

**The CULT domain of cereblon: a
pharmacological target and teratogenicity
gateway**

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Iuliia Boichenko
aus Schevchenkove, Kiliya Region
Odesa Gebiet, Ukraine

Tübingen
2016

Tag der mündlichen Qualifikation: 06.06.2016

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Andrei Lupas

2. Berichterstatter: Prof. Dr. Alfred Nordheim

SUMMARY

Half a century ago thalidomide caused one of the biggest pharmaceutical tragedies known by now. It was widely prescribed to pregnant women as a sedative, but displayed teratogenic properties, causing limb malformations and other developmental defects in more than 10,000 babies worldwide. Nevertheless, thalidomide and its derivatives are used nowadays in treatment of leprosy and multiple myeloma. The protein cereblon was identified as a primary target of thalidomide in the cell. As a substrate receptor, cereblon (CRBN) is linked *via* the damaged DNA binding protein 1 (DDB1) and cullin 4A to the E3 ubiquitin ligase machinery. The drug binds to the C-terminal region of cereblon, also referred to as CULT domain (cereblon domain of unknown activity, binding cellular ligands and thalidomide). This domain represents the most conserved part of cereblon and is also found solely in single-domain proteins in bacteria and, as a secreted form, in eukaryotes. Based on its ligand binding properties, its high degree of conservation and its intracellular as well as extracellular localization, a common interest arose in understanding the functional role of the CULT domain *in vivo*.

The CULT domain carries a number of highly conserved cysteine and tryptophan residues within its amino acid sequence. In the solved crystal structure of the bacterial CULT domain, four conserved cysteines stabilize the protein fold by coordinating a zinc ion. Three invariant tryptophan residues build an aromatic cage, to which the ligand binds. Considering the structural similarity of uridine and thalidomide, we tested uridine binding to the hydrophobic pocket and could show an identical mode of binding. So far, uridine represents the only natural ligand, for which an interaction with the CULT domain has been shown. Further studies demonstrated that parts of the CULT domain fold upon ligand binding, thus stabilizing the protein. The pocket is highly similar to the aromatic cages found in histone readers that recognize methylated lysine or arginine residues in chromatin. This structural similarity suggests analogous ligands for cereblon, which include a distinct type of post-translational modifications.

We developed a FRET-based *in vitro* assay for testing and characterizing ligand binding to cereblon. The determination and comparison of the substrate affinities for three CULT domains, from *Homo sapiens*, *Magnetospirillum gryphiswaldense*, and the secreted *Caenorhabditis elegans*, revealed similar values for the same ligands on a relative scale. This study convincingly confirmed the bacterial protein as a robust model system for both: (i) testing the specificity of the CULT domain to its ligands; and (ii) deciphering the function of cereblon proteins. Using the FRET-assay, we showed that various therapeutically relevant pharmaceuticals display affinities to the CULT domain with binding constants in the micromolar range. These off-

target effects were further validated by applying an *in vivo* assay in zebrafish embryos to test the teratogenic potential of these compounds mediated *via* their interaction with cereblon.

Searching for proteins interacting with cereblon *in vivo*, we identified transcription termination factor Rho to bind to the bacterial CULT domain. The assumption of a possible role of cereblon in transcriptional regulation is further supported by the fact that *hCRBN* interacts with LSD1, a demethylase which is essential for transcriptional regulation in multicellular organisms. Taken together, our data imply a potential function of cereblon in transcriptional repression and/or activation, particularly during limb formation.

ZUSAMMENFASSUNG

Der Contergan-Skandal in der Mitte des letzten Jahrhunderts war eine der größten Tragödien in der Geschichte der Arzneimittelentwicklung und -zulassung. Die Behandlung von Schwangerschaftsübelkeit mit dem Beruhigungsmittel Contergan basierend auf dem Wirkstoff Thalidomid, führte zu fatalen Fehlbildungen bei über 10000 Neugeborenen weltweit. Aufgrund ihrer sowohl entzündungshemmenden als auch immunsuppressiven Eigenschaften, finden Thalidomid und seine Derivate trotz ihrer teratogenen Eigenschaften auch heute noch eine breite Anwendung in der Behandlung von Lepra und Multiplem Myelom.

Fast 60 Jahre dauerte es, bis das Protein Cereblon, als ein zelluläres Target von Thalidomid identifiziert wurde. Als Substratrezeptor ist Cereblon durch Wechselwirkung mit DDB1 (*damaged DNA binding protein 1*) mit dem Cul4A E3 Ubiquitin-Ligase-Komplex verbunden. Durch die Bindung von Thalidomid an die C-terminale Domäne von Cereblon, auch als CULT-Domäne (*cereblon domain of unknown activity, binding cellular ligands and thalidomide*) bezeichnet, wird die Substraterkennung von Cereblon moduliert. Die CULT-Domäne ist hochkonserviert. Man findet ausschließlich aus dieser Domäne bestehende Homologe sowohl in Bakterien als auch in einer Gruppe von eukaryotischen sekretierten Proteinen. Aufgrund ihrer Eigenschaft Thalidomid zu binden, ihres hohen Konservierungsgrades und ihres sowohl extrazellulären als auch intrazellulären Vorkommens, besteht großes Interesse, die funktionellen Aspekte der CULT-Domäne *in vivo* zu verstehen.

Mit Lösung der Kristallstruktur einer bakteriellen CULT-Domäne konnten wir zeigen, dass streng konservierte Cysteinreste durch Bindung eines Zinkions zur Stabilisierung des Proteins beitragen. Ebenfalls hoch konservierte Tryptophanreste bilden eine Substratbindetasche, in der Thalidomid gebunden wird. Weitere Untersuchungen ergaben, dass Uridin, welches in seiner Struktur Thalidomid ähnelt, auf die gleiche Art und Weise gebunden wird. Es repräsentiert damit den bisher einzigen bekannten Cereblonliganden natürlicher Herkunft. Strukturelle Studien zeigten, dass Ligandenbindung das Protein stabilisiert, da einige Teile des Proteins erst dadurch falten. Die Bindetasche der CULT-Domäne zeigt eine starke Ähnlichkeit zu ebenfalls aromatischen, käfigartigen Bindestellen in anderen Proteinen mit verschiedenen Arten von Liganden. Dazu gehören z.B. die sogenannten *Histone reader*, welche methylierte Lysin- bzw. Argininreste binden.

Wir haben einen fluoreszenzbasierten *in vitro* Assay (FRET-Assay) entwickelt, welcher die Identifizierung und nachfolgende Charakterisierung von Cereblonliganden ermöglicht. Mit Hilfe dieses Assays zeigten wir in einer vergleichenden Analyse der CULT-Domänen aus *Magnetospirillum gryphiswaldense*, *Caenorhabditis elegans* und dem Menschen ähnliche Ligandenspezifitäten als auch -affinitäten. Dies ermöglicht die Anwendung des, im Vergleich zu seinen Homologen, sehr robusten bakteriellen Proteins als: (i) Modellsystem zur Identifizierung potentieller, pharmakologisch interessanter Cereblonliganden und (ii) zur Entschlüsselung der zellulären Funktion der CULT-Domäne. Unter Nutzung des bakteriellen Modellsystems konnten wir mit Hilfe des FRET-Assays die Bindung von therapeutisch relevanten Pharmaka an die CULT-Domäne mit Affinitäten im mikromolaren Bereich zeigen. Mögliche teratogene Auswirkungen dieser unerwünschten Wechselwirkung zwischen Cereblon und Arzneimittelwirkstoff als auch zwischen Cereblon und weiteren identifizierten Verbindungen wurden anschließend in einem *In vivo*-Assay in Zebrafischembryonen analysiert.

In Bakterien durchgeführte Experimente zur Identifizierung zellulärer Bindeproteine der CULT-Domäne zeigten eine Interaktion mit dem *Transcription termination factor Rho*. Eine mögliche regulatorische Funktion von Cereblon in der Kontrolle der Transkription wird durch den Befund unterstützt, dass für das humane Homologe eine Interaktion mit LSD1, einer für multizelluläre Organismen essentiellen Demethylase, gezeigt werden konnte.

Zusammengefasst lassen alle vorliegenden Ergebnisse und Beobachtungen eine potentielle Beteiligung von Cereblon an Vorgängen der transkriptionellen Regulation, insbesondere während der Embryonalentwicklung, vermuten.

CONTENTS

SUMMARY.....	3
CONTENTS.....	7
ABBREVIATIONS.....	10
1 INTRODUCTION.....	13
1.1 The teratogenic phenotype induced by thalidomide.....	13
1.2 Chemistry and pharmacokinetics of thalidomide.....	14
1.3 Development of thalidomide derivatives.....	15
1.4 Hypotheses of thalidomide teratogenicity.....	16
1.5 Species- and tissue-specificity of thalidomide action.....	19
1.6 Cereblon is a primary target in myeloma treatment.....	21
1.7 Cereblon is a protein with various functions.....	25
1.8 The domain structure of cereblon.....	27
1.9 The CULT domain.....	28
1.10 Aims of the thesis.....	31
2 MATERIALS AND METHODS.....	33
2.1 Cloning, expression, and protein purification.....	33
2.2 Electrophoretic mobility shift assay (EMSA).....	34
2.3 DNA pull down assay.....	34
2.4 Histone code peptide microarray.....	35
2.5 Analytical size-exclusion chromatography.....	35
2.6 <i>In vivo</i> assay in zebrafish.....	36
2.7 Immunoprecipitation of FLAG-MsCI4 and mass spectrometric analysis.....	37
2.8 Protein co-immunoprecipitation from bacterial cell extract.....	38
2.9 Western blotting.....	38
2. 10 Mammalian cell culture.....	39
2.10.1 Cloning procedure for mammalian expression.....	39
2.10.2 Mammalian cell culture maintenance and transfection.....	39

2.10.3 FLAG-CRBN immunoprecipitation and mass spectrometric analysis	40
2.10.4 Protein co-immunoprecipitation from mammalian cell extract.....	41
2.11 Generation of <i>C. elegans</i> transgenes.....	42
2.12 Nematode maintenance and synchronization.....	42
2.13 Preparation of <i>C. elegans</i> cell extracts.....	43
2.14 Anti-CeCRBN _{sec} Δ1-15 antibody production.....	43
3 RESULTS AND DISCUSSION.....	45
3.1 Thalidomide mimics uridine binding to an aromatic cage in cereblon.....	45
3.1.1 Synopsis.....	45
3.1.2 Own contributions.....	46
3.2 Structural dynamics of the cereblon ligand-binding domain.....	47
3.2.1 Synopsis.....	47
3.2.2 Own contributions.....	47
3.3 A FRET-based assay for the identification and characterization of cereblon ligands.....	48
3.3.1 Synopsis.....	48
3.3.2 Own contributions.....	48
3.4 Additional results and discussion.....	50
3.4.1 The CULT domain: structural insight.....	50
3.4.2 What are the physiological ligands of the CULT?.....	53
3.4.3 Cereblon-mediated teratogenicity of pharmacological compounds.....	58
3.4.4 Interaction of the MsCI4 with transcription termination factor Rho.....	62
3.4.5 Proteomic analysis of CULT-mediated <i>h</i> CRBN binding partners.....	68
3.4.6 Cereblon co-immunoprecipitates demethylase LSD1.....	74
3.4.7 Investigation on the sub-cellular localization of <i>C. elegans</i> secreted cereblon protein.....	79
3.4.8 Contributions.....	83
4 CONCLUSIONS.....	86

5 BIBLIOGRAPHY.....	88
6 ACKNOWLEDGEMENTS.....	97
Appendix I.....	98
Appendix II.....	101
Appendix III.....	102
Appendix IV.....	103
Appendix V.....	104
CURRICULUM VITAE.....	108
PUBLICATIONS.....	109

ABBREVIATIONS

AER	apical ectodermal ridge
AIDS	acquired immune deficiency syndrome
AMPK	AMP-activated protein kinase
ATP	adenosine triphosphate
ARNSMR	autosomal recessive non-syndromic mental retardation
Bmp	bone morphogenic proteins
BSA	bovine serum albumin
CD	circular dichroism
cDNA	complementary DNA
CeCRBN _{sec}	secreted cereblon protein from <i>C. elegans</i>
KG-1	human myeloid cell line
CRBN	cereblon
CK1 α	casein kinase 1 α
CoREST	corepressor for REST
CSNK1A1	casein kinase 1, alpha 1 coding gene
CtBP	C-terminal binding protein
Cul4	cullin 4
CULT	cereblon domain of unknown activity, binding cellular ligands and thalidomide
p21/waf1	cyclin-dependent kinase inhibitor 1/wild type activating fragment-1
DDB1	DNA damage-binding protein 1
DCAF	DDB1- and Cul4-associated factors
DDB2	DNA damage-binding protein 2
Dkk-1	Dickkopf Wnt signaling pathway inhibitor 1
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECL	enhanced chemiluminescent
EDTA	ethylenediaminetetraacetate
EHTM	euchromatic histone-lysine N-methyltransferase
EMSA	electrophoretic mobility shift assay
Fas	fas-associated death domain protein
FG	ferrite-glycidyl methacrylate
Fgf	fibroblast growth factor
FRET	fluorescence resonance energy transfer
gfp	green fluorescent protein
H3K4	Lysine 4 in histone 3
H929	human caucasian IgA-producing plasmacytoma
HDAC	histone deacetylase

