Journal of Structural Biology 188 (2014) 225–232

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 4 November 2014

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).

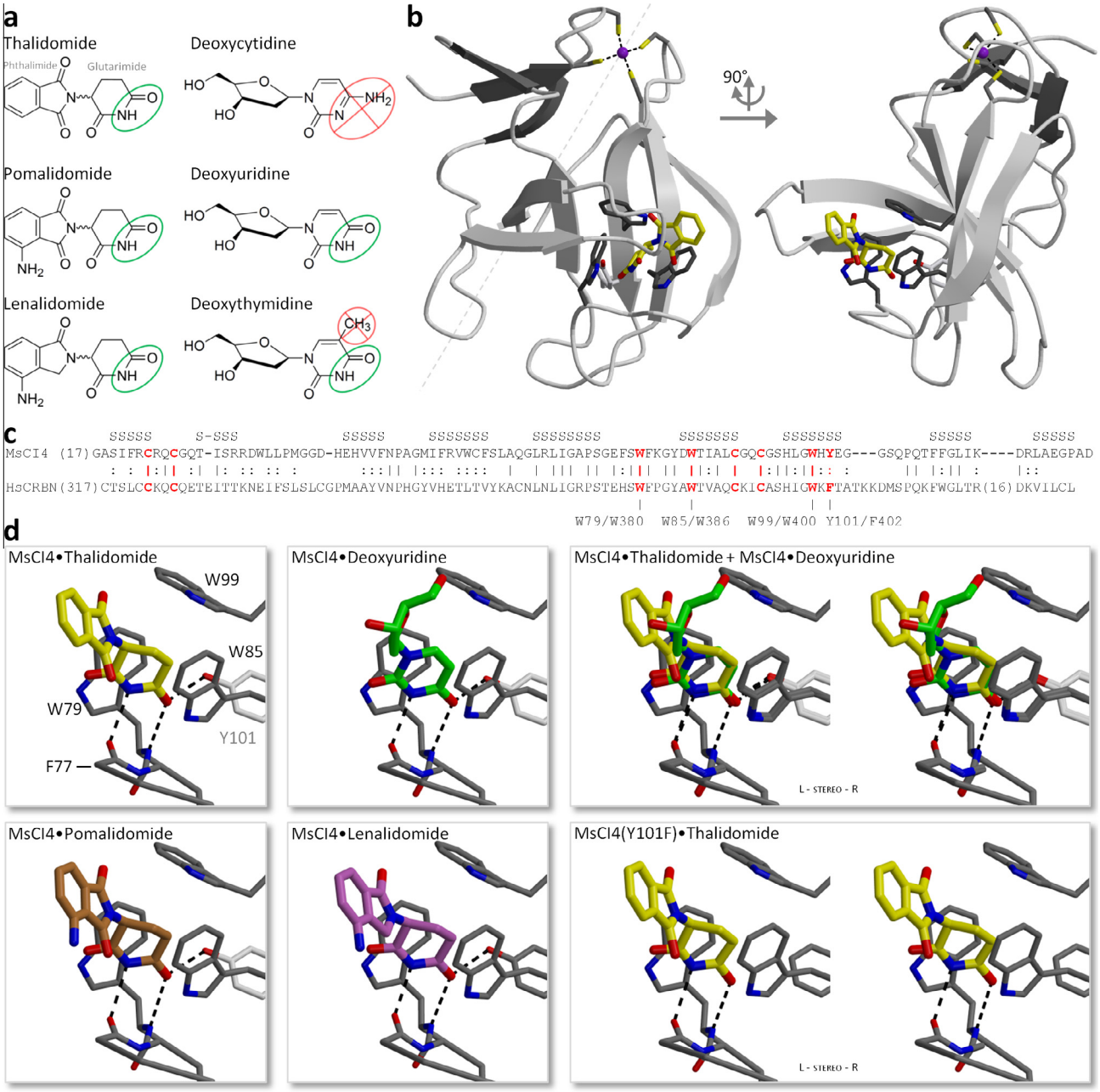

Fig.1. How cereblon recognizes its ligands. (a) The structure of thalidomide, its derivatives, and related pyrimidine nucleosides. Encircled in green are the functional groups that are recognized by cereblon; encircled and crossed out in red are groups that are incompatible with the binding pocket. Pomalidomide and lenalidomide have modified phthaloyl rings but the same glutarimide ring as thalidomide and are therefore bound in the same way. The uracil moiety of uridine is bound in the same way as the glutarimide ring of thalidomide, whereas the thymine moiety of thymidine is too bulky and the cytosine moiety of cytidine cannot form the necessary hydrogen bonds. (b) Cartoon representation of one monomer of MsCI4 in complex with thalidomide in two orientations. Shown in detail are the three tryptophan residues forming the boundaries of the binding pocket, Y101 (white), and four conserved cysteines that coordinate a structural zinc ion. In the left orientation, an approximate twofold symmetry axis is apparent, which relates the beta sheet in dark gray color to the beta sheet in light gray color; this axis passes through the zinc binding site. (c) Sequence alignment of MsCI4 with the thalidomide-binding domain of human Cereblon, together with the secondary structure of the crystal structure. The residues in the binding site are highlighted in bold red, the cysteines of the zinc binding site in red. Identical residues are marked with "|", similar residues with ":". The sequence identity is 35% over 104 aligned positions. (d) Detailed view of the binding pocket of MsCI4 in complex with different ligands, including a superposition of the pocket in complex with thalidomide and deoxyuridine, and a detailed view of the MsCI4^{Y101F} mutant pocket in complex with thalidomide. Highlighted are the hydrogen bonds formed between the ligands and the protein.

2.5. In vivo assay

