Journal of Structural Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Thalidomide mimics uridine binding to an aromatic cage in cereblon

Marcus D. Hartmann [a], Iuliia Boichenko [a], Murray Coles [a], Fabio Zanini [b], Andrei N. Lupas [a], Birte Hernandez Alvarez [a,*]

[a] Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
[b] Evolutionary Dynamics and Biophysics Group, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 21 October 2014
Accepted 24 October 2014
Available online xxxx

Keywords:
Immunomodulatory drug
IMiD
Revlimid
Pomalyst

ABSTRACT

Thalidomide and its derivatives lenalidomide and pomalidomide are important anticancer agents but can cause severe birth defects via an interaction with the protein cereblon. The ligand-binding domain of cereblon is found, with a high degree of conservation, in both bacteria and eukaryotes. Using a bacterial model system, we reveal the structural determinants of cereblon substrate recognition, based on a series of high-resolution crystal structures. For the first time, we identify a cellular ligand that is universally present: we show that thalidomide and its derivatives mimic and compete for the binding of uridine, and validate these findings *in vivo*. The nature of the binding pocket, an aromatic cage of three tryptophan residues, further suggests a role in the recognition of cationic ligands. Our results allow for general evaluation of pharmaceuticals for potential cereblon-dependent teratogenicity.

1. Introduction

Thalidomide was originally introduced for its sedative and anti-emetic properties in the 1950s, but banned from the market in the early 1960s due to teratogenic effects that had led to severe developmental defects in about 10,000 newborns. In the decades after its withdrawal, it was also discovered to possess anti-inflammatory, immunomodulatory and anti-angiogenic properties, thus promising valuable treatment for a broad range of clinical conditions (Bartlett et al., 2004; Franks et al., 2004). Especially, its success as an anti-cancer agent paved the road for a renaissance and promoted the development of a new class of immunomodulatory drugs (IMiDs) based on thalidomide as a lead compound. However, many important and promising derivatives like lenalidomide (CC-5013, Revlimid) and pomalidomide (CC-4047, Pomalyst) potentially inherited thalidomide's teratogenic properties, severely limiting their use.

Thalidomide consists of a phthaloyl ring and a glutarimide ring with a chiral carbon; it racemizes *in vivo* and only the (S)-enantiomer is thought to be teratogenic (Bartlett et al., 2004; Franks et al., 2004). Lenalidomide and pomalidomide have the same architecture with modified phthaloyl moieties (Fig. 1). In 2010, Handa and co-workers identified the protein cereblon as a primary target of thalidomide (Ito et al., 2010). They showed that cereblon associates with damaged DNA binding protein 1 (DDB1), a core component of the DDB1/cullin4 (CUL4) E3 ubiquitin ligase complex which is known as a key player in the nucleotide excision repair pathway. These E3 ligase complexes employ a large number of different DDB1-CUL4-associated factors (DCAFs) with different substrate specificity as substrate receptors (Iovine et al., 2011). Cereblon constitutes a novel DCAF for this complex, altering its ubiquitin ligase activity upon thalidomide binding, which may in turn cause its teratogenic effects. The region critical for DDB1 binding was found to reside in the middle part of the cereblon protein, whereas the binding site for thalidomide was mapped to the C-terminal 104 amino acids: the two point mutants of human cereblon hCrbn^{Y384A} and hCrbn^{W386A} and especially the double mutant hCrbn$^{YW/AA}$ had significantly lowered thalidomide-binding activity. Also lenalidomide and pomalidomide were shown to bind to cereblon, competing for the same binding site (Lopez-Girona et al., 2012). Moreover, it was found that the degradation of the transcription factors Ikaros and Aiolos by the Cereblon/DDB1/CUL4 E3 ligase is stimulated by these IMiDs in multiple myeloma cells (Gandhi et al., 2014; Kronke et al., 2014). These downstream factors can however only represent a subset of the natural targets, as they do not occur outside the animal kingdom, whereas cereblon occurs throughout eukaryotes, except fungi.