HEK293T	human embryonic kidney cells
HUVEC	human umbilical vein endothelial cells
U266	human myeloma cells
HIC2	hypermethylated in cancer 2 protein
HMM	Hidden Markov Model
hpf	hours post fertilization
HRP	horseradish peroxidase
IFN β	interferon β
IKZF	Ikaros family zinc finger protein
IL	interleukine
IMiD	immunomodulatory drug
IGF-1	insulin-like growth factor 1
IRF4	interferon regulatory factor 4
IPTG	isopropyl- β -D-1-thiogalactopyranoside
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LDS	lithium dodecyl sulfate
LON	N-terminal domain of the ATP-dependent protease La
LSD1	Lysine-specific histone demethylase 1
LSM12	Sm-like protein 12
MsCI4	<i>Magnetospirillum gryphiswaldense</i> cereblon isoform 4
MANT	N-methylantraniloyl
MBT	malignant brain tumor domain
Me (1,2,3)	methylation (mono-, di-, tri-)
RMe2a	Arginine methylation (asymmetrical)
RMe2s	Arginine methylation (symmetrical)
MEIS2	myeloid ecotropic viral integration site 1 homolog 2
MeK	Lysine methylation
MsrB	Methionine sulfoxide reductase
Mis18	MIS18 kinetochore protein homolog A
MM	multiple myeloma
mRNA	messenger RNA
MS	mass spectrometry
MDS	myelodysplastic syndrome
NF-kB	nuclear factor kB
NK-cell	natural killer cells
NMR	nuclear magnetic resonance
Nrf2	NF-E2-related factor 2
NF-E2	nuclear factor-erythroid 2 transcription factor
PATZ	POZ-, AT-hook-, and zinc finger-containing protein
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDB-ID	Protein Data Bank-Identification
PHD	plant homeodomain

PKCB1	Protein kinase C-binding protein 1
PMSF	phenylmethylsulfonyl fluoride
PWWP	a conserved Pro-Trp-Trp-Pro motif
REST	RE1-silencing transcription factor
rCRBN	rat cereblon
RIG-I	regulatory domain of retinoic acid-induced gene-1
RMSD	root-mean-square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
<i>Sf9</i> cells	<i>Spodoptera frugiperda</i> cells
SILAC	stable isotope labeling by amino acids
SPIB	Spi-B transcription factor
SUMO	small ubiquitin-like modifier proteins
SUV39H1	suppressor of variegation 3–9 homolog 1; histone-lysine N-methyltransferase
TBS	Tris-buffered saline
TEV	tobacco etch virus protease
TGF- β 1	transforming growth factor beta 1
TNF α	tumor necrosis factor alpha
UTR	untranslated region
UV	ultraviolet
VEGF	vascular endothelial growth factor
bFGF	basic fibroblast growth factor
CIC-2	voltage-gated chloride channel
WD-40	40 amino acids repeat terminating in a tryptophan and aspartic acid
WIZ	widely-interspaced zinc finger-containing protein
Wnt	wingless type proteins
zCRBN	zebrafish cereblon
ZNF	zinc finger

1 INTRODUCTION

Thalidomide was elaborated in 1954 as a sedative by the pharmaceutical company Grünenthal. From 1957 to 1962 it was prescribed to pregnant women for the treatment of nausea. During these years the drug was distributed to over 40 countries in Europe, Australia, Japan, but banned in the USA because of concerns about peripheral neuropathy initiation. After its teratogenic effects were reported, thalidomide was immediately withdrawn from the market. By then, over 10,000 children worldwide were born with multiple defects, which turned to be the most tragic pharmaceutical catastrophe registered so far (Lenz, 1988). The use of thalidomide during the first trimester of pregnancy caused numerous birth anomalies of limbs (amelia and phocomelia), heart disorders, malformations of the inner and outer ear, ocular abnormalities in babies. In contrast, these effects were not detected on rodents, demonstrating that difference between species requires additional examination. This tragedy illustrated a sharp importance of a more rigorous and systematic testing of pharmaceuticals for developmental toxicity before their distribution into the market. Interest in thalidomide was reestablished due to its immunomodulatory, anti-inflammatory, anti-angiogenic properties, which are efficient in treatment of erythema nodosum leprosum, cancer, autoimmune diseases, and AIDS (Franks et al., 2004). On the account of its teratogenicity the prescription of thalidomide is strictly controlled by the System for Thalidomide Education and Prescribing Safety (S.T.E.P.S.) program in the USA (Zeldis et al. 1999). Nevertheless, the usage of thalidomide during pregnancy still causes many malformations in the newborn in developing countries due to a poor medical surveillance (WHO 2014; BBC News, 2013).

1.1 The teratogenic phenotype induced by thalidomide

As for any drug, the dose and exposure period are very important to balance its positive and negative properties. In 1960s it was demonstrated that the timing of exposure and the dosage of thalidomide played a critical role in evoking teratogenic effects (Lenz, 1988). The susceptible period of fetal development to drug treatment is between 20 and 34 days postfertilization. The window of exposure in humans for malformations of arms is 21-32 days after fertilization, for legs 27-32 days, for thumb hypoplasia 21-28 days, for outer ear

disorders 20-24 days, and for inner ear 24-34 days (Miller and Strömmland, 1999). Amelia (birth defect of absence one or more limbs) and phocomelia (birth defect of truncated limb(s)) were the most common thalidomide-induced abnormalities. Ear disorders were also often observed and varied from microtia to anotia. Along with main limb anomalies, other defects were reported, namely kidney malformations, dental disorders, central nervous system, and ocular anomalies. Furthermore, mental retardation, Duane syndrome, and autism were also recorded in affected kids, even though the rate was low. The risk of birth defects was estimated to be 10 % – 50 %. Even a single 50 mg dose was sufficient to cause severe harm, when the drug was consumed by pregnant woman during the first trimester of pregnancy (Newman, 1986). The mortality rate for the newborn “thalidomide babies” was 40 % mostly due to serious internal organs disorders.

1.2 Chemistry and pharmacokinetics of thalidomide

Thalidomide consists of two imide rings: phthalimide and glutarimide. It has two conformations, S- and R-enantiomers, which rapidly racemize under physiological conditions and are impossible to be separated from each other. The S-enantiomer is thought to be teratogenic, whereas R-enantiomer has sedative effect. Because of thalidomide’s low solubility in water, dimethyl sulfoxide (DMSO) is used as a solvent for its application in experiments. *In vivo*, it undergoes non-enzymatic hydrolytic breakdown, generating multiple chemically inactive metabolites. Additionally, thalidomide is hydroxylated by cytochrome P450 in the liver like most of the drugs during their metabolism in the body. The highest concentration of thalidomide in the blood is reached 3-6 hours after being administered. It is eliminated in the urine with the average half-life of 7.3 hours (Chung et al., 2004). Due to thalidomide’s poor solubility, its absorption rate in the gastrointestinal tract is much slower than its elimination.

The conventional dose of thalidomide applied for treatment of multiple myeloma and leprosy is 50-800 mg per day, which leads to the average concentration in the body between 3 and 48 μM according to single-compartment model (Chung et al., 2004). This amount of the drug is enough to cause both sedative and teratogenic effects. In rabbits the anti-angiogenic and the teratogenic activity was observed at 100-400 mg/kg dose, leading to approximate body concentration between 385 and 1,540 μM . Fin malformation in zebrafish was induced

by thalidomide at concentrations of 200-400 μM in E3 medium (Ito et al., 2010). The general concentration of the drug in the body is hard to estimate due to its unknown penetration properties through the epidermis. Seemingly, the working concentration of thalidomide varies even among thalidomide-sensitive species.

1.3 Development of thalidomide derivatives

The discovery of immunomodulatory and anti-cancer activity of thalidomide has awakened growing interest in an elaboration of structural analogues, which are less harmful and preserve the healing properties. In the last decade, multiple preclinical and clinical studies showed their high potential in the treatment of multiple myeloma and other malignancies. This demonstrated a vital importance for the development of immunomodulatory drugs (IMiDs) as well as thorough investigation of the mechanism of their anti-proliferative activity.

Today, the second generation of IMiDs is effectively used in the treatment of thalidomide-resistant myeloma pointing out the divergent characteristics of these compounds. Lenalidomide and pomalidomide, the most effective substitutes of thalidomide known by now (Fig. 1.1), possess less negative clinical consequences, although the teratogenicity issue still remains unresolved. While sharing important immunomodulatory properties, they differ in their efficiency and toxicity profiles in the treatment of different diseases. For instance, in zebrafish embryo, the effective immunomodulatory concentration of pomalidomide does not confer fins malformation compared to the other IMiDs (Mahony et al., 2013).

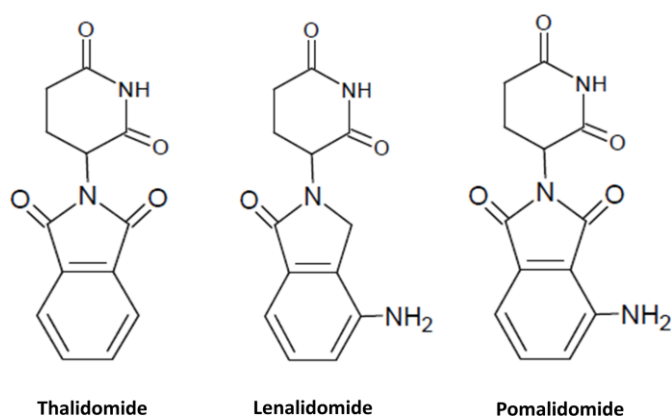


Figure 1.1. Structures of thalidomide, lenalidomide, and pomalidomide.

1.4 Hypotheses of thalidomide teratogenicity

More than half a century scientists worldwide tried to reveal the mystery of thalidomide teratogenicity. To elucidate molecular mechanisms of limb malformations, more than 30 hypotheses have been proposed, including:

- 1) oxidative damage
- 2) intercalation and inhibition of DNA synthesis
- 3) inhibition of angiogenesis
- 4) cereblon binding
- 2) DNA oxidation
- 3) modification of cytokine metabolism
- 4) downregulation of adhesion molecules
- 5) ascorbic acid synthesis
- 6) acylation of macromolecules
- 7) folic acid antagonism
- 8) glutamate metabolism misbalance.

Oxidative damage. Thalidomide has been shown to promote generation of reactive oxygen species (Hansen and Harris, 2004). Nuclear factor *κ*B (NF-*κ*B) is a redox-sensitive transcriptional factor, which activity is abnormally regulated upon thalidomide treatment. Fgf 8 and Fgf 10, the downstream targets of NF-*κ*B and essential factors for limb growth signaling, display a reduced expression in the mesoderm beneath the apical ectodermal ridge (AER) in rabbits after exposure to the drug (Hansen et al., 2002). Bone morphogenetic proteins (Bmp) and Dickkopf-1 (Dkk-1) are the other cytokines, which play an essential role in embryogenesis, and are upregulated by thalidomide (Knobloch et al., 2007). Furthermore, Bmps downregulate Fgf signaling in limb development (Pajni-Underwood et al., 2007). Importantly, they are expressed in the distal epidermal cells in the region of the AER and particularly necessary for limb patterning. Under physiological conditions NF-*κ*B negatively regulates abundance of Bmps, which are upregulated upon thalidomide exposure (Hansen and Harris, 2004). By mediating Bmps expression, thalidomide blocks Wnt and Akt signaling pathways, which control cell proliferation and survival (Knobloch et al., 2008; Grotewold and Ruther, 2002). The disturbance of these signaling pathways has been hypothesized to cause apoptosis of progenitor cells in limb buds.

DNA oxidation represents another harmful consequence of oxidative stress since the genetic code is under the threat of unpredicted irrevocable changes. Although this hypothesis seems to be plausible and additionally clarifies anti-proliferative activity of the drug, it does not explain the tissue-specificity of thalidomide-induced effects.

Anti-angiogenesis. The inhibition of blood vessel formation by thalidomide entailed limb deformities in different species from zebrafish to mammals (D'Amato et al., 1994, Therapontos et al., 2009, Ito et al., 2010). This statement advocates not only the teratogenic properties of thalidomide, but also the anti-cancer medication since many anti-neoplastic drugs are targeted to block vessel formation in tumors. Handa and colleagues showed that a violation of angiogenesis was rather a secondary consequence of the inhibition of vasculogenesis since transcriptional changes of Fgf8/Fgf10 and morphological abnormalities preceded marginal blood vessel formation (Ito et al., 2010).

DNA intercalation. Thalidomide was reported to interact with GC-rich promoters by intercalation, which leads to a disturbed transcriptional activity of the associated genes. For example, insulin-like growth factor 1 (IGF-1) and Fgf2 transcription was decreased as a result of thalidomide treatment in chicks. (for a review see e.g. Stephens et al., 2000). These proteins promote angiogenesis, therefore DNA intercalation theory is also consentient with anti-angiogenic activity of the drug.

Cereblon binding. As described above, none of the existing hypotheses clarifies all the aspects of thalidomide teratogenicity, although they substantially contribute to the knowledge about the mechanism of thalidomide action. They are unable to give an explanation to species specificity and, what is more important, provide no insight into direct target(s) of thalidomide in the cell. Identification of a primary target of thalidomide was a crucial step towards revealing the molecular mechanisms of caused teratogenicity. To accomplish this, Handa and co-workers developed ferrite-glycidyl methacrylate (FG) beads, which had a carboxyl-type derivative of thalidomide attached to amino-group of FG (Ito et al., 2010). These beads allowed the isolation of target molecules from the cell extract, using their magnetic properties. The extracts of human cell lines were used for affinity purification of the thalidomide-binding molecules. Two proteins of the apparent molecular mass of ~55 kDa and 127 kDa were eluted with free thalidomide after binding procedures, which signified that this interaction is thalidomide specific. Mass spectrometric analysis revealed that these were cereblon (~ 55kDa) and DNA damage-binding protein1 (DDB1, ~ 127 kDa). Further studies showed that CRBN is the direct target of thalidomide, whereas DDB1 indirectly interacts with

thalidomide *via* CRBN (Fig. 1.2). *Via* its interaction with DDB1, CRBN is connected to E3 ubiquitin ligase machinery, but the functional relevance of this interaction was unclear (Ito et al., 2010). In contrast to the intensively studied DDB1, the precise cellular role of cereblon was poorly understood. It is an adaptor protein associated with DDB2, which links two cellular processes: DNA excision repair and ubiquitination (Wittschieben and Wood, 2003). Further research demonstrated that DDB1 is a subunit of a cullin 4 (Cul4)-based E3 ubiquitin ligase machinery. The complex consists of Cul4 (Cul4A or Cul4B), regulators of cullins, DDB1, and a substrate receptor (Groisman et al., 2003), which binds to different substrates that are targeted for ubiquitin-dependent degradation and are subsequently processed by the proteasome.