A zebrafish pectoral fin deformity assay was performed as described by Ito et al. (2010) with the following modifications: (1) The zebrafish cereblon (zcrbn) gene, as annotated at http://www.ensembl.org, was amplified by PCR from cDNA; (2) The examined species was *Danio rerio* ^{leot1}; (3) zcrbn mRNA was dissolved in nuclease-free water at 200 ng/μl; (4) The length of the pectoral fins was measured using self-developed software, which allows curvature determination in a semi-automatic fashion. The deformity was classified as "no effect" for fins exceeding 85% of the control length, "mild" for fins corresponding to 75–85% of the control length, "medium" for fins shortened to 60–75%, and "severe" for fins shorter than 60% of the control length.

Table 1
Crystallization conditions and cryo protection.

Structure	Protein solution additives	Reservoir solution (RS)	Cryo solution
MsCI4·thalidomide	3 mM thalidomide 3% (v/v) DMSO	0.1 M magnesium chloride 0.1 M sodium citrate pH 5 15%(w/v) PEG 4000	RS + 15% (v/v) PEG 300
MsCI4·pomalidomide	3 mM pomalidomide 3% (v/v) DMSO	0.1 M sodium acetate pH 4.6 15 %(w/v) PEG 20,000	RS + 15% (v/v) PEG 300
MsCI4·lenalidomide	3 mM lenalidomide 3% (v/v) DMSO	15 %(w/v) PEG 6000 5 %(w/v) glycerol	RS + 10% (v/v) PEG 300
MsCI4·deoxyuridine	3 mM deoxyuridine	0.1 M citric acid pH 3.5 25 %(w/v) PEG 3350	–
$MsCI4^{Y101F}$·thalidomide	3 mM thalidomide 3% (v/v) DMSO	0.2 M sodium chloride 0.1 M phosphate-citrate pH 4.2 20 %(w/v) PEG 8000	RS + 10% (v/v) PEG 300

Table 2
Data collection and refinement statistics.

	MsCI4·thalidomide	MsCI4·pomalidomide	MsCI4·lenalidomide	MsCI4·deoxyuridine	$MsCI4^{Y101F}$·thalidomide
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions *a*, *b*, *c* (Å)	56.5, 60.1, 88.5	56.5, 60.1 88.5	56.7, 59.7, 88.1	56.8, 60.0, 88.4	56.8, 60.0, 88.4
Resolution (Å)	37.3–1.40 (1.48–1.40)[a]	37.3–1.45 (1.54–1.45)[a]	37.3–1.85 (1.96–1.85)[a]	37.4–1.80 (1.91–1.80)[a]	37.4–1.90 (2.01–1.90)[a]
R_{merge}	7.4 (73.2)	3.8 (44.9)	7.3 (41.6)	8.5 (50.5)	8.6 (77.3)
$I/\sigma I$	12.46 (1.73)	16.32 (2.45)	8.41 (1.96)	8.71 (2.33)	15.06 (2.59)
Completeness (%)	99.6 (97.7)	99.5 (99.1)	98.7 (98.0)	98.8 (97.4)	99.9 (99.5)
Redundancy	6.69 (6.39)	3.40 (3.37)	2.86 (2.81)	2.87 (2.74)	6.99 (6.91)
Refinement					
Resolution (Å)	37.3–1.40	37.3–1.45	37.3–1.85	37.4–1.80	37.4–1.90
No. reflections	57,028	51,290	24,829	27,173	23,212
R_{work}/R_{free}	0.16/0.19	0.17/0.20	0.19/0.24	0.16/0.22	0.18/0.24
PDB code	4V2Y	4V2Z	4V30	4V31	4V32

[a] Highest resolution shell is shown in parenthesis.

