proteins. After complete incorporation, the unlabeled (light) and the labeled (heavy) lysates are treated with the control compound and a compound of interest attached to beads. Following incubation and washing procedures, the beads are mixed equally, eluted and subsequent quantification via LC–MS yields a light/heavy ratio (SILAC ratio). For true targets ratios up to 10 and higher may be observed, while unspecific binders usually display low or equal isotope ratios. This method allows a much more precise delineation of specific versus unspecific binding events compared to traditional mass spectrometry procedures.[5] In addition, a recent improvement of the procedure termed spike-in SILAC,[6] whereby the heavy labeled sample is used as an internal/spike-in standard, allows a better quantification of proteomes across whole tissues or organisms.

One of the main drawbacks of linker employment is the requirement of a comprehensive structure activity relationship study with respect to the observed phenotype. This analysis reveals suitable positions for linker attachment at retained bioactivity. To avoid functionalization of the compound, linker-free (or label-free) methods were developed allowing the direct use of a small molecule identified from a phenotypic screen in target identification campaigns. For example, one such method evaluates the thermodynamic stabilization of the target proteins upon ligand binding by means of the melting temperature (T_m) shift during heat denaturation. The same property was exploited in the SILAC-pulse proteolysis procedure, whereby the thermodynamic stabilization of the protein structure upon interaction with a ligand circumvents its digestion by thermolysin during chemical denaturation with urea (see Linker-free target identification methods for a detailed description).

All target identification procedures described above yield a list of proteins, regardless of the presence of the linker or not. At this point, various biochemical, biophysical and cellular experiments are required to confirm the small molecule–target interaction and the contribution of the target to the observed phenotype. This process is often time consuming because it necessitates the examination of individual proteins from multiple perspectives. Thus, straightforward confirmation methods able to easily identify the target responsible for the phenotype of interest are in high demand. Advances in next-generation DNA sequencing and gene editing tools such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) generate robust procedures for confirmation of small molecule binding proteins. On one hand, massively paralleled DNA-sequencing allows the identification of resistance mutations acquired by the target protein upon compound treatment over several generations using cytotoxicity as readout. On the other hand, the recently described CRISPR/Cas9 gene editing method grants the precise introduction of resistance point mutation(s) within the wild type allele(s). The CRISPR-engineered clones should be resistant upon compound treatment. However, DNA sequencing might not be able to detect mutations within protein targets that are prone to epigenetic silencing upon small molecule action.

The digest addresses target identification techniques from the linker perspective and thereby describes the developments of linker-free and linker-based methods within recent years. In addition, newly reported target confirmation approaches are discussed. Although alternative techniques are currently available,[7–10] these are beyond the scope of the current digest. For an overview the reader is referred to more comprehensive reviews.[11–13]

Linker-based target identification techniques: Table 1 shows examples of chemical probes representing various concepts with the linker molecule ranging from oligoethyleneglycol units of different length to more recent designs such as single strand DNA. Choosing the linker molecule with an appropriate tag mandates the spectrum of experimental procedures that can be exploited in the target identification campaign.

SCH51344 (Entry 1). SCH51344[14] is a small molecule inhibitor of anchorage-independent growth of various tumor cell lines and hinders Ras transformation acting via a novel MAPK independent mechanism.[15] To reveal potential binding partners of this compound, Huber et al. adopted a chemical proteomics procedure which employs SCH51344 attached to sepharose beads via a short linker (Table 1, entry 1) and competition with the linker-free molecule.[16] For affinity-based enrichment the immobilized derivative P1 was incubated with cell lysates followed by washing, elution and digestion procedures and subsequent mass spectrometric analysis of digested peptides to generate a first list of potential target proteins. Comparison with a control experiment, in which the immobilized compound is supplemented with free ligand as competitor, facilitated the identification of MTH1 (7,8-dihydro-8-oxoguanine triphosphatase, also known as NUDT1) as primary target of SCH51344. The target was further confirmed in vitro in experiments such as isothermal titration calorimetry (ITC) and MTH1 catalytic assay. Compound binding to MTH1 was also proven in cellular lysates by means of immunoblotting using SCH51344 attached to beads and increasing concentrations of the free ligand. Moreover, transfection with short interfering RNA against MTH1 hampered colony formation, whereas stable knockdown by means of short hairpin RNAs successfully recapitulated the effect of the inhibitor. Additionally, when MTH1 was overexpressed, reduced sensitivity towards treatment with the compound was observed. The authors also found that the dual c-MET/ALK inhibitor crizotinib,[17] currently in various clinical trials,[18] targets MTH1 as well. By comparing the chemical proteomic profiles of both (R)- and (S)-crizotinib the authors could confirm that only (S)-crizotinib targets MTH1 in cellular lysates, highlighting an enantiomer specific binding event. Subsequent enzymatic studies showed that (S)-crizotinib is a nanomolar inhibitor of MTH1 while the (R)-enantiomer is 16-fold less potent with a half maximal inhibitory concentration of the compound in the micromolar range. Comparing the chemical proteomic profiles of active (S)-crizotinib with SCH51344 yielded MTH1 as the only shared target among the two structurally unrelated small molecules. The study highlights the importance of comparing protein binding profiles of unrelated small molecules to identify common mechanisms of action.

E-3810 (Lucitanib) (Entry 2). SILAC-based quantitative chemical proteomics was successfully employed in an elegant identification of E-3810 (Lucitanib) targets.[19] E-3810 is an anti-angiogenic multikinase inhibitor with anti-tumor activity[20] currently in phase II clinical trials.[21] Although the compound inhibits VEGFR and FGFR kinases in the nanomolar range, it might hit other targets which were not previously revealed by in vitro profiling possibly due to the limited number of kinases tested. Colzani et al. generated chemical probe P2 (Table 1, entry 2) and determined its efficacy in isolating FGFR2 from cellular lysates. Subsequently, a serial competition SILAC experiment was performed,[22] whereby either the light or heavy lysate was incubated with the compound immobilized on agarose resin and in which E-3810 was added at three different concentrations. Thus, a total of four SILAC experiments were concomitantly performed and the determined SILAC ratios allowed to distinguish specific over unspecific binders and background contaminants. Furthermore, the delineation of strong