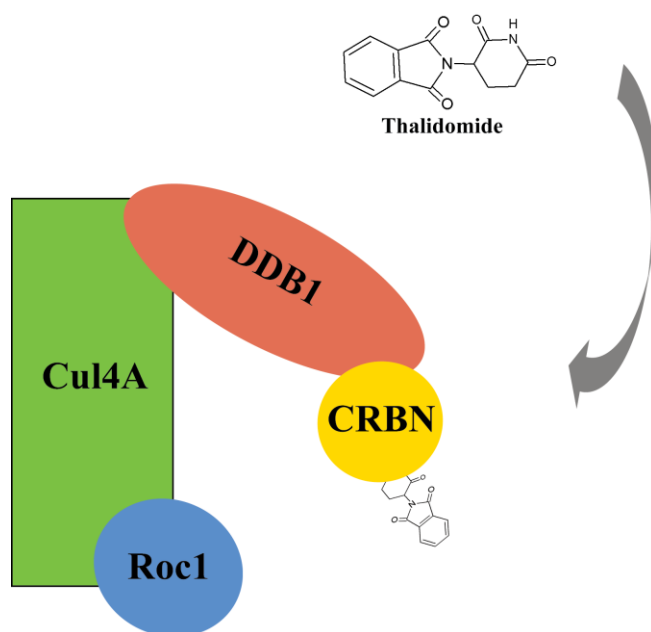


Figure 1.2. Cereblon functions as a component of E3 ubiquitin ligase complex (adapted from Ito et al., 2010).

This mission is crucial for sustaining a well-being of the cell as an excess of certain protein molecules in the specific time window might lead to a physiological misbalance. E3 ubiquitin ligase complexes control such processes as cell cycle, embryogenesis, immune response, and cancerogenesis (Bedford et al., 2011; Cohen and Tcherpakov, 2010). DDB1 is one of the adaptor proteins employed by Cul4, which binds numerous substrate receptors expanding the range of action of the whole machinery. They are called DDB1- and Cul4-associated factors (DCAF). Mostly, DCAFs contain a specific WD-40 domain (WDXR), for instance DDB2 and Cockaine syndrome A, but not cereblon. Nevertheless, CRBN is capable

to compete with DDB2 for binding to DDB1, presenting a novel type of substrate receptors (Ito et al., 2010). Thalidomide treatment blocks CRBN auto-ubiquitination, suggesting that the drug inhibits E3 ubiquitin ligase. The connection between the thalidomide-induced birth defects and CRBN was demonstrated *in vivo* in *Danio rerio* embryos. The zebrafish cereblon ortholog (zCRBN) shares ~ 70 % sequence identity to human CRBN. Both thalidomide exposure and *zcrbn* knockdown, cause pectoral fin defects, otic vesicles disorder, and reduction of Fgf8 expression in the fin buds in *Danio rerio* embryos. Tetrapod limbs and zebrafish fins are homologous in terms of early patterning and gene expression, albeit the skeletal structures are quite different in the adult forms. The thalidomide-induced phenotype was rescued in zebrafish and chicks by injection of the *zcrbn*^{YW/AA} mRNA encoding binding-deficient mutant of CRBN, but not by injection of wild type *zcrbn* mRNA. Although this undoubtedly confirms that CRBN is a direct *in vivo* target of thalidomide, the issues about its tissue-specific teratogenic effects and rodents' resistance to the drug are still debated. The key question: What are the cellular cereblon substrate(s)?; also remains poorly explored. Albeit there are few CRBN substrates found, they provide insight neither into CRBN-mediated thalidomide teratogenicity nor into universal cereblon function (Krönke et al., 2013; Lu et al., 2013; Gandhi et al., 2013; Krönke et al., 2015).

1.5 Species- and tissue-specificity of thalidomide action

Thalidomide administration leads to limbs deformities not only in humans, but also in monkeys, rabbits, and chicks (Knobloch and Ruther, 2008). Additionally, zebrafish embryos demonstrated fins malformations upon drug treatment (Ito et al., 2010). Rabbits and monkeys suffer from amelia and phocomelia, whereas in chicks only amelia occurs (Knobloch and Ruther, 2008). The fibroblast growth factor 8 (Fgf8) is expressed in the apical ectodermal ridge (AER) and plays an important role in limbs/fins development in vertebrates. Thalidomide application decreased expression of Fgf8 in the AER in susceptible animals, signifying a common teratogenic effect of the drug in different species. Rodents, however, are not susceptible to the thalidomide even at the exceptionally high doses (up to 4g/kg) (Brent, 1964). Given the high level of cereblon evolutionary conservation and binding of thalidomide to mice cereblon, the reason for such a difference between species still remains ambiguous (PDB-ID 4TZC, deposited, to be published by Chamberlain et al). Mutational analysis

revealed a single amino acid difference between human and mouse cereblon to provoke the blockage of lenalidomide-dependent CK1 α substrate ubiquitination in a myeloid cell line (Krönke et al., 2015). Although the role of this substrate in thalidomide teratogenicity is completely unknown, these data demonstrate differences in the responsiveness to lenalidomide that are based on a minor variation in the amino acid sequence and have a striking influence on cellular function. Complementary, the mystery of the species susceptibility to thalidomide is being discussed with various cereblon-independent hypotheses.

Firstly, the differences in pharmacokinetics have to be considered as a cause of species specificity. Thalidomide is quickly metabolized to many side products *in vitro* and *in vivo*, giving rise to non-teratogenic breakdown compounds (Melchert, List et al., 2007). Orally administered thalidomide has a half-life of 0.5 hours in mice, 2.2 hours in rabbits, and 7.3 hours in human plasma, supporting the idea of species-specific pharmacokinetic variance (Chung F et al., 2004). However, the drug does not cause limb defects in mice even with increased dosage (Janer, Slob., et al 2008).

Secondly, the requirement of prior bioactivation of thalidomide in the liver was suggested for its activity, which may be distinct in different species. This observation is based on the fact that the drug displays its anti-angiogenic activity in the presence of rabbit or human liver microsomes, but not the microsomes from the rat liver. In contrast, thalidomide-induced morphological and transcriptional changes were reported to be independent from its anti-angiogenic properties (Ito et al., 2010). Moreover, the bioactivation model contradicts the study showing thalidomide's direct interaction with cereblon, which has been proven multiple times by different groups (Ito, et al., 2010; Lopez-Girona et al., 2012; Zhu et al., 2011). In chicks and zebrafish, the teratogenic effects were caused by exposure of these animals to the drug at the developmental stages, when liver is not yet functional (Field et al., 2003; Ito et al., 2010).

Thirdly, given a short time-window of thalidomide action as a teratogenic agent, the minor differences in gestational development could cause distinct species response to the thalidomide exposure (Janer, Slob et al., 2008). Finally, generation of reactive oxygen species upon thalidomide treatment varies from species to species as well (Parman, Wiley et al., 1999).

Taken together, these hypotheses discuss various possibilities of the origin of species-specific thalidomide susceptibility, but none of them has provided its credibility yet. The

details of thalidomide teratogenicity remain to be elucidated, but numerous studies have paved the way for clarification the aspects of the tragedy.

The tissue specificity of thalidomide might descend from CRBN binding. In 48-hpf zebrafish embryos *zcrbn* expression is detectable at high level in the brain, pectoral fin, and head vasculature, but not in somites, which is consistent with detected drug-induced teratogenic effects (Ito et al., 2010). In comparison, in the adult human and mouse tissues CRBN is ubiquitously expressed (Su et al., 2004). Concurrently, CRBN is required, but might be not sufficient, for thalidomide teratogenicity.

1.6 Cereblon is a primary target in myeloma treatment

Thalidomide and related immunomodulatory drugs possess anti-tumor activity with particularly high efficiency in the treatment of multiple myeloma (MM). About 30 % of MM patients show therapeutic effects after treatment with thalidomide or its derivatives lenalidomide and pomalidomide. In combination with dexamethasone or bortezomib the efficacy of IMiDs was more than doubled, yielding a response rate for thalidomide > 60 % and even higher for lenalidomide > 80 % (Weber et al., 2007; Lacy et al., 2009). Despite such a remarkable achievement in clinical usage, the molecular mechanisms of anti-proliferative, anti-angiogenic, pro-apoptotic, and immunomodulatory effects of IMiDs remain elusive. After discovery of CRBN as a primary target of thalidomide in the cell, it was crucial to understand whether the anti-tumor activity of IMiDs is mediated by CRBN. Indeed, CRBN-deficient human multiple myeloma cells displayed IMiDs resistance, demonstrating that CRBN presence is an absolute must for the IMiDs therapeutic effect (Zhu et al., 2011). Moreover, CRBN expression level in the patients with reported lenalidomide resistance was significantly reduced (by 20 % - 90 %) before and after IMiD treatment, which again affirms that the response to the drug correlates with the CRBN level in the cell. Stewart and co-workers demonstrated that acquired lenalidomide and pomalidomide resistance is a consequence of CRBN expression reduction in the MM1.S cell line (Zhu et al., 2011). Additionally, to ascertain the role of CRBN in gained lenalidomide resistance, H929 cell line was exposed to increasing doses of the drug (Lopez-Girona et al., 2012). This caused a decrease of CRBN mRNA level accompanied with elevated anti-proliferative resistance, which correlates with published data. *Crbn* knockdown results in a significant mutilation of

IMiD-induced effects including IRF4 downregulation and cell cycle arrest, which mimics the usage of binding-deficient mutant CRBN^{YW/AA} (Zhu et al., 2011).

Fluorescent thermal melt shift studies illustrated that lenalidomide and pomalidomide bind cereblon with ten times higher affinity (Lopez-Girona et al., 2012). They inhibit autoubiquitination of CRBN within the DDB1-containing E3 ligase machinery similarly to thalidomide, which is abolished in the binding-deficient mutant CRBN^{YW/AA}. All three compounds induce expression of a key cell cycle regulator – the cyclin dependent kinase inhibitor 1 (p21/waf1), which controls cyclin dependent kinase activity (Lopez-Girona et al., 2012). The anti-angiogenic effect results from a reduced secretion of VEGF and bFGF from stromal and myeloma cells inhibiting endothelial cell adhesion and movement. Furthermore, IMiDs cause G0/G1 cell cycle arrest and induce apoptosis in multiple myeloma cells *via* an increased apoptosis protein-2 expression, elevated sensitivity to Fas induction, and activation of caspase-8. Interferon regulatory factor 4 (IRF4), a transcription factor regulating interferon signaling, is downregulated by IMiDs, causing death of the myeloma cell. Thalidomide and its derivatives also mediate production of TNF α , NF- κ B, IL-2, IL-6, which are essential for cell growth, adhesion, survival, and proliferation of MM cells (for a review see e.g. Shortt et al., 2013). IMiDs further stimulate T-cell activity, escalating the NK-T- and NK-cell cytotoxicity against cancer cells (Lopez-Girona et al., 2012). CRBN seems to be the main trigger of thalidomide, lenalidomide, and pomalidomide functioning, although the broad spectrum of their action has initially challenged the notion of a single target. These compounds might also possess some secondary effects as a result of promiscuous binding to off-target factors, which could explain the differences in their clinical influence.

Despite the pleiotropic effects of IMiDs, their main target in the cell remains CRBN within the DDB1-containing E3 ubiquitin ligase machinery. Altered ubiquitination of cereblon substrates seems to be a plausible explanation of IMiD-induced changes in the cell, although the knowledge about cereblon substrates is limited.

The homeobox transcription factor MEIS2 was identified as a first endogenous target of cereblon, the degradation of which was 2-fold decreased upon IMiD treatment (Fischer et al., 2014). MEIS2 is implicated in many aspects of human development, which partially overlap with thalidomide-affected organs. However, the phenotypes caused by MEIS2 overexpression/mutations differ from thalidomide-induced teratogenic effects.

In the 2014 study, Ebert and co-workers used SILAC-based quantitative mass-spectrometric analysis to characterize lenalidomide caused ubiquitination and degradation in

the multiple myeloma cells (Krönke et al., 2014). Two B-cell specific transcription factors, namely IKZF1 and IKZF3, showed decreased abundance upon compound treatment. Remarkably, these proteins interact with cereblon only in the presence of lenalidomide, which causes their cullin-dependent ubiquitination and subsequent degradation. CRBN knockdown or expression of the CRBN^{YW/AA} mutant abolishes IKFZ1 and IKFZ3 degradation and induces IMiD resistance in multiple myeloma cells. Strikingly, single amino acid change in IKFZ1 Q146H (or IKFZ3 Q147H) abrogates lenalidomide-dependent ubiquitination, providing evidence why lenalidomide possesses differential sensitivity for Ikaros family members (IKFZ2 and IKFZ4 are lenalidomide invulnerable) (Krönke et al., 2014). These findings are completely consistent with the data of Kaelin, Chopra, and their co-workers, who used another approach in determination of cereblon substrates, but deduced similar conclusions (Lu et al., 2014; Gandhi et al., 2013). The degradation of IKFZ1/3 is associated with a decreased level of IRF4 expression and enhanced T-cell IL-2 production as they regulate promotor regions of these genes. This data supports previous studies of anti-proliferative activity of IMiDs suggesting that IKFZ1/3 targets modulate the therapeutic response. IKFZ1 and IKFZ3 play an important role in B-cell biology, and their disposal from the cell explains the efficacy of lenalidomide activity in cancer treatment.