3. Results

3.1. Cereblon isoform 4 from M. gryphiswaldense

For crystallization and *in vitro* experiments, we designed constructs of full-length cereblon and isolated thalidomide binding domains from different eukaryotes as well as one bacterial representative, cereblon isoform 4 from *M. gryphiswaldense* MSR-1 (MsCI4; 124 residues). While the work with the eukaryotic constructs was cumbersome and did not yield crystals, the bacterial protein was highly rewarding: With soluble expression in *E. coli* at high yields and facile purification with standard chromatography methods, it poses a very robust model system. Further, we expressed and purified two mutants of MsCI4. The double mutant $MsCI4^{YW/AA}$ with the mutations Y83A and W85A is the equivalent to the binding deficient double mutant $hCrbn^{YW/AA}$ with mutations Y384A and W386A. The other mutant is $MsCI4^{Y101F}$, with a substitution in the binding pocket as described later on. To confirm that MsCI4 indeed binds to thalidomide, we employed a simple NMR-based assay that relies on chemical shift changes that occur upon ligand binding. Thereby we found that wild-type MsCI4 and $MsCI4^{Y101F}$ both have binding constants for thalidomide in the low micromolar range, whereas the binding constant of the double mutant $MsCI4^{YW/AA}$ was shifted to the millimolar range. Crystallization trials were performed for MsCI4 and $MsCI4^{YW/AA}$, both in apo form and in presence of thalidomide. Only the trial with MsCI4 and thalidomide was successful; crystals appeared within 2 days in a wide range of conditions.

3.2. Crystal structures in complex with thalidomide, lenalidomide and pomalidomide

The best dataset collected from these crystals could be scaled to a resolution of 1.4 Å in space group $P2_12_12_1$. The structure could be solved by exploiting the anomalous signal of the structural zinc ions of the β-tent fold; three such ions were located in the asymmetric unit, belonging to 3 MsCI4 molecules. All three monomers are bound to thalidomide and have the same overall conformation, resembling the expected β-tent fold with two orthogonal β-sheets (Fig. 1).

The thalidomide binding site is formed by the three conserved tryptophan residues W79, W85 and W99, corresponding to W380, W386 and W400 of hCrbn, which constitute 3 sides of a rectangular aromatic cage. W79 and W85 form opposite walls of the cage, sandwiching the glutarimide ring at van der Waals distance, and are orthogonal to W99. Opposite of W99, the glutarimide ring forms two hydrogen bonds to backbone atoms, one between the amino group and F77O and the other between the distal keto group and W79N. A third hydrogen bond is formed between the distal keto group and the hydroxyl group of a tyrosine, Y101, corresponding to F402 of hCrbn, which forms the bottom of the cage. (Fig. 1). From the racemic mixture of thalidomide used in the crystallization trial, the (S)-enantiomer is specifically selected by the geometry of the pocket. In further co-crystallization experiments we obtained the structures of the MsCI4·lenalidomide and the MsCI4·pomalidomide complex in the same crystal form as MsCI4·thalidomide. As lenalidomide and pomalidomide are derivatives of thalidomide that are only modified in the phthaloyl moiety, and the interactions with the binding pocket are exclusively mediated via the glutarimide ring, the binding modes of the three compounds are identical (Fig. 1).

Of all interactions that the ligands form with the binding pocket, the hydrogen bond with Y101 is specific for MsCI4, as Y101 corresponds to a phenylalanine (F402) in hCrbn. This led us to confirm the validity of our results in the point mutant $MsCI4^{Y101F}$, replacing the tyrosine to phenylalanine and thereby creating a binding pocket that should be chemically indistinguishable from hCrbn. NMR-based measurements of this mutant with

thalidomide yielded a binding constant in the same range as for wild-type MsCI4. This is further underpinned by a crystal structure obtained for the MsCI4^{Y101F}·thalidomide complex, which is identical to the wild-type complex, except for lacking a hydrogen bond to the substituted tyrosine (Fig. 1).

3.3. Uridine as a cellular ligand

Given the binding mode of thalidomide, what can we infer about a potential natural ligand? As cereblon is supposed to act as a DCAF of DDB1, a DNA related ligand seemed plausible. In this context, the binding mode and the structural similarity of the glutarimide ring to pyrimidine bases suggested to us that the natural substrate of the thalidomide-binding domain might be a pyrimidine nucleoside – uridine, cytidine or thymidine (Fig. 1a). Of these, only uridine and thymidine are theoretically able to form the same hydrogen bonds within the binding pocket as glutarimide does. Furthermore, the cage formed by the three tryptophan residues might sterically discriminate between thymidine and uridine, as the additional methyl group of thymidine would clash with the W99 side chain. We therefore assumed that the binding site would selectively bind uridine. To test this hypothesis, we assayed MsCI4 and MsCI4^{Y101F} with different nucleobases, nucleosides and deoxynucleosides using NMR. Both proteins indeed bind uracil, uridine and deoxyuridine with binding constants in the low micromolar range, comparable to the results obtained for thalidomide. In contrast, a lower limit for the binding constants for a panel of other pyrimidine nucleobases, nucleosides and deoxynucleosides could be established in the millimolar range (Fig. 2). In support of these findings, we subsequently obtained the structure of the MsCI4·deoxyuridine complex by co-crystallization, again with the same crystal form as for MsCI4·thalidomide. As anticipated, the uracil ring of the nucleoside is bound in the same way as the glutarimide ring of the IMiD compounds (Fig. 1).