By sequence analysis (Lupas et al., 2014), human cereblon is a multi-domain protein: an N-terminal intrinsically unstructured region is followed by a LON domain and a C-terminal domain that contains the thalidomide-binding site defined by Ito et al. (2010).

* Corresponding author at: Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstr. 35, D-72076 Tübingen, Germany. Fax: +49 7071 601 349.
E-mail address: birte.hernandez@tuebingen.mpg.de (B. Hernandez Alvarez).

http://dx.doi.org/10.1016/j.jsb.2014.10.010
1047-8477/

The C-terminal domain, which we named CULT (cereblon domain of unknown activity, binding cellular ligands and thalidomide), is also found as the sole domain in two other protein families, one comprising secreted proteins of eukaryotes and the other cytosolic proteins of bacteria. CULT domains contain three strictly conserved tryptophan residues, one of which (W386 in hCrbn) was identified as critical for thalidomide binding (Ito et al., 2010).

Searches for distant homologs show that the CULT domain is related to proteins sharing a common fold formed by two four-stranded, antiparallel β-sheets that are oriented at approximately a right angle, and pinned together at their tip by a zinc ion (Lupas et al., 2014). We have named this fold the β-tent for the prominent arrangement of its β-sheets. Proteins of this fold show considerable functional and structural divergence, including an absence of most sequence motifs characteristic for CULT domains; however, the location of the substrate binding site appears to be conserved among all members of the fold (Lupas et al., 2014).

Here we determined crystal structures of a bacterial CULT domain from *Magnetospirillum gryphiswaldense* MSR-1, bound to thalidomide and other ligands. With a sequence identity of 35%, it is as similar to the human thalidomide binding domain as domains from other eukaryotic clades. Substantiated by an NMR-based assay and *in vivo* data, the structures provide the molecular basis for thalidomide teratogenicity and, for the first time, reveal potential natural ligands that are universally present in all organisms with cereblon proteins.

2. Experimental procedures

2.1. Cloning

The gene encoding MsCI4 was synthesized codon-optimized for expression in *Escherichia coli*. It was cloned in pETHis_1a, a modified pETM vector (sequence available at http://www.embl.de/pepcore/pepcore_services/strains_vectors/vectors/bacterial_expression_vectors/index.html) for overexpression of the protein in *E. coli* with a N-terminal hexa-histidine tag followed by a cleavage site of TEV (*Tobacco Etch Virus*) protease.

Mutants MsCI4$^{YW/AA}$ and MsCI4^{Y101F} were constructed using QuikChange® Site-Directed Mutagenesis Kit (Stratagene) with wild type MsCI4 in pETHis_1a as a template and mutagenic primers. The correctness of all clones was verified by DNA sequencing.

2.2. Protein expression and purification

All proteins were expressed in *E. coli* C41 (DE3) cells. After induction of protein expression in the logarithmic phase at A_{600} = 0.6, the cultures were shaken for 4 h at 37 °C. Cells were pelleted, resuspended in 20 mM Tris, pH 7.5, 100 mM NaCl, 5 mM 2-Mercaptoethanol, 4 mM $MgCl_2$, DNase I and Protease Inhibitor Cocktail (Roche Applied Science), and lysed using a French pressure cell. After centrifugation of the extract, the supernatant was applied on a NiNTA agarose column equilibrated in 20 mM Tris, pH 7.9, 300 mM NaCl, 5 mM 2-Mercaptoethanol. Histidine tagged proteins were eluted with a gradient of 0–0.5 M imidazole. MsCI4 wildtype or mutant protein containing fractions were pooled and dialyzed against 20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM 2-Mercaptoethanol. The histidine tag was cleaved overnight at 4 °C by TEV protease. The protein mixture was loaded on a NiNTA column to which the histidine tagged TEV protease and the cleaved linker bound. Cleaved MsCI4 proteins were found to be in the flow through. They were pooled and concentrated to 14 mg/ml in 20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM 2-Mercaptoethanol.