Casein kinase 1A1 (CK1 α) is another IMiD-induced cereblon substrate. Lenalidomide promotes its ubiquitination by E3 ubiquitin ligase, resulting in CK1 α degradation in myeloid cell line KG-1 (Krönke et al., 2015). The CK1 α mRNA is transcribed from CSNK1A1 gene, which deletion causes myelodysplastic syndrome (MDS) because of haploinsufficiency. Lenalidomide is highly efficient in the treatment of MDS promoting remission in more than 50 % of the cases. The ubiquitination and subsequent reduction of CK1 α level in the cell leads to apoptosis. Interestingly, it is not targeted for degradation by mouse cereblon due to an amino acid sequence difference. Only 5 variant residues were found in the thalidomide-binding domain of mouse and human cereblon. A substitutional analysis of non-conserved positions in human CRBN with the matching amino acids from mouse CRBN revealed that the exchange of human V387 for mouse isoleucine I391 disrupted the lenalidomide-induced substrate degradation. The same effect was also observed for lenalidomide-triggered ubiquitination of IKFZ1 and IKFZ3 in HEK293T cells with human CRBN V387I (Krönke et al., 2015). The bulkier isoleucine residue causes a spatial arrangement, disengaging hydrogen bond formation between cereblon and IKFZ or CK1 α and preventing their ubiquitination. Taken together, these data illustrate that lenalidomide selectively regulates IKFZ1/3 and

CK1 α levels, unraveling a new mechanism of action of IMiDs - cereblon gain of function effect. Consequently, cereblon was considered as a precursory biomarker for the efficiency of IMiD therapy (Gandhi et al., 2014).

Recent findings suggest two possible explanations of IMiD-induced teratogenicity: (I) inhibition of substrate binding with cereblon and their increased abundance in the cell; (II) creation of a new binding surface upon IMiD harboring, which causes ubiquitination of unusual substrates and their subsequent degradation. Besides, these substrates may vary in different cellular lineages, leading to tissue-specificity effects.

Lenalidomide and pomalidomide in cancer treatment. The clinical potency of lenalidomide is higher than thalidomide, but lower than pomalidomide, which has been shown in a number of cellular systems and *in vivo* (Lopez-Girona et al., 2012; Mahony et al., 2013).

Lenalidomide belongs to the second-generation of IMiDs possessing pleiotropic effects in cancer treatment, including increased anti-proliferative and immune cell stimulation activities. It downregulates IRF4 and SPIB transcription factors in a cereblon-dependent manner, which leads to an elevated IFN β production and blockage of the pro-survival NF- κ B signaling pathway in the diffuse large B-cell lymphoma. Additionally, lenalidomide reduces expression of anti-apoptotic factors; displays anti-angiogenic activity by modulation of TNF α , VEGF, and endothelial cell migration (Dredge et al., 2005); modifies cytokine production, thereby enhancing cytotoxic activity of CD8 $^+$ T-cells (for a review see e.g. Martiniani et al., 2012).

Pomalidomide is the third-generation IMiD displaying a higher potency in anti-myeloma activity in comparison to thalidomide and lenalidomide due to: (I) lower toxicity and reduced side effects at its working dose, (II) fast metabolism prior to excretion, and (III) efficiency in patients refractory to thalidomide and lenalidomide (Lacy et al., 2011). In comparison to its structural analogs, pomalidomide exhibits its anti-inflammatory activity at notably lower concentration that is non-teratogenic in zebrafish and chicken (Mahony et al., 2013). Although pomalidomide was recently licensed for the clinical usage, US Food and Drug Administration indicates that it may be teratogenic in some mammalian species.

Thalidomide does not induce limb deformities in rodents, but does it in rabbits. Some other developmental defects, namely rib, eye disorders were reported after thalidomide administration to pregnant rats (Parkhie and Webb, 1983). Concurrently, lenalidomide did not cause teratogenic effects in rats and rabbits, but affected monkeys' limb development (Celgene, unpublished data). Pomalidomide provoked developmental defects in rats and

rabbits, but was reported to be safe for zebrafish and chicken, when used at the anti-inflammatory working dose.

The molecular mechanism of IMiDs action is very complex including the divergence and complementarity in the treatment of different diseases. Unfortunately, the valuable properties of IMiDs appear to be inseparable from their harmful clinical side effects, such as nerve-damage, teratogenicity, myelosuppression, since they are caused by the same molecular targets. Considering their exceptional efficiency in multiple myeloma treatment, a new search of harmless IMiDs is highly beneficial. Alternatively, identification of cereblon downstream targets that can be therapeutically manipulated by novel drugs could significantly improve the clinical picture of multiple myeloma patients.

1.7 Cereblon is a protein with various functions

Human *crbn* contains 11 exons and encodes a 442–amino acid protein. CRBN is highly conserved from plants to humans and was originally identified as a candidate gene for autosomal recessive mild mental retardation (Higgins et. al., 2004). An R419X non-sense mutation in CRBN causes the truncation of the C-terminal 24 amino acids, which is linked to autosomal recessive non-syndromic mental retardation (ARNSMR), memory and learning defects. The mutant protein preserves its ability to form a fully functional E3 ubiquitin ligase complex with Cul4A and DDB1 since it ubiquitinates AMPK α 1 and calcium activated potassium channel (Xu et al., 2013). At the same time, R419X mutation enhances CRBN autoubiquitination and its subsequent proteasomal degradation, causing its decreased abundance in the cell. Interestingly, thalidomide reduces ubiquitination of wild type CRBN, but not CRBN R419X, indicating that it can't be considered as a drug against mental retardation. However, proteasome inhibitors are able to restore physiological level of CRBN R419X in the cells, thus appearing to be potential drugs for intellectual disability treatment (Xu et al., 2013).

Immunoprecipitation from rat brain lysate showed rat CRBN (rCRBN) interacts directly with the cytosolic C-terminus of the calcium activated potassium channel α subunit (Jo et al., 2005). *In situ* hybridization experiments represented the ubiquitous expression pattern of *rcrbn* mRNA with a prominent signal in brain, liver, kidney, and testis. Notably, the highest level of CRBN expression was detected in the hippocampus, cerebellum, and cortex,

what overlaps with mRNA distribution of calcium activated potassium channel (Jo et al., 2005; Aizawa et al., 2011). The cell surface expression and assembly of the channel tetramers are suppressed by rCRBN overexpression, suggesting an explanation of the involvement of cereblon in cognitive processes (Jo et al., 2005; Rajadhyaksha et al., 2011; Liu et al., 2014). Moreover, the voltage-gated chloride channel (CIC-2) was identified as CRBN binding partner, and shown to be functionally modulated in the retina (Hohberger, Enz, 2009; Chen et al., 2015).

CRBN is widely expressed in all tissues in the organism (Jo et al., 2005; Hohberger and Enz, 2009; Aizawa et al., 2011). Although it plays an important role in the functioning of the nervous system, the regulation of its expression in neuronal cells is only poorly understood. Herold and co-workers illustrated that mental disorders can be inflicted by hypoxia (Tenyi et al., 2008). To follow up this idea, Park and colleagues monitored the regulation of CRBN expression under conditions of hypoxia and reoxygenation (Lee et al., 2010). They registered an increase of CRBN mRNA and protein expression levels in mouse neuroblastoma cells, which was stimulated by transcription factor Nrf2. Nrf2 (NF-E2-related factor 2) is a transcription factor that not only binds cereblon promoter, but also activates antioxidant response elements, which protect the cell from the oxidative damage. Thus, CRBN is speculated to be one of the components of the antioxidant defense in the cell, but the experimental evidence supporting this hypothesis is missing (Lee et al., 2010).

Opportune protein degradation is crucial for the function of the central nervous system since it defines a synaptic plasticity and signal transmission. In addition to its involvement in ubiquitination process, CRBN was demonstrated to inhibit the 26S proteasome activity. By interacting with the $\beta 7$ subunit of the 20S core particle, CRBN disrupts its assembly in a neuroblastoma cell line (Lee et al., 2012). Many neuron disorders result from proteasome dysfunction and cereblon-mediated mental retardation might also be one of them in the light of cereblon R419X non-sense mutation. CRBN also binds to the $\alpha 1$ subunit of AMP-activated protein kinase (AMPK) and inhibits its activity, leading to dramatic changes in the consumption and production of ATP in the cell (Lee et al., 2011). Similarly, *crbn* knockout in mice causes a significant energy imbalance, especially when mice are on the fat-rich diet (Lee et al., 2013).

To elucidate the physiological role of cereblon, conditional and germ-line *crbn* knockout mice were generated (Rajadhyaksha et al., 2011). They are affected in hippocampus-dependent associative memory and learning, thus representing a good *in vivo*

model for studying functionality of these disorders. Park and colleagues studied influence of the whole body *crbn* deficiency in mice under a normal and a high-fat diet (Lee et al., 2013). *Crbn*-deficient mice showed a hyperphosphorylation of AMPK leading to a significant elevation of its activity, which demonstrated a negative regulatory effect of cereblon on AMPK.

1.8 The domain structure of cereblon

Cereblon received its name due to the presence of a LON domain, its high expression in the cerebellum and its important role in brain development. CRBN appears in all eukaryotes except fungi. Its size varies from 400 to 600 amino acids and it occurs only once in each genome (Lupas et al., 2015). Cereblon has a conserved LON domain located 50-100 amino acids from the N-terminus (Fig. 1.3). The LON domain comprises the N-terminal part of the ATP-dependent Lon protease, which does not possess the protease or ATPase activity, but is crucial for protein-protein interaction. The LON domain consists of two regions: an N-terminal six β -stranded pseudo-barrel, covered on one side by a helix (LON-N); and a helical bundle of four to five helices (LON-C). These two parts are separated by a stretch of amino acid residues, which was identified by Handa and co-workers as DDB1-binding region (Ito et al., 2010). A deletion analysis demonstrated that the cereblon mutant, which lacks the middle region between 187 to 260 amino acids, lost its capacity to interact with DDB1.

At the very N-terminus cereblon homologs have an extension of up to 100 amino acid residues, which is the least conserved part of the protein. The region flanking the LON domain at the N-terminus gains conservation from early eukaryotes to plants and humans. It has a notable FDxxLPxxHxYLG motif with yet unpredicted function (Lupas et al., 2015).

The C-terminal part of CRBN, which is most conserved among all species, is responsible for thalidomide binding (Ito et al., 2010). This region contains numerous strictly conserved residues preserved from early eukaryotes to humans. Sequence analysis additionally identified other proteins comprising only the thalidomide-binding domain, which was named CULT for cereblon domain of unknown activity, binding cellular ligands and thalidomide.

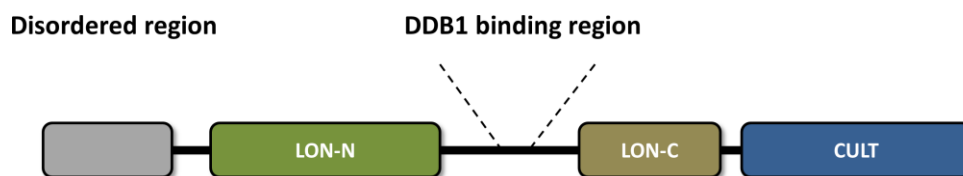


Figure 1.3. Schematic view of the domains of cereblon (modified from Lupas et al., PLoS Computational Biology, 2015).

1.9 The CULT domain

Regardless the high level of conservation and the wide distribution of the CULT domain, pointing on a basic function in nature, its role in the organism remains so far uncovered. Detailed bioinformatic analysis with a focus on the thalidomide-binding domain showed that, beside LON-containing cereblon proteins, other groups comprising a CULT domain do exist (Fig. 1.4). The two biggest clusters are: (I) prokaryotic proteins from δ -proteobacteria, a few representatives from α - and γ -proteobacteria, and one spirochete; and (II) animal proteins from simplest multicellular organisms up to fish (Lupas et al., 2015).

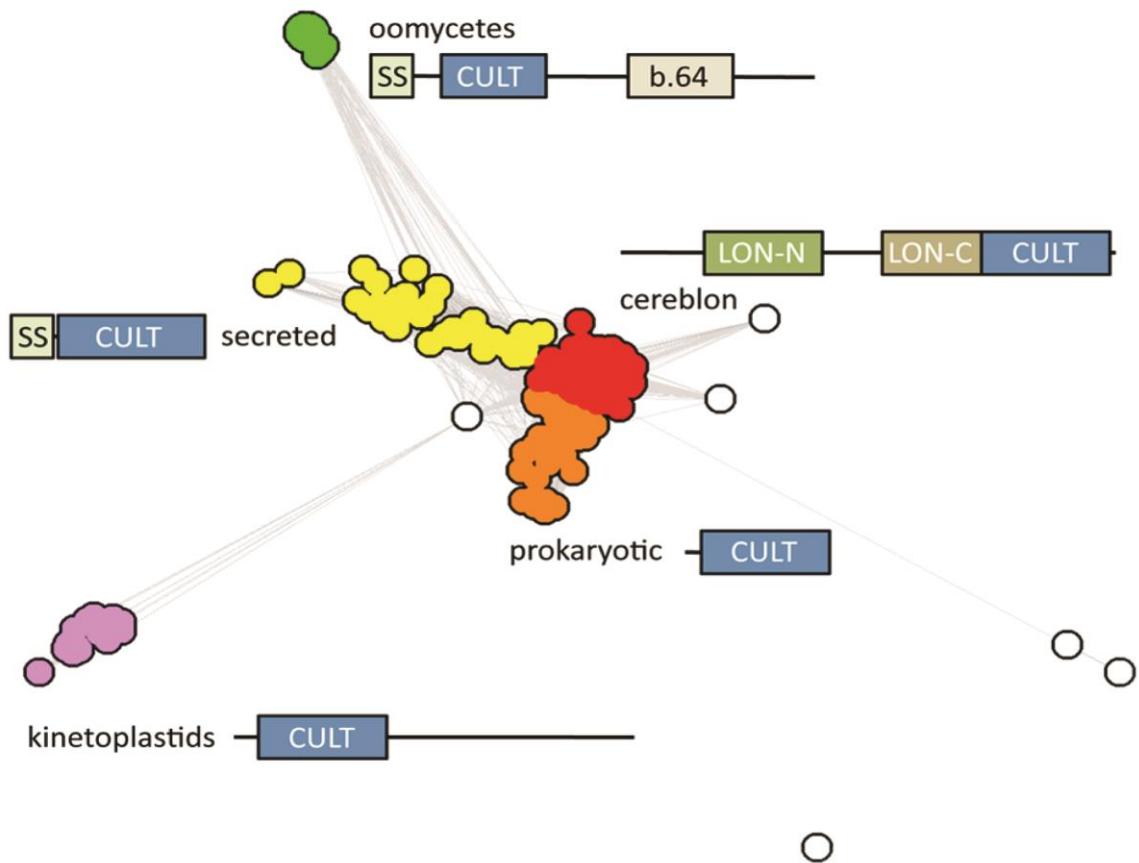


Figure 1.4. Cluster map of CULT domain proteins (adapted from Lupas et al., PLoS Computational Biology, 2015)

The bacterial cereblon proteins consist exclusively of the CULT domain, whereas their animal homologs additionally contain an N-terminal secretion signal sequence (Hostomska et al., 2009). The other two groups are evolutionary more distant and provide different context for the CULT domain. In oomycetes it is positioned between the N-terminal signal sequence and a carbohydrate-binding domain and in kinetoplastids the CULT precedes a C-terminal region with an unknown architecture and function. The multiple sequence alignment of representatives from the main groups of cereblon proteins illustrates a high level of conservation among different clusters (Fig. 1.5). Two CxxC cysteine motifs are a distinctive feature of a zinc binding site, which compose such a site in domains structurally homologous to the CULT domain. Further, the multiple sequence alignment shows three invariant tryptophan residues with their conserved neighborhood.