To verify these findings *in vivo*, we examined the effects of thalidomide and pyrimidine nucleosides on zebrafish development based on a pectoral fin deformity assay reported by Ito et al. (2010). Therein we found that embryos treated with uridine exhibit an identical incidence of malformations as embryos treated with thalidomide, whereas treatment with cytidine or thymidine did not show any effect. As for thalidomide, the effect of uridine treatment could be rescued by the injection of zcrbn$^{YW/AA}$ mRNA, encoding the binding-deficient double mutant of zebrafish cereblon (Fig. 3).

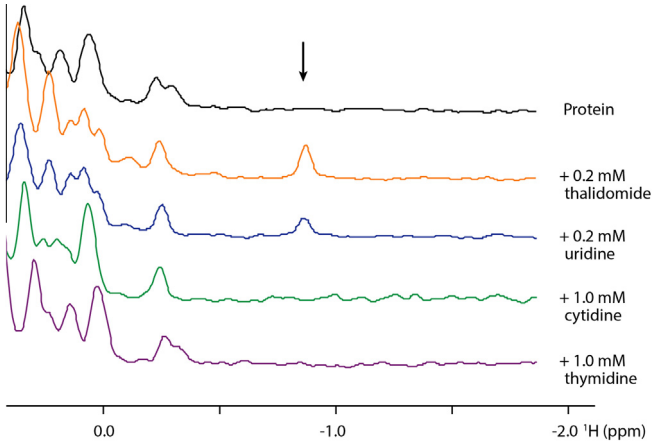

Fig.2. Thalidomide and nucleosides in the NMR-based ligand binding assay. The traces show 1D proton spectra on 50 μM MsCI4 at 600 MHz. A detail is shown of the region containing methyl groups that are up-field shifted by close contacts with aromatic groups, and are thus presumably within the protein core. Spectra are shown in the absence of ligand (black) or with various ligands at the indicated concentrations, with changes in signal positions and intensities indicative of binding. For thalidomide (orange) and uridine (blue), a signal appears upon binding (black arrow) that is characteristic of their specific binding mode. This signal does not appear for cytidine (green) and thymidine (purple). Similar spectra with varying concentrations of ligands were used to estimate binding constants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. The aromatic cage

While the inferences on uridine were driven by structural similarity on the ligand side, consideration of the nature of the binding pocket suggests yet another class of natural ligands: Aromatic cages of similar architecture are a characteristic, if not defining feature of many binding sites for compounds with methylated ammonium or guanidinium groups. Therein, ligand binding is dominated by cation-π interactions between the cationic ligand and the many π systems of the aromatic cage (Dougherty, 1996). Examples are found in binding sites for betaine- and choline-derived ligands in metabolism and neurotransmission, and in binding sites for the readout of posttranslational modifications of lysine and arginine residues. The latter are especially prominent in histone tails, where different methylation states of specific lysine and arginine residues have a direct effect on chromatin structure and gene expression, and are passed on to daughter cells as epigenetic markers (Gayatri and Bedford, 2014; Yun et al., 2011). In Fig. 4 we compare the aromatic cage of cereblon to representative exemplars of cages with specificity for defined methylation states of lysine and arginine residues, especially of PHD-, Tudor- and Tudor-like domains (Li et al., 2006; Su et al., 2014). While the majority of these cages are built from a mixture of phenylalanine, tyrosine and tryptophan residues, in individual cases they are dominated by tryptophan residues as in cereblon. Such an example is the cage mounted on an ankyrin repeat structure in Fig. 4. Therein, the tryptophan residues form hydrogen bonds with their indole residues that stabilize the cage structure, which could not be formed by phenylalanine or tyrosine. Likewise, the strictly conserved tryptophan residues in the cereblon cage appear to serve a similar stabilizing function. The binding of suitable ligands to the cage remains to be shown; our first attempts, assaying the binding of betaine- and choline derived ligands as well as individual lysine and arginine residues in different methylation states by NMR, were inconclusive.