2.3. NMR spectroscopy

Comparison of spectra of MsCI4 alone and in presence of thalidomide revealed significant chemical shift changes for many resonances, including several prominent, upfield-shifted methyl groups. One such methyl group shifts from −0.31 to −0.89 ppm on binding of thalidomide, providing a trivial assay for ligands employing a thalidomide-like binding mode. Typically, 1D proton spectra were acquired on 50 μM protein samples both alone and in the presence of 10–500 μM ligand. Uracil, uridine and deoxyuridine induce these characteristic chemical shift changes at the lowest concentrations tested. Other pyrimidine nucleobases and nucleosides tested (cytosine, cytidine, deoxycytidine, thymidine and deoxythymidine) did not, with concentrations of at least 500 μM. For these compounds ligand-detected experiments (STD (Meyer and Peters, 2003) and water-LOGSY (Dalvit et al., 2000)) were also used to probe for binding where characteristic chemical shift changes were not detected. Here ligand protein concentration ratios of 10:1 were employed with ligand concentrations of 500 μM. A negative result in these experiments used to place a conservative lower limit on binding affinity at twice this concentration, i.e. ~1 mM. These assays were repeated for the MsCI4^{Y101F} mutant (thalidomide, uracil, uridine and deoxyuridine) and for the MsCI4$^{YW/AA}$ double mutant (thalidomide and uridine).

2.4. Crystallography

Crystallization trials were performed at 294 K in 96-well sitting drop plates with 50 μl of reservoir solution and drops containing 400 nl of protein solution in addition to 400 nl of reservoir solution. The protein solution in the individual co-crystallization trials contained the additives listed in Table 1, in addition to 17 mg/ml of protein in 20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM 2-Mercaptoethanol. Most of the different co-crystallization trials yielded crystals in multiple similar conditions. The conditions for the crystals used for structure determination are listed in Table 1. All crystals were loop mounted and flash-cooled in liquid nitrogen. Where necessary, crystals were transferred into a separate drop of cryo-solution as indicated in Table 1 prior to flash-cooling. All data were collected at beamline X10SA (PXII) at the SLS (Paul Scherrer Institute, Villigen, Switzerland) at 100 K using a PILATUS 6 M detector (DECTRIS) at a wavelength of 1 Å. Diffraction images were processed and scaled using the *XDS* program suite (Kabsch, 1993). The first structure, MsCI4·thalidomide, was solved exploiting the anomalous signal of the structural zinc ions. Three zinc sites, belonging to the three monomers in the asymmetric unit, were identified with SHELXD (Sheldrick, 2008). Density modification with SHELXE yielded an electron density map of excellent quality, which was subsequently traced with ARP/WARP (Perrakis et al., 1999). The structures of MsCI4·pomalidomide, MsCI4·lenalidomide, MsCI4·deoxyuridine and MsCI4^{Y101F}·thalidomide were subsequently solved based on the MsCI4·thalidomide coordinates. All structures were completed by cyclic manual modeling with Coot (Emsley and Cowtan, 2004) and refinement with REFMAC5 (Murshudov et al., 1999). Analysis with Procheck (Laskowski et al., 1993) showed excellent geometries for all structures. Data collection and refinement statistics are summarized in Table 2. All molecular depictions were prepared using MolScript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997). All coordinates and structure factors were deposited in the Protein Data Bank (PDB) under the accession codes 4V2Y (MsCI4·thalidomide), 4V2Z (MsCI4·pomalidomide), 4V30 (MsCI4·lenalidomide), 4V31 (MsCI4·deoxyuridine) and 4V32 (MsCI4^{Y101F}·thalidomide).

Please cite this article in press as: Hartmann, M.D., et al. Thalidomide mimics uridine binding to an aromatic cage in cereblon. J. Struct. Biol. (2014), http://dx.doi.org/10.1016/j.jsb.2014.10.010