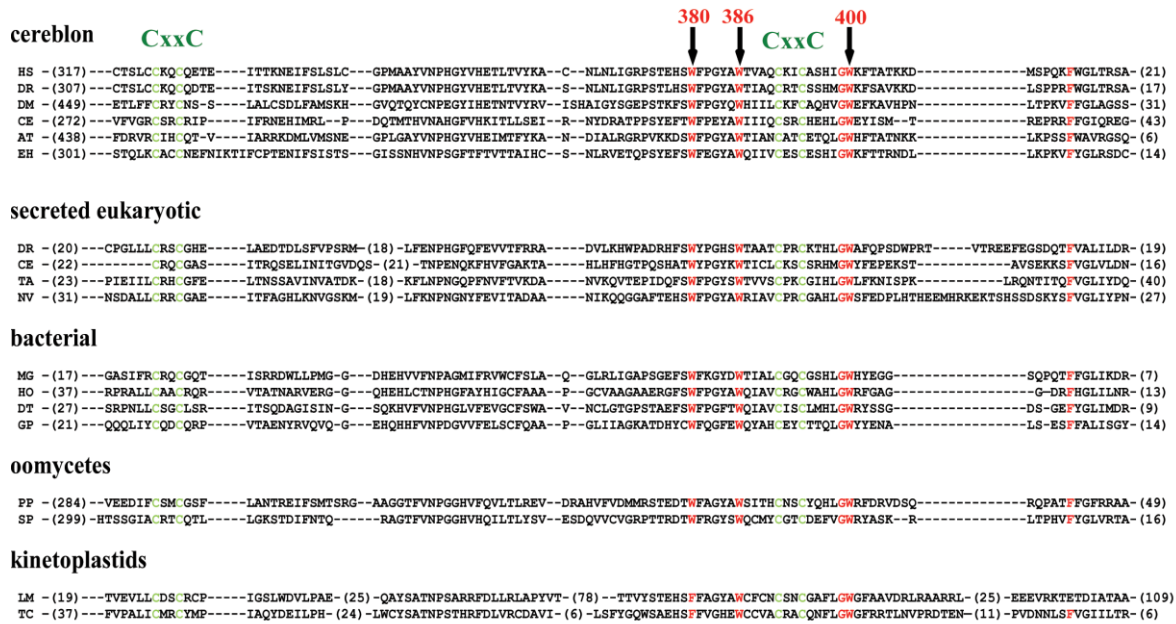


Figure 1.5. Multiple alignment of CULT domains from representatives of the groups shown in Fig. 1.4 (based on the results of a PSI-Blast search with the CULT domain of human cereblon). The sequences are: (cereblon) HS - *Homo sapiens*, DR - *Danio rerio*, DM - *Drosophila melanogaster*, CE - *Caenorhabditis elegans*, AT - *Arabidopsis thaliana*, EH - *Entamoeba histolytica*; (secreted eukaryotic) DR - *Danio rerio*, CE - *Caenorhabditis elegans*, TA - *Trichoplax adhaerens*, NV - *Nematostella vectensis*; (bacterial) MG - *Magnetospirillum gryphiswaldense*, GP - *gamma proteobacterium BDW918*, HO - *Haliangium ochraceum*, DT - *Desulfonatronospira thiodismutans*; (oomycete) PP - *Phytophthora parasitica INRA-310*, SP - *Saprolegnia parasitica CBS 223.65*; (kinetoplastid) LM - *Leishmania major strain Friedlin*, TC - *Trypanosoma cruzi Dm28c* (modified from Lupas et al., PLoS Computational Biology, 2015).

The CULT domain belongs to the group of β -tent fold proteins (Lupas et al., 2015). Members of this group share a common arrangement of two four-stranded, antiparallel β -sheets oriented at a right angle with a zinc ion at their tip providing the name for the whole group. Lupas and co-workers performed the search for remote homologs using profile Hidden Markov Model (HMM) comparison in HHpred and yielded a protein family of eukaryotes, yippee, as the best match for the CULT domain (Lupas et al., 2015). Mammalian yippee proteins are involved in signal transduction and anti-proliferative activity, but the mechanism of their action is still obscure (Kelley et al., 2010). Along with yippee, another eukaryotic protein Mis18 (absent in plants), which is implicated in centrosome assembly, was found

(Hayashi et al 2004; Fujita et al., 2007). Methionine sulfoxide reductase (MsrB) and the regulatory domain of retinoic acid-induced gene-1 (RIG-I) are two distant CULT homologs with known structure and completely distinct functions. MsrB is a universal protein found in all domains of life, that guards the cellular environment from oxidative damage by reducing methionine sulfoxide to methionine, whereas RIG-I is found only in animals and senses viral RNA (Lee B. et al., 2009; Leung and Amarasinghe, 2012). This data suggests that CULT domain, yippee, MsrB, RIG-I and other homologs not mentioned here can be grouped together based on their structural, but not functional similarities.

1.10 Aims of the thesis

Cereblon is a protein widely spread from plants to humans. Its C-terminal highly conserved CULT domain binds the drug thalidomide and mediates its teratogenic effects.

However, the natural ligands of this domain as well as its function are still unknown. The presence of homologs entirely composed of the CULT domain up to fish, together with the high grade of conservation points at the importance of this domain in nature and is indicative for a conserved function. At the beginning of the work within the framework of this thesis, the structure of a bacterial CULT domain from *M. gryphiswaldense* was just solved by our lab. Choosing the thalidomide-binding domains from human CRBN and the secreted *C. elegans* homolog, the first goal of this work was to solve the structure of a eukaryotic CULT domain to provide a basis for comparative structural and functional analyses.

Another important aim was the identification of so far unknown natural ligand(s) of the CULT domain. Based on its high stability, availability and easiness in handling, MsCI4 was used in our lab as a robust model system not only for structural studies, but also for the development of a FRET-based *in vitro* assay for determination of binding affinities of ligands to the CULT domain. Within the framework of this thesis, we aimed to use this assay to comparatively study the affinities of the CULT domains from human, worm, and bacteria for thalidomide and structurally related low-molecular weight substances, including various pharmacological compounds. Based on these *in vitro* studies, we intended to apply an *in vivo* assay using zebrafish embryos to validate the teratogenic potential of the identified ligands.

Aiming at the elucidation of the function of the CULT domain and its involvement in cellular processes, we further planned the use of MsCI4 in pull down analyses in the presence

or absence of thalidomide in order to identify binding candidates of the CULT domain in the cell.

Expecting that the additional domains of human CRBN, the N-terminal unstructured region and the LON domain containing the DDB1 binding region, play an important role in protein functionality and could provide a clue towards the understanding of the cellular role of the CULT domain in the eukaryotic cell, we further extended our studies beyond the bacterial model system. In order to gain first insight into the cereblon basic function in the eukaryotic cell, we outlined comparative proteome analysis of the proteins bound to human cereblon with and without drug treatment.

2 MATERIALS AND METHODS

In this chapter I am describing the materials and methods used during my PhD, but not presented in our publications. The according unpublished results are summarized in chapter 3.4

2.1 Cloning, expression, and protein purification

The coding DNA of target genes or gene regions was amplified by polymerase chain reaction (PCR) with sequence specific oligonucleotides carrying restriction enzyme sites, using the plasmid or cDNA as a matrix (Saiki et al., 1988). After running a 1 % agarose gel (Sambrook et al., 1989), the according PCR product was extracted from the gel (Qiagen kit), cleaved with appropriate restriction enzymes and subsequently ligated into an expression vector treated with the same enzymes unless otherwise specified. Cloning details for genes expressed for structural studies are summarized in the Appendix I.

FLAG-MsCI4 and MsCI4^{YW/AA} was constructed by inserting an annealed primer pair encoding the FLAG-tag and flanked by NcoI sites in the NcoI site of the available expression clone containing MsCI4 in pETHis1a (Hartmann et al., 2014). The resulting construct contains a 6xHis tag followed by a TEV cleavage site, a FLAG tag, and MsCI4 (Fig. 2.1.1).

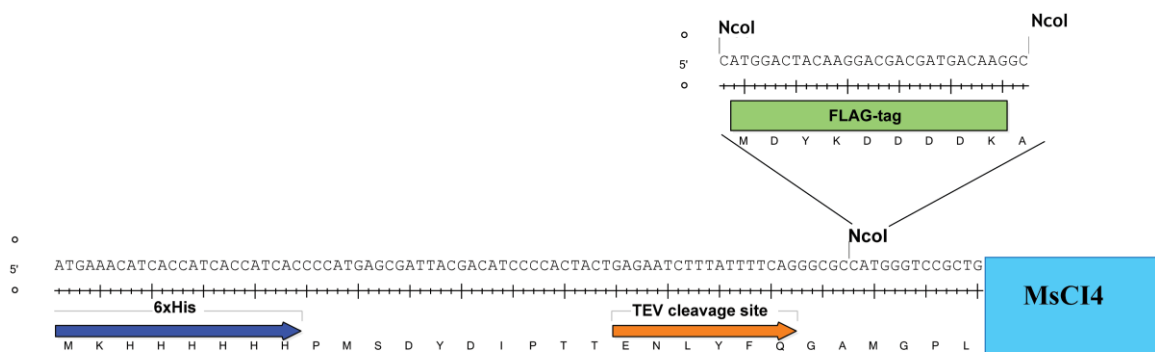


Figure 2.1.1. Insertion of the FLAG-tag DNA coding sequence into clone containing MsCI4 in pETHis1a (schematic overview). Only the coding DNA strand is depicted on the picture.

The genes of candidate binders of MsCI4 were amplified, using the *M. gryphiswaldence* genomic DNA, and cloned in the petHis1a vector. Further details of cloning and expression concerning these constructs are summarized in the Appendix II.

Plasmids were delivered into *E. coli* competent cells by heatshock. A positive clone selection was performed based on the antibiotic resistance and correctness was confirmed by DNA sequencing. Overexpression of the genes in *E. coli* was optimized with respect to temperature and isopropyl- β -D-1-thiogalactopyranoside (IPTG) concentration.

The standard purification protocol included loading soluble fractions of cellular extracts in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM 2-mercaptoethanol on a nickel affinity column followed by cleavage of the tag and a second step of affinity purification unless otherwise specified. SDS-PAGE analysis was performed at each stage to monitor purity of the sample (Laemmli, 1970).

Eukaryotic proteins were produced in the baculovirus-insect cell expression system as described here (Boichenko et al., 2016). For DDB1 isolation, additional ion exchange and size exclusion chromatography steps were required to obtain pure protein.

Folding of the proteins and formation of secondary structures was further validated by NMR or CD spectroscopy (Jasco CD-Spectropolarimeter J-810) (Kelly et al., 2005)

2.2 Electrophoretic mobility shift assay (EMSA)

Protein was mixed with RNA or either single-stranded or double-stranded DNA (optionally containing single uracil nucleotides) in a molar ratio of 1:10 (Appendix III). The mixture was incubated for 30 minutes in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 % glycerol, 0.5 mM 2-mercaptoethanol at room temperature. Afterwards samples were run on the NativePAGE (Invitrogen) or a 2 % agarose gel. Gels were stained, using Stain G (Serva) or 0.5 μ g/ml ethidium bromide solution (Hellman and Fried, 2007).

2.3 DNA pull down assay

The His-tagged protein was mixed with either single- or double-stranded biotinylated DNA (optionally containing single uracil nucleotides) in a ratio of 1:10 in the presence or

absence of thalidomide (Appendix III). The mixture was incubated for 1 hour in the binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 % glycerol, 0.5 mM 2-mercaptoethanol) at room temperature. Samples were further applied on streptavidin sepharose (Sigma) equilibrated with the binding buffer. The beads were washed 3 times with 1 ml wash buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 % glycerol, 0.5 mM 2-mercaptoethanol, 0.5 % Nonidet P-40). Afterwards LDS sample buffer was added to the beads and a Bis-Tris gel (Invitrogen) was run followed by Western blot analysis using anti-His antibodies (Wu, 2006).

For some experiments, the procedure was adapted in a way that cell lysate instead of binding buffer was added to the samples. Bacterial lysate was prepared as described by method 2.7. For studying the human protein, HEK293T cells overexpressing HA-tagged CRBN were used. The mammalian cell lysate was prepared as described by method 2.10.4. All steps were carried out at 4 °C.