4. Discussion

In this study we have established a simple and robust bacterial model system for the characterization of the CULT domain, the thalidomide binding domain of cereblon. Due to its easy handling, the bacterial protein is highly amenable to exhaustive biochemical characterization, enabling us to screen for ligands on a large scale. Here we report the first results of our characterization and screening efforts, which are driven by the insights derived from the presented high-resolution crystal structures.

With uridine, we unravel a natural cellular ligand that is universally present. As the previously identified downstream factors Ikaros and Aiolos are restricted to animals, they can only account for a subset of the total range of physiological substrates of cereblon. In contrast, uridine bears physiological implications for all organisms with cereblon proteins. We have shown that of all possible nucleosides, cereblon specifically binds uridine, that uridine causes the same teratogenic phenotype as thalidomide, and that it competes for the same binding site. Cereblon has been implicated in different activities, as in the modulation of ion channels (Liu et al., 2014), the regulation of AMP-activated protein kinase (Lee et al., 2011), and in the general development of the nervous

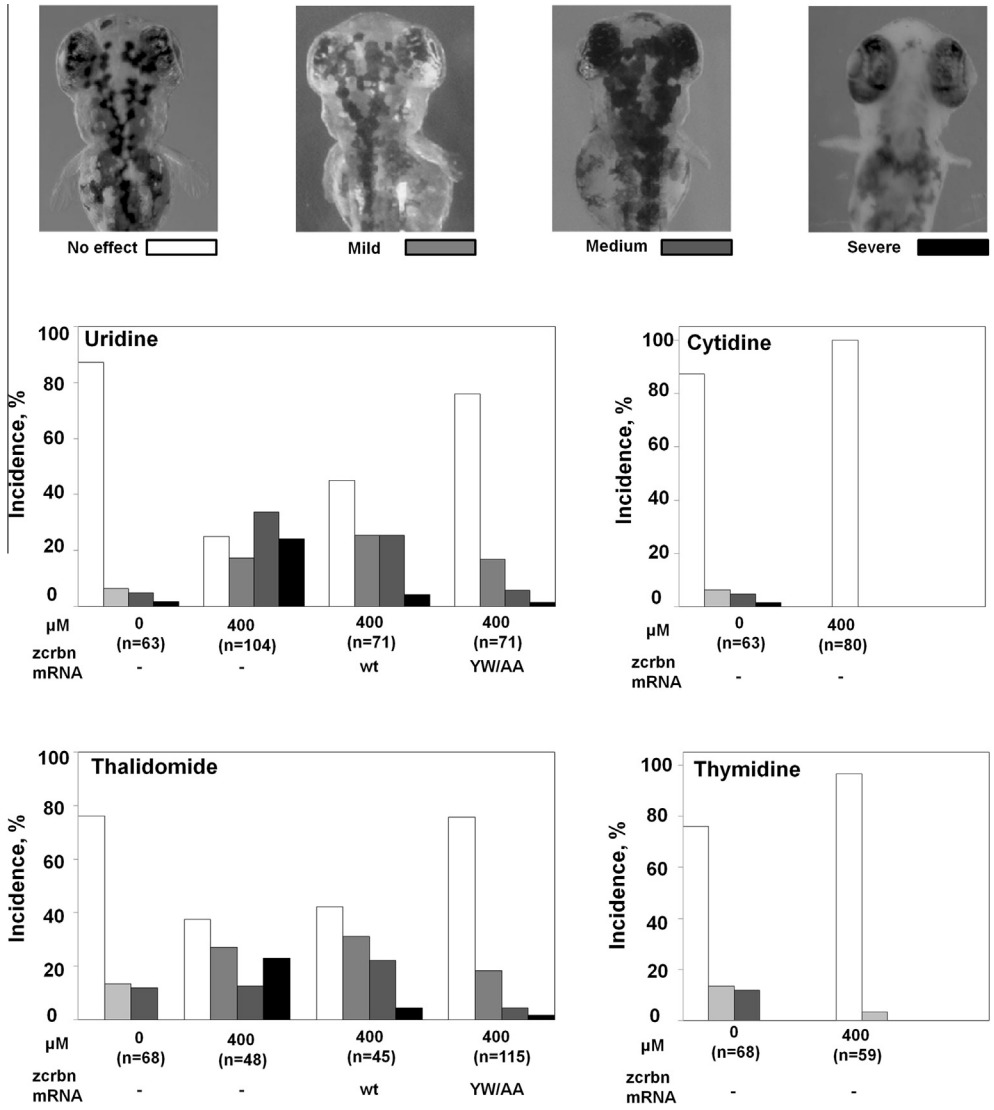

Fig.3. Zebrafish pectoral fin deformity resulting from thalidomide and nucleoside treatment. The images illustrate the diagnostic criteria of the pectoral fins of embryos at 75 h post fertilization. Uridine treatment causes similar developmental defects as evoked by thalidomide, whereas cytidine and thymidine treated embryos remain unaffected. As for thalidomide, the effect of uridine treatment can be rescued by injecting zcrbn$^{YW/AA}$ mRNA, but not by injecting zcrbn wt mRNA.

system (Chang and Stewart, 2011), in which the role of uridine binding is unclear. However, taken together with the fact that cereblon is a cofactor of DDB1, a central regulator of the cellular response to DNA damage, our results suggest that cereblon recognizes uracil in DNA. Uracil occurs in DNA because of misincorporation of dUMP or deamination of cytosine, and is usually excised by uracil-DNA glycosylases (Visnes et al., 2009). Therefore, one possible scenario for the role of cereblon is in DNA maintenance and quality control. However, as the DDB1-CUL4 E3 ligase system is only present in eukaryotes, bacterial cereblon can only express this function in another context. Indeed, MsCI4 alone showed no affinity for uracil-containing DNA.

Looking from another perspective, following the comparison of the aromatic cages, it is conceivable that the natural ligands of cereblon are methylated lysine or arginine residues in histone tails or transcription factors. It is established that the DDB1/CUL4 E3 ubiquitin ligase associates with several DCAFs that are core components of histone methylation complexes (Higa et al., 2006). Moreover, and more recently, the DCAF "DCAF1" was shown to recognize monomethylated lysines within a specific sequence motif, targeting histone and non-histone proteins for methyl-dependent ubiquitination by the DCAF1/DDB1/CUL4 ligase. (Lee et al., 2012). With an analogous involvement in the recognition of such modifications, cereblon in its E3 ubiquitin ligase complex could act in a DNA context, in which the binding of uridine would rather pose an ancillary effect. However, our binding studies in this direction were so far inconclusive. It should be noted that also Chamberlain et al. (2014) (see below) were not able to demonstrate the binding of small cationic ligands to a thalidomide binding domain.

The validity and accuracy of the bacterial model system, which is based on the high sequence similarity and essentially perfect conservation of the aromatic cage, is further supported by two recent studies: At the time of submission, Fischer et al. (2014), and later, during review of this work, Chamberlain et al. (2014) published crystal structures of cereblon proteins in complex with IMiD drugs from chicken, human and mouse with the respective resolutions of 3 Å, 3 Å and 1.9 Å. Fig. 5 shows representative structures of the thalidomide binding domains from these three organisms superimposed onto the MsCI4·thalidomide complex, all with RMSD values <1 Å and a virtually identical geometry of the aromatic cage. Together with our *in vivo* experiments, verifying

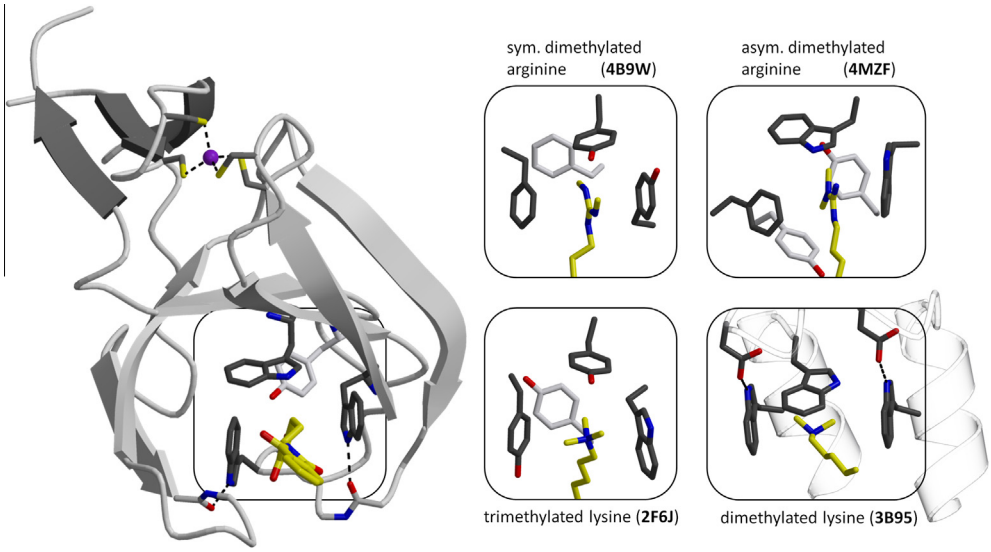

Fig.4. The aromatic cage of the thalidomide binding pocket compared to the aromatic cages for the readout of specific methylation states of lysines and arginines with the bound substrates symmetrical dimethylated arginine, asymmetrical dimethylated arginine, trimethylated lysine and dimethylated lysine. PDB codes are given in parenthesis. For the cages of cereblon and the ankyrin repeat on the bottom right, hydrogen bonds of the tryptophan side chains are indicated.

the relevance of ligands identified in the bacterial system for eukaryotic cereblon, this further highlights the major advantage of MsCI4 in allowing for high-resolution binding studies in a simple and minimal model system.