2.4 Histone code peptide microarray

The peptide microarray slide (JPT Peptide Technologies) was incubated with 0.1 mg/ml His-tagged protein solution in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM 2-mercaptoethanol for 4 hours at 4 °C. Subsequently, the slides were washed 5 times with 1 ml TBST buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween) and incubated with mouse anti-His antibodies (1/1,000; Dianova) in 2 % BSA in TBST. Following 5 steps of a washing procedure, anti-mouse-Cy5 antibodies were applied in 2 % BSA in TBST (1/5,000; Abcam). After an additional washing step, the signal was recorded on the GenePix microarray scanner (Molecular Devices) using fluorescent signal. Data were analyzed with a PepSlide_Analyzer_Win-1.5.8 (SICASYS Software GmbH).

2.5 Analytical size-exclusion chromatography

Appropriate amounts of either RNA or DNA was incubated with protein in a molar ratio of 1:1 in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM 2-mercaptoethanol for 10 minutes. 200 µl of sample were further loaded on a Superdex 200 10/300 G column. RNA, DNA or

protein solutions were run separately as controls. Elution was recorded by UV absorption at 260 and 280 nm.

2.6 *In vivo* assay in zebrafish

Compounds. Glutarimide, hydantoin, succinimide, ethosuximide, γ -butyrolactam, δ -valerolactam, and sodium barbiturate (all from Sigma) were dissolved in water to make 400 mM stock solution.

Cloning. The zebrafish cereblon (*zcrbn*) gene annotated at <http://www.ensembl.org> was amplified by nested PCR from cDNA, using primer pairs 5'-AGAGCGGATTGTAAACACACGGC, 5'-TGTGTGAAGGCAACAACCTTGTGTTC for the first reaction set up and 5'-CGGATCCGATGGCTGCTGAGAGG, 5'-CTCTAGAGTCACAGGCACAGCAAGCGTG for the second. The PCR product was digested with BamHI and XbaI and cloned in the pCS2+ vector. All clones were verified by DNA sequencing. zCRBN possesses double mutation Y389A/W391A, which corresponds to the human CRBN^{Y384A/W386A} and MsCI4^{Y83A/W85A}. The positions of zCRBN mutation are different from described by Handa and co-workers since gene annotation was taken from <http://www.ensembl.org>, but not NCBI database (Gene ID: 445491) that results in 15 amino acid extended N-terminus of the protein (Ito et al., 2010).

Fish treatment. Zebrafish were maintained at 28°C on a 14-hours light / 10-hours dark cycle. The breeding was done as previously described (Nüsslein-Volhard C, Dahm R, 2002). The examined species was *Danio rerio*^{leot1}. The stock solutions of compounds were diluted in E2 medium to a final concentration of 400 μ M. 2 hours post fertilization, embryos were dechorinated in E2 medium containing 2 mg/ml pronase (Sigma) for 8 min at room temperature, washed five times, and immediately transferred to the medium containing the potential teratogen. Fish developed for 75 hours post fertilization in the drug-containing E2 medium, which was exchanged every 12 hours.

Microinjections of capped mRNA. Microinjections were carried out by injection of zygotes of 1-cell stage embryos under a microscope (Zeiss, Stemi 2000), using 275 Pa (40 psi) injecting pressure for 100 ms (World Precision Instruments, Pneumatic PicoPump PV820). After linearization of *zcrbn*_pCS2+ with NotI, capped mRNA was synthesized with mMessage mMachine kit and cleaned up *via* lithium chloride precipitation (both Ambion).

mRNA was dissolved in nuclease free water at 200 ng/μl and 20% (v/v) phenol red solution (Sigma- Aldrich, P0290-100ML).

Measurement of pectoral fins. 3 days old embryos were anesthetized and fixed in the 1% agarose for further imaging with Zeiss, SteREO Discovery. The length of the pectoral fins was measured using self-developed software, which allows curvature determination in a semi-automatic fashion. The diagnostic criteria were estimated as follows: "no effect" for fins exceeding 85% of the fin length of the control; "mild" for fins corresponding to 75-85% of the control length; "medium" for fins shortened to 60 to 75% of length; and "severe" for fins shorter than 60% of the control length.

2.7 Immunoprecipitation of FLAG-MsCI4 and mass spectrometric analysis

Bacterial cells were resuspended in the lysis buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 4 mM MgCl₂, 0.5 % Nonidet P-40, 0.5 mM DTT, 1 mM PMSF, protease inhibitors (EDTA-free, GE Healthcare), DNase I). Samples were sonicated, using a Branson Sonifier (G. Heinemann), centrifuged at 48,700 x g for 45 minutes at 4 °C to clear the lysates. The supernatants were collected and protein concentration was determined (Pierce BCA kit). In parallel, FLAG-MsCI4 or FLAG-MsCI4^{YW/AA} was coupled to anti-FLAG M2 magnetic beads (Sigma) equilibrated with binding buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 4 mM MgCl₂, 0.5 % Nonidet P-40). Equal amounts of cell extract were mixed with coupled beads in binding buffer, optionally in the presence of 50 μM thalidomide or the appropriate concentration of DMSO. 100 μl of the beads suspension was used for each sample condition. Samples were incubated for 4 hours at 4 °C with gentle mixing. After incubation, the resin beads were washed 3 times with 1 ml wash buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 4 mM MgCl₂, 0.5 % Nonidet P-40) to remove non-specifically bound proteins. Target molecules were eluted by incubation of the beads with 200μl of 0.1 M Glycine-HCl, pH 2.0 for 5 minutes with gentle shaking. Protein precipitation was performed overnight at -20 °C following addition of 800 μl of cold 100 % acetone. Solutions were centrifuged at 17,000 x g at 4 °C, pellets were collected and dried. After solubilization, protein concentration was determined by Bradford method (Bradford, 1976). Tryptic digestion of 12 μg/sample was performed at the Proteome Center (Universität Tübingen) with on the stage methylation labeling (Boersema et al., 2009). After mixing LC-

MS/MS analysis on a Proxeon Easy-nLC coupled to an LTQ-Orbitrap XL was conducted. The data were processed, using MaxQuant software (version 1.5.1.0). and had a setting of 1 % for the false discovery rate (Käll L., et al., 2008).

2.8 Protein co-immunoprecipitation from bacterial cell extract

E. coli cells overexpressing the candidate binding protein were resuspended in lysis buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 0.5 % Nonidet P-40, 5 % glycerol, 1 mM PMSF, protease inhibitors (EDTA-free (GE Healthcare), DNase I)). Samples were sonicated, using a Branson Sonifier (G. Heinemann), and centrifuged at $48,700 \times g$ for 45 minutes at 4 °C to clear the lysates. The supernatants were pre-cleared by incubation with 20 μ l protein A/G magnetic beads (Sigma) for 1 hour at room temperature. FLAG-MsCI4 or FLAG-MsCI4^{YW/AA} was added to the pre-cleared lysates at a final concentration of 0.8 mg/ml and incubated for 2 hours at room temperature. In parallel 40 μ l protein A/G beads were coupled to 10 μ g rabbit anti-FLAG antibodies (Dianova) for each sample. Subsequently, the coupled beads were combined with bacterial lysates containing overexpressed FLAG-MsCI4 or FLAG-MsCI4^{YW/AA} and incubated for 2 hours at room temperature. The beads were washed 3 times with 1 ml wash buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 0.5 % Nonidet P-40, 5 % glycerol) to remove non-specifically bound proteins. Target molecules were eluted by incubation of the beads with 200 μ l of 0.1 M Glycine-HCl pH 2.0 for 5 minutes with gentle shaking. Following addition of 800 μ l of cold 100 % acetone, proteins were precipitated overnight at -20 °C. Subsequently, solutions were centrifuged at $17,000 \times g$ at 4 °C, pellets were collected, dried, and dissolved in the LDS or SDS sample buffer (Invitrogen). Samples were loaded on the Bis-Tris gels (Invitrogen) and further analyzed by Western blotting.

2.9 Western blotting

Proteins were separated by SDS-PAGE and transferred onto a nitrocellulose membrane (pore size 0.45 μ m, Whatman) in 25 mM Tris, 150 mM Glycine, and 10 % methanol for 1 hour at 30 V, using an XCell II Blot Module (Invitrogen). The membrane was blocked with

5 % BSA in PBS with 0.1 % Tween 20 (PBST) for 2 hours at room temperature. Proteins were detected by incubation of the membrane for 16 hours at 4 °C with primary antibodies in 2 % BSA in PBST. Membrane was washed three times with PBST and incubated for 1 hour at room temperature with HRP-conjugated secondary antibodies in 2 % BSA in PBST. After washing with PBST, bands were developed with Clarity Western ECL substrate (Bio-Rad) and imaged with a FUSION Xpress (Peqlab).

The antibodies and their dilutions are listed in Appendix IV.

2. 10 Mammalian cell culture

2.10.1 Cloning procedure for mammalian expression

The coding DNA of target genes was amplified by PCR using the plasmid or human cDNA as a matrix for the reaction and gene specific oligonucleotides carrying sites for restriction enzymes. The PCR product was extracted from the gel (Qiagen kit), cleaved with appropriate restriction enzymes and subsequently ligated into an expression vector treated with the same enzymes. Cloning details for human genes used for mammalian expression are summarized in the Appendix V. Plasmids were delivered into *E. coli* competent cells by heatshock. A positive clone selection was performed based on the antibiotic resistance and confirmed by DNA sequencing.

2.10.2 Mammalian cell culture, maintenance, and transfection

Adherent HEK293T or HELA cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % fetal bovine serum, 100 µg/ml penicillin, 100 µg/ml streptomycin, and 2 mM L-Glutamine (all from Invitrogen) at 37 °C in a humidified atmosphere with 5 % CO₂ (DuBridge et al., 1986).

For transient transfection, HEK293T cells were seeded in medium without antibiotics in 100 mm dishes or 75cm² flasks (both Greiner Bio-One) and cultured till 90 % confluency. DNA and lipofectamine 2000 (Invitrogen) were mixed separately with OPTIMEM (Invitrogen) medium (unless other is specified), following the manufacturer's protocol. Two

solutions were combined, incubated at room temperature for 20 minutes for lipid-DNA complex formation and added dropwise to the cells. The medium was exchanged after 6 hours. For smaller dishes the amount of DNA was adjusted accordingly. Cells were harvested after 24 hours of transfection by centrifugation at 1,000 x g and frozen at -20 °C.

For SILAC-based mass spectrometric analysis HEK293T were grown for 6 passages in appropriate medium containing isotope-labeled amino acids. For each condition, cells from a 25 cm² flask were harvested by spinning them down at 1,000 x g. The pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 6 M urea, 2 M thiourea, 1 % N-Octylglucoside, 4 mM MgCl₂ and 10 µg/ml DNase I), vortexed, and incubated for 10 minutes at room temperature. Supernatants were prepared by centrifugation at 2,800 x g, 10 °C for 30 minutes and further given for an isotope incorporation check to the Proteome Center (Universität Tübingen). Transfection was performed with samples containing ≥ 96 % of labeled amino acids. The procedure was the same as it is described above with one modification: instead of OPTIMEM the SILAC medium (Invitrogen) deprived from amino acid isotopes was used for preparation of lipid-DNA complex. 20 µM thalidomide or appropriate DMSO concentration was used for cell treatment.

2.10.3 FLAG-CRBN immunoprecipitation and mass spectrometric analysis

Cells were resuspended in 500µl lysis buffer (20 mM Tris-HCl pH 7.5, 130 mM NaCl, 20 mM KCl, 10 % glycerol, 5 mM EDTA, 0.5 % Triton X-100, 20 µM thalidomide or appropriate DMSO concentration, 1 mM PMSF, phosphatase inhibitor cocktail 2 (Sigma, dilution 1:100), protease inhibitors (EDTA-free, GE Healthcare)), incubated on ice for 5 minutes, homogenized by 5–10 passages through a needle, centrifuged at 17,000 × g for 5 minutes at 4 °C to clear the lysates and the supernatants were collected. Protein concentration was determined to adjust it equally for all the samples (Pierce BCA kit). The cell extract was added to anti-FLAG beads (Sigma) preliminary equilibrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl and incubated for 4 hours at 4 °C with gentle mixing to capture the FLAG-CRBN. 100 µl of the beads suspension was used for each condition. Following incubation, the resin beads were washed 3 times with 1 ml wash buffer (20 mM Tris-HCl pH 7.5, 130 mM NaCl, 20 mM KCl, 10 % glycerol, 5 mM EDTA, 0.5 % Triton X-100, 20 µM thalidomide or appropriate DMSO concentration) to remove non-specifically bound

proteins. Target molecules were eluted by incubation of the beads with 0.1 M Glycine-HCl pH 2.0 for 5 minutes with gentle shaking. After determination of protein concentration (Pierce BCA kit), the eluates samples were further analyzed by MS analysis at the Proteome Center (Universität Tübingen).

2.10.4 Protein co-immunoprecipitation from mammalian cell extract

Overexpression was done separately for each protein of interest. 10^7 cells per sample were resuspended in lysis buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 10 % glycerol, 1 mM EDTA, 0.5 % Nonidet P-40, 1 mM PMSF, phosphatase inhibitor cocktail 2 (Sigma, dilution 1:100), protease inhibitors (EDTA-free, GE Healthcare)). The lysates were incubated on ice for 5 minutes, homogenized by 5–10 times passages of the cells through a needle, cleared by centrifugation for 30 minutes at 17,000 x g and 4 °C. The protein extracts with HA-CRBN and its FLAG- or Myc-tagged potential binding partner were combined and added to anti-HA magnetic beads (Sigma) preliminary equilibrated in the buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 10 % glycerol, 1 mM EDTA, 0.5 % Nonidet P-40). 40 µl of the beads suspension was used for each condition. Appropriate samples overexpressing HA- or FLAG-tagged LacZ were prepared as controls. Extracts were incubated for 4 hours at 4 °C with gentle mixing to capture the HA-CRBN. The resin beads were washed 3 times with 1 ml wash buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 10 % glycerol, 1 mM EDTA, 0.5 % Nonidet P-40) to remove non-specifically bound proteins. Target molecules were eluted by incubating the beads with 200 µl of 0.1 M Glycine-HCl pH 2.0 for 5 minutes with gentle shaking. Proteins were precipitated overnight at -20 °C after addition of 800 µl of cold 100 % acetone. Pellets were collected by centrifugation at 17,000 x g at 4 °C, dried and dissolved in the LDS or SDS sample buffer (Invitrogen). Samples were loaded on Bis-Tris or Tris-Acetate gels (Invitrogen) and further analyzed by Western blotting.