Using the structural insight gained by the presented crystal structures, we were able to derive certain structural rules for cereblon-binding ligands. This is exemplified by the selectivity for uridine over thymidine and cytosine, showing that the ligand needs to form specific hydrogen bonds while conforming to sterical requirements set by the aromatic cage. As glutarimide and uracil alone also bind to this pocket, it is conceivable that pharmaceuticals with accessible glutarimide- or uracil-like moieties could have teratogenic effects by binding to cereblon with the same binding mode. Moreover, it must be considered that pharmaceuticals might have metabolic products that fulfill the requirements for cereblon binding. In fact, in the performed *in vivo* assay, we cannot exclude that at least part of the effect is due to metabolic products that retained the glutarimide or uracil moiety. To gain more detailed insight into the specific structural parameters for cereblon-mediated teratogenicity, a systematic study of ligand specificity in needed. With a more precise knowledge about the structural requirements, pharmaceuticals could be rationally designed or modified to specifically bind - or not to bind - to cereblon. With this study we provide a robust protein construct amenable for high-throughput screening and a proof-of-principle to assess and refine these structural determinants.

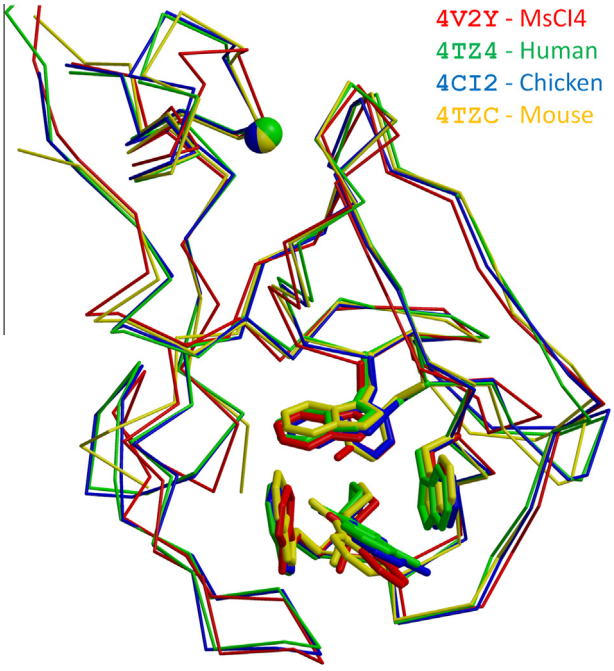

Fig.5. A comparison of MsCI4·thalidomide to IMiD-bound thalidomide binding domains from human, chicken and mouse, which were published shortly before submission (chicken) or during peer review (human, mouse) of this work. All structures superimpose with an RMSD < 1 Å.

5. Author contributions

M.D.H. designed the project, performed the crystallographic analysis and wrote the manuscript; I.B. performed the *in vivo* studies; M.C. performed NMR spectroscopy; F.Z. developed image-analysis software for the *in vivo* studies; A.N.L. designed the project and performed the bioinformatic analysis; B.H.A. designed the project and performed and supervised all biochemical work; All authors discussed the results and critically read the manuscript.

Acknowledgments

We gratefully acknowledge the skillful assistance of Kerstin Bär, Silvia Deiss and Julia Franke during sample preparation and of Reinhard Albrecht with crystallographic data collection. We thank the beamline staff at the Swiss Light Source for their excellent support and thank Uwe Irion and Christian Söllner for assistance in conducting *in vivo* experiments. This work was supported by institutional funds of the Max Planck Society.

References

Bartlett, J.B., Dredge, K., Dalgleish, A.G., 2004. The evolution of thalidomide and its IMiD derivatives as anticancer agents. Nat. Rev. Cancer 4, 314–322.

Chang, X.B., Stewart, A.K., 2011. What is the functional role of the thalidomide binding protein cereblon? Int. J. Biochem. Mol. Biol. 2, 287–294.
Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Vulpetti, A., Sundstrom, M., 2000. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J. Biomol. NMR 18, 65–68.
Chamberlain, P.P., Lopez-Girona, A., Miller, K., Carmel, G., Pagarigan, B., Chie-Leon, B., Rychak, E., Corral, L.G., Ren, Y.J., Wang, M., Riley, M., Delker, S.L., Ito, T., Ando, H., Mori, T., Hirano, Y., Handa, H., Hakoshima, T., Daniel, T.O., Cathers, B.E., 2014. Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs. Nat. Struct. Mol. Biol. 21, 803–809.
Dougherty, D.A., 1996. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science 271, 163–168.
Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.
Fischer, E.S., Böhm, K., Lydeard, J.R., Yang, H., Stadler, M.B., Cavadini, S., Nagel, J., Serluca, F., Acker, V., Lingaraju, G.M., Tichkule, R.B., Schebesta, M., Forrester, W.C., Schirle, M., Hassiepen, U., Ottl, J., Hild, M., Beckwith, R.E., Harper, J.W., Jenkins, J.L., Thomä, N.H., 2014. Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. Nature 512, 49–53.
Franks, M.E., Macpherson, G.R., Figg, W.D., 2004. Thalidomide. Lancet 363, 1802–1811.
Gandhi, A.K., Kang, J., Havens, C.G., Conklin, T., Ning, Y., Wu, L., Ito, T., Ando, H., Waldman, M.F., Thakurta, A., Klippel, A., Handa, H., Daniel, T.O., Schafer, P.H., Chopra, R., 2014. Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4(CRBN.). Br. J. Haematol. 164, 811–821.
Gayatri, S., Bedford, M.T., 2014. Readers of histone methylarginine marks. Biochim. Biophys. Acta 1839, 702–710.
Higa, L.A., Wu, M., Ye, T., Kobayashi, R., Sun, H., Zhang, H., 2006. CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. Nat. Cell Biol. 8, 1277–1283.
Iovine, B., Iannella, M.L., Bevilacqua, M.A., 2011. Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. Int. J. Biochem. Cell Biol. 43, 1664–1667.
Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., Handa, H., 2010. Identification of a primary target of thalidomide teratogenicity. Science 327, 1345–1350.
Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26, 795–800.
Kraulis, P.J., 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.
Kronke, J., Udeshi, N.D., Narla, A., Grauman, P., Hurst, S.N., McConkey, M., Svinkina, T., Heckl, D., Comer, E., Li, X., Ciarlo, C., Hartman, E., Munshi, N., Schenone, M., Schreiber, Carr, S.A., Ebert, B.L., 2014. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. Science 343, 301–305.
Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Lee, J.M., Lee, J.S., Kim, H., Kim, K., Park, H., Kim, J.Y., Lee, S.H., Kim, I.S., Kim, J., Lee, M., Chung, C.H., Seo, S.B., Yoon, J.B., Ko, E., Noh, D.Y., Kim, K.I., Kim, K.K., Baek, S.H., 2012. EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. Mol. Cell 48, 572–586.
Lee, K.M., Jo, S., Kim, H., Lee, J., Park, C.S., 2011. Functional modulation of AMP-activated protein kinase by cereblon. Biochim. Biophys. Acta 1813, 448–455.
Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., Patel, D.J., 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature 442, 91–95.
Liu, J., Ye, J., Zou, X., Xu, Z., Feng, Y., Zou, X., Chen, Z., Li, Y., Cang, Y., 2014. CRL4A(CRBN) E3 ubiquitin ligase restricts BK channel activity and prevents epileptogenesis. Nat. Commun. 5, 3924.
Lopez-Girona, A., Mendy, D., Ito, T., Miller, K., Gandhi, A.K., Kang, J., Karasawa, S., Carmel, G., Jackson, P., Abbasian, M., Mahmoudi, A., Cathers, B., Rychak, E., Gaidarova, S., Chen, R., Schafer, P.H., Handa, H., Daniel, T.O., Evans, J.F., Chopra, R., 2012. Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. Leukemia 26, 2326–2335.
Lupas, A.N., Zhu, H., Korycinski, M., 2014. The thalidomide-binding domain of cereblon defines the CULT domain family and is a new member of the β-tent fold. PLOS Comp. Biol., in press.
Merritt, E.A., Bacon, D.J., 1997. Raster3D: photorealistic molecular graphics. Methods Enzymol. 277, 505–524.
Meyer, B., Peters, T., 2003. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew. Chem. Int. Ed. Engl. 42, 864–890.
Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. 55, 247–255.
Perrakis, A., Morris, R., Lamzin, V.S., 1999. Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 6, 458–463.
Sheldrick, G.M., 2008. A short history of SHELX. Acta Crystallogr. A 64, 112–122.
Su, X., Zhu, G., Ding, X., Lee, S.Y., Dou, Y., Zhu, B., Wu, W., Li, H., 2014. Molecular basis underlying histone H3 lysine–arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. Genes Dev. 28, 622–636.
Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G., Krokan, H.E., 2009. Uracil in DNA and its processing by different DNA glycosylases. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 563–568.
Yun, M., Wu, J., Workman, J.L., Li, B., 2011. Readers of histone modifications. Cell Res. 21, 564–578.