2.11 Generation of *C. elegans* transgenes

The gene of secreted *C. elegans* cereblon protein is annotated as R08B4.3. PCR fragments were amplified from wild type *C. elegans* genomic DNA as depicted in Fig. 2.11.1. A R08B4.3::GFP reporter was constructed by fusion PCR to obtain the R08B4.3 amplicon in frame with *gfp*, which was generated from the pPD95.75 (Hobert, 2002). Three different promoter regions were used separately for designing the constructs: natural promoter of R08B4.3, *ges-1*, and *myo-3*. PCR products were visualized by standard agarose gel electrophoresis. After purification from the gel (Qiagen kit), DNA was injected into gonads of adult worms together with the pRF4 plasmid (*rol-6*) as a coinjection marker as described (Mello et al., 1991). The expression of the coinjection marker restrains worm movement to a circle trajectory. All constructs were verified by sequencing.

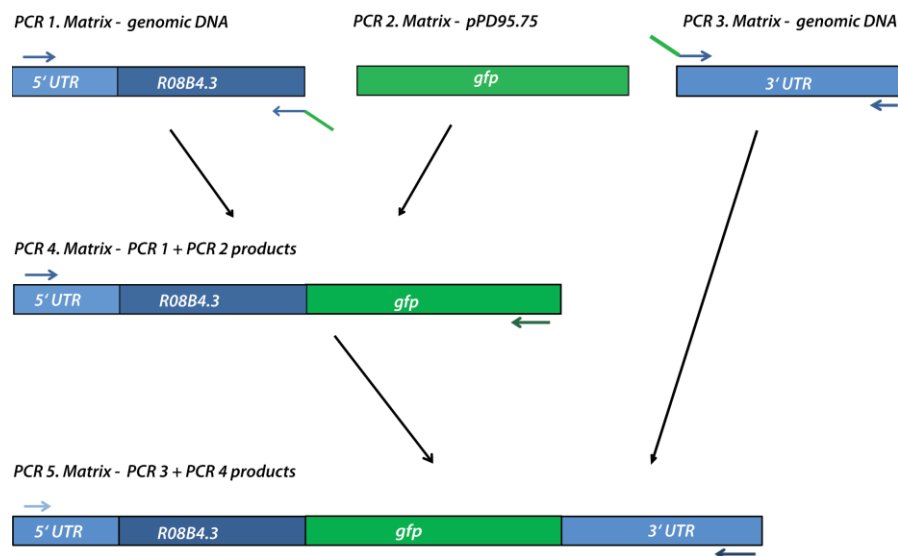


Figure 2.11.1. Outline of fusion PCR protocol for R08B4.3::GFP reporter construction.

2.12 Nematode maintenance and synchronization

Nematodes (*C. elegans* N2 Bristol strain) were maintained on NGM (Nematode Growing Media) agar plates seeded with *Escherichia coli* OP50 as food source and stored at 20 °C. Synchronized populations of *C. elegans* were obtained by hypochlorite treatment. Worms were allowed to grow to young adult stage and then injected into gonads. The

transgenes showing the features of marker expression were picked for the F2 generation, which were subsequently checked for the GFP signal that is indicative for intended construct expression.

2.13 Preparation of *C. elegans* cell extracts

Synchronized populations of *C. elegans* were resuspended in extraction buffer (50 mM HEPES pH 7.3, 130 mM NaCl, 20 mM KCl, 0.5 % Triton X-100, 10 % glycerol, 4 mM MgCl₂, 5 mM 2-Mercaptoethanol, 1 mM DTT, 1 mM PMSF and protease inhibitors (EDTA-free, GE Healthcare)), sonicated, and centrifuges for 30 minutes at 17,000 x g and 4 °C. Supernatants were collected and total protein concentrations were determined, using Pierce BCA kit. 30 µg of protein extract per lane of a 15 % SDS-PAGE was loaded and further analysed by Western blotting.

2.14 Anti-CeCRBN_{sec}Δ1-15 antibody production

Antigen preparation. Protein expression was performed at 37 °C for 4 hours and induced with 1 mM IPTG. Cells were pelleted, resuspended in 6 M Guanidinium chloride, 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 4 mM MgCl₂, DNase I and protease inhibitors (EDTA-free, GE Healthcare), and lysed, using a French pressure cell. The sample was applied on a nickel affinity column equilibrated with 6 M Guanidinium chloride, 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol. The His-tagged protein was eluted with a gradient of 0–0.5 M imidazole. Protein containing fractions were pooled and dialyzed against 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM 2-mercaptoethanol. The His-tag was cleaved overnight at 4 °C by TEV protease. The protein mixture was loaded on a nickel affinity column to bind the His-tagged TEV protease and the cleaved linker. The target protein was found in the flow through, pooled and concentrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM 2-mercaptoethanol, 0.1M L-Arginine. SDS-PAGE analysis was performed at each stage to monitor purity of the sample.

Rabbit immunization. Animals were screened before immunization by Western blot analysis by testing the pre-immune serum against CeCRBN_{sec}Δ1-15 at a dilution of 1:500. As

the tested serum did not show any pre-existing reaction against our protein of interest, one of the rabbits was subsequently injected with 1 mg/ml protein solution. An 8-week immunization schedule for the generation of the antibodies was applied, which included one priming injection followed by 2 biweekly injections and one booster injection followed by the blood collection.

Antibody purification. Serum was prepared after incubation of the sample at 37 °C for 1 hour and centrifugation at 1,000 x g for 15 minutes. The presence of antibodies in the serum was checked by Western blotting using purified *CeCRBN_{sec}Δ1-15* as an antigen. 65 days after initial injections, rabbit's blood was collected and used for serum preparation and further antibodies purification. IgG fraction was extracted overnight by ammonium sulfate (40 % of saturated solution) precipitation at 4 °C. Pellet was collected by centrifugation for 30 minutes at 3,000 x g and 4 °C, resuspended in and dialyzed against PBS buffer to get rid of ammonium sulfate leftovers. The antibody purification was performed, using a NHS-activated column (GE Healthcare) with *CeCRBN_{sec}Δ1-15* coupled to it following manufacturer's instructions. The working dilution of anti-*CeCRBN_{sec}Δ1-15* antibodies was determined by Western blotting and was as high as 1/1,000.

3 RESULTS AND DISCUSSION

3.1 Thalidomide mimics uridine binding to an aromatic cage in cereblon

Marcus D. Hartmann, Iuliia Boichenko, Murray Coles, Fabio Zanini, Andrei N. Lupas, Birte Hernandez Alvarez

Journal of Structural Biology. 2014; 188(3): 225–32

3.1.1 Synopsis

Thalidomide and related immunomodulatory drugs (IMiDs) lenalidomide and pomalidomide possess anti-cancer activity with a particularly high efficiency in treatment of multiple myeloma (MM). In addition to the therapeutic properties of thalidomide, its derivatives also inherited the teratogenic properties of the compound. Therefore, their usage is strictly controlled. The protein cereblon, which together with DDB1 and Cul4A forms an E3 ubiquitin ligase complex, was identified as a direct target of thalidomide in the cell. The C-terminal region of cereblon is responsible for thalidomide binding. It shares a high degree of conservation among bacteria and eukaryotes. This region was named CULT domain (cereblon domain of unknown activity, binding cellular ligands and thalidomide). Using high-resolution structures of a bacterial model system, we demonstrate how cereblon proteins bind their ligand. The domain carries a peculiar binding pocket built from three invariant tryptophan residues forming an aromatic cage for ligand docking. Additionally, the structure of “humanized” MsCI4^{Y101F} mutant was solved to provide an indistinguishable binding pocket of bacterial CULT from human cereblon. Furthermore, we show that IMiDs mimic the binding of uridine to the CULT domain, revealing the first physiological ligand of cereblon proteins universally present in all domains of life. These findings are further supported by an *in vivo* experiment demonstrating fins malformations in zebrafish embryos upon uridine treatment, analogous to an experiment performed with thalidomide by Handa and coworkers (Ito et al., 2010). The injection of the thalidomide binding-deficient *zcrbn*^{YW/AA} mRNA, but not *zcrbn* wild type mRNA, rescues the phenotype, demonstrating that this effect is indeed cereblon-dependent. Based on these data and the knowledge about the interaction of human cereblon with DDB1, we supposed a possible involvement of cereblon in the recognition of uracil in DNA. However, experiments investigating this possibility failed to show any binding of the CULT domain to

RNA as well as uridine containing DNA so far (see chapter 3.4.2). Alternatively, the architecture of the binding pocket suggests binding of cationic ligands such as methylated lysine or arginine residues. It is conceivable that cereblon recognizes such a modification of its substrates similar to histone readers, which recognize named post-translational modifications in chromatin.

The high sequence similarity between the bacterial and the eukaryotic CULT domain supported by their structural conservation points at the fact, that the bacterial protein represents an accurate model system for studying the CULT domain.

3.1.2 Own contributions

I performed the *in vivo* studies published in this paper. These included cloning procedures of *zcrbn* wild type and mutant gene, *in vitro* transcription, breeding of the fish, injections of mRNA into embryos, imaging of embryos, and data analysis. Silvia Deiss helped me to mount the embryos in the agarose before imaging. Together with Fabio Zanini (AG Neher) we discussed the algorithm of data analysis and he wrote an image analysis program. Finally, I prepared the according figures and contributed to the preparation of the manuscript.

3.2 Structural dynamics of the cereblon ligand binding domain

Hartmann MD, Boichenko I, Coles M, Lupas AN, Hernandez Alvarez B

PLoS ONE.2015; 10(5): e0128342.

3.2.1 Synopsis

The protein cereblon, the primary target of immunomodulatory drug thalidomide in the cell, has been structurally characterized in several studies (Hartmann et al., 2014; Fischer et al., 2014; Chamberlain et. al., 2014). Its thalidomide-binding region, named CULT domain, possesses an aromatic binding pocket formed by three invariant tryptophan residues, which is structurally conserved from bacteria to eukaryotes. Structural studies show the glutarimide ring of different ligands, including thalidomide, lenalidomide, pomalidomide, and uridine to bind to the pocket in the same fashion. With MsCI4 from *M. gryphiswaldense*, we have established a bacterial protein as a proper model system for the structural characterization of the thalidomide-binding domain on the basis of high sequence and structure similarity. In the present study, we aimed to extend our knowledge about possible conformational changes upon cereblon substrate recognition by comparison of the ligand-free and the ligand-bound state of the bacterial CULT domain. Analysing soaking experiments, we show that the CULT domain adopts a different conformation upon ligand binding. Washing out the ligand from the crystals causes unfolding of about one third of the domain. The effect of ligand binding on protein stability was examined by biophysical assays revealing an increase in thermal stability of the domain upon interaction with thalidomide or uridine. Our data indicate that a large proportion of the CULT domain folds in the presence of its ligand. Using our bacterial model system, we further show, that the C-terminal truncation of 24 amino acids in cereblon, which causes mild mental retardation in humans, does not affect ligand binding to the pocket. Finally, our studies unravel the structural flexibility of the CULT domain upon substrate binding, providing a further piece in the puzzle towards deciphering cereblon function in the cell.

3.2.2 Own contributions

I performed thermal shift assay studies, analysed the data, prepared the according figures, and contributed to the preparation of the manuscript.

3.3 A FRET-based assay for the identification and characterization of cereblon ligands

Iuliia Boichenko, Silvia Deiss, Kerstin Bär, Marcus D. Hartmann, Birte Hernandez Alvarez

Journal of Medicinal Chemistry; 2016; DOI: 10.1021/acs.jmedchem.5b01735

3.3.1 Synopsis

IMiD binding modulates the affinity of cereblon for its substrates. Degradation of the homeobox transcription factor MEIS2, which is implicated in many aspects of human development, is impaired by IMiD treatment (Fisher et al., 2014). Furthermore, IMiDs have been shown to stimulate the degradation of IKFZ1/3 and CK1 α in a cereblon-dependent manner, thus triggering the anti-cancer effect of the drugs. Therefore, based on its functional modulation upon ligand binding, cereblon represents on one hand a useful tool for targeted protein degradation that is highly valuable in cancer treatment. On the other hand, cereblon mediates the teratogenic effects of IMiDs. In the present study, we establish a FRET-based *in vitro* assay for testing the binding affinities of cereblon to potential ligands. Members of the three main groups of proteins containing a CULT domain, namely the thalidomide-binding domain from *H. sapiens*, MsCI4 from *M. gryphiswaldense*, and the secreted cereblon homolog from *C. elegans*, were taken for a comparative analysis exploring their ligand binding properties. The “humanized” MsCI4^{Y101F} mutant was used since it has an indistinguishable binding pocket from human cereblon. We demonstrate that the three proteins, on a relative scale, show identical substrate affinities with similar specificities for selected ligands, including succinimide, which was identified as a new binding moiety for cereblon proteins. Based on its high stability, availability, and easiness in handling, we demonstrate that, using MsCI4 as a model protein, our assay presents a robust tool to screen for new cereblon ligands and also to test novel pharmacologically interesting compounds for their teratogenic potential mediated *via* cereblon.

3.3.2 Own contributions

I designed the construct and cloned the gene of the secreted *C. elegans* protein. I set up the baculovirus insect cell expression system, which was used for production of the two

eukaryotic proteins (the human CULT domain and the secreted *C. elegans* homolog). Further I optimized expression and developed a purification scheme for the two eukaryotic proteins from *Sf9* cells. Finally, I contributed to the preparation of the manuscript.

3.4 Additional results and discussion

In this chapter I am describing unpublished results obtained during my PhD work.

3.4.1 The CULT domain: structural insight

The mystery of thalidomide teratogenicity has been explored for many years, but only recently the protein cereblon was identified as a cellular target of the drug (Ito et al., 2010). It is widely spread among eukaryotes, except fungi, and functions as a substrate receptor of the CUL4-DDB1 E3 ubiquitin ligase complex.

IKFZ1/3 and CK1 α are degraded upon treatment of the myeloma cells with lenalidomide, which is the structural homolog of thalidomide (Krönke et al., 2014; Lu et al., 2014; Gandhi et al., 2013; Krönke et al., 2015). Thalidomide binding to cereblon caused opposite effects for the homeobox transcription factor MEIS2, which is implicated in various aspects of human development (Fisher et al., 2014). The modulatory effects of IMiDs on cereblon function point at a key role of the thalidomide-binding domain in cereblon function. This significance is further supported by the sequence conservation of this domain from bacteria to eukaryotes.

The key to understanding protein function is its structure. It provides information about a three-dimensional conformation of the molecule, revealing specific features, which refer to a comprehension of the role of the protein in the cell. Therefore, the first goal was to obtain the structure of the CULT domain as a rational basis for its functional insight.

Expression and purification of cereblon homologs from bacterial and insect cells

Aiming to solve the structure of the CULT domain, we chose representatives from the three main subgroups (Fig. 1.4) of cereblon homologous proteins: human full length CRBN, the secreted *C. elegans* lacking its secretion signal sequence (*CeCRBN*_{sec} Δ 1-15), and the bacterial MsCI4. The bacterial cereblon homolog was successfully expressed in *E. coli* and purified under native conditions (Hartmann et al., 2014). However, the eukaryotic proteins appeared to be insoluble when expressed in *E. coli* and subsequent refolding experiments of the proteins that were purified under denaturing conditions failed. To overcome this problem, we produced domains of human cereblon separately in *E. coli*, but yielded again insoluble proteins for all of them (Fig. 3.4.1.1A and Appendix I).

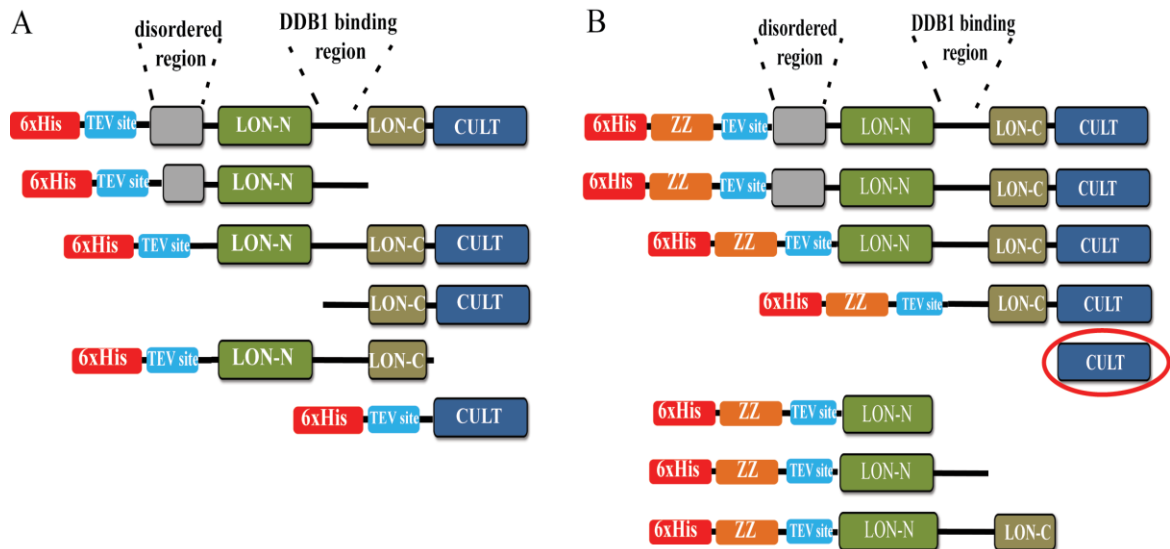


Figure 3.4.1.1. Constructs of human cereblon expressed in *E. coli* (A) and insect cells (B). Successfully purified human CULT domain is circled in red.

Recent studies reported the annelid *Alvinella pompejana*, which inhabits hydrothermal vent sites, to be an attractive metazoan target for structural biology research. Since proteins from thermophiles are stable and more obedient for crystallization experiments, we expressed the LON-containing cereblon homolog from *Alvinella pompejana*, which shows 50 % sequence identity to human CRBN, fused to His- or His-SUMO-tags in *E. coli* strains C41 (DE3) and Lemo 21 (DE3). Compared to other *E. coli* (DE3) strains, protein expression in the Lemo 21 strain is additionally controlled by the natural inhibitor of T7 RNA polymerase – lysozyme, which production is regulated by adding L-rhamnose to the expression medium at levels from 0 to 2 mM. However, expression using these *E. coli* strains did not increase the solubility of the proteins. In a next step, we set up the baculovirus-insect cell expression system, which is more suitable for producing eukaryotic proteins. *CeCRBN_{sec}Δ1-15* containing a C-terminal His-linker was successfully expressed and purified by Ni affinity chromatography from *Sf9* cells. After cleavage of the tag by TEV protease, the protein was again loaded on the affinity chromatography column. The flow through fraction was collected for further concentration of the pure *CeCRBN_{sec}Δ1-15* protein (Fig. 3.4.1.2A). Contrary, *hCRBN* was only poorly expressed and highly sensitive towards proteolytic degradation. In order to rectify these difficulties, another construct was made: the His-tag was fused to the N-terminus and two additional Z-domains (so called ZZ-tag, Gülich et al., 2000) were inserted

between the His-tag and the protein to yield higher expression and solubility (Fig. 3.4.1.1). Additionally, to increase *hCRBN* stability, it was co-expressed and co-purified with its binding partner *hDDB1* and/or thalidomide. Since none of these measures could improve the protein production, we designed a series of truncated constructs of *hCRBN* for expression in insect cells (Fig. 3.4.1.1). Finally, human *CRBN* Δ 1-315, which exclusively consists of the CULT domain, was successfully expressed and purified from insect cells (Fig. 3.4.1.2B) (Boichenko et al., 2016). Detailed information of all constructs designed for cereblon expression trials are listed in the Appendix I.

All three proteins: MsCI4, *CeCRBN*_{sec} Δ 1-15, and the human thalidomide-binding domain *hCRBN* Δ 1-315 were set up for crystallization with thalidomide. The eukaryotic proteins did not yield any crystals, whereas the crystal structures of the bacterial CULT domain from *M. gryphiswaldense* bound to thalidomide and other ligands were solved (Hartmann et al., 2014).

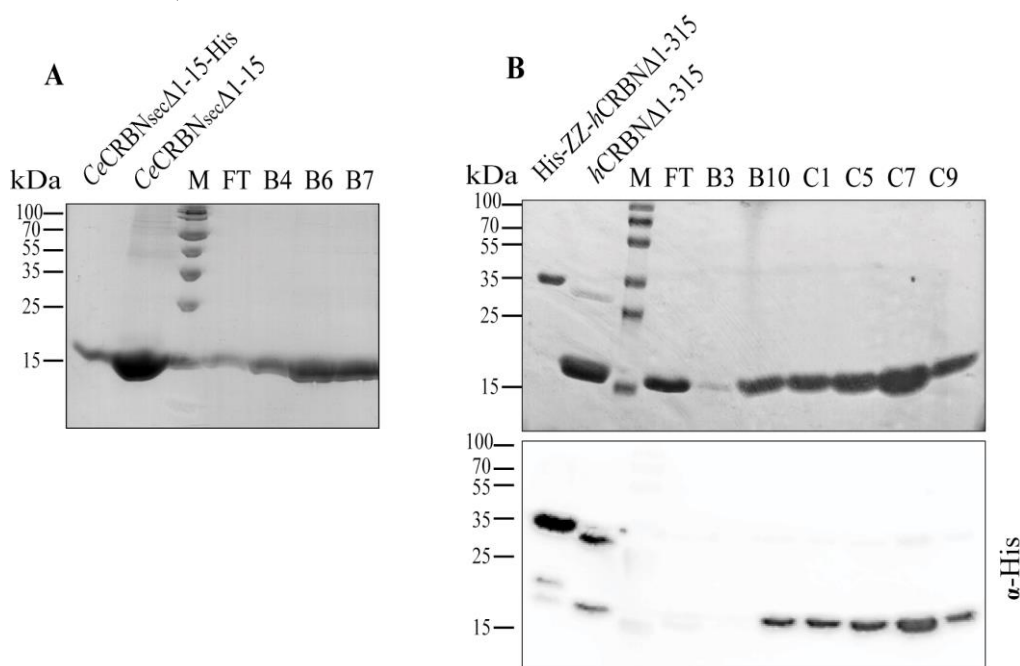


Figure 3.4.1.2. Purification of *CeCRBN*_{sec} Δ 1-15 (A) and *hCRBN* Δ 1-315 (B) by Ni-affinity chromatography after cleavage of the His-tag by TEV protease. A. *CeCRBN*_{sec} Δ 1-15 was found in the FT (flow through). B. *hCRBN* Δ 1-315 was also found in the FT (flow through). As the size of His-ZZ-tag is similar to the size of *hCRBN* Δ 1-315, fractions were also analyzed by Western blotting. M – marker.

At the same time, two other studies published the structures of chicken (Fischer et. al., 2014) and human (Chamberlain et. al., 2014) cereblon in complex with *hDDB1*. Additionally,

the structure of mouse thalidomide-binding domain in complex with thalidomide was deposited to the PDB (PDB-ID 4TZC, deposited by Chamberlain et al).

Taken together, all available structures of CULT domains, from human, chicken, mouse, and bacteria, show a high degree of its conservation. The RMSD over 100 Ca positions between the human and the bacterial protein is less than 1Å (Hartmann et al., 2014, Fig. 5), emphasizing MsCI4 appears to be a proper model system to clarify the function of the CULT domain.

3.4.2 What are the physiological ligands of the CULT domain?

Does the CULT domain recognize uracil in DNA?

The CULT domain is highly conserved from bacteria to human, especially amino acid residues involved in hydrophobic pocket composition. Although recent structural studies show that the substrate specificity towards certain drugs might vary in dependence of single amino acid residues at the entrance of the pocket (Krönke et al., 2015). Notably, thalidomide-binding domain displays the same binding mode for its ligands. Therefore, we suppose that the CULT domain recognizes the same or at least similar ligand(s) in pro- and eukaryotes.

Searching among natural compounds for cereblon ligand, we noticed that pyrimidines resemble glutarimide moiety of thalidomide (Fig. 3.4.2.1). Indeed, uracil, but not cytidine or thymine, was demonstrated to bind the CULT domain of cereblon (Hartmann et. al., 2014). A possible physiological role of uridine binding by cereblon is not known so far. Given the interaction between cereblon and DDB1, a multifaceted protein that is also involved in DNA repair (Scrima et al., 2008), it could be involved in processes of uracil recognition and elimination from damaged DNA. For secreted cereblon proteins, a similar involvement in the detection of viral DNA or RNA in the extracellular space is conceivable.

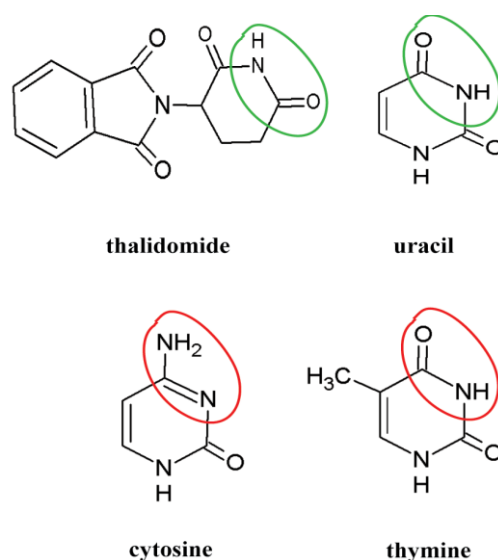


Figure 3.4.2.1. Structural formulas of thalidomide and pyrimidines. The groups, which were supposed to bind the CULT domain, are circled in red; the groups, which bind the CULT domain, are marked in green.

In order to prove these assumptions, a number of experiments were performed. Possible binding of the CULT domain to uracil-containing DNA or RNA was tested in SEC experiment, EMSA assays, and pull downs with biotinylated single- or double-stranded DNA as bait. Tested proteins include MsCI4, *Ce*CRBN and *h*CRBN. The pull down experiments were performed in binding buffer and for the bacterial and the human proteins also in the presence of the appropriate cell extract in order to provide other unknown components that are probably essential for an interaction. However, all performed experiments did not give any hint for an interaction between the CULT domain and nucleic acids.

Does the CULT domain specifically recognize protein modifications?

As already mentioned (see chapter 3.1.1), the binding pocket of the CULT domain resembles an aromatic cage of other, unrelated protein domains such as PHD, chromo, WD40, Tudor, MBT, Ankyrin Repeats, and PWWP. They recognize different methylation states of lysine and arginine residues: unmethylated (me0), mono- (me1), di- (me2), and tri- (me3) methylated for lysine and unmethylated (me0), mono- (me1), dimethylated symmetrically (me2s) and asymmetrically (me2a) for arginine, respectively (Fig. 3.4.2.2 and Fig. 3.4.2.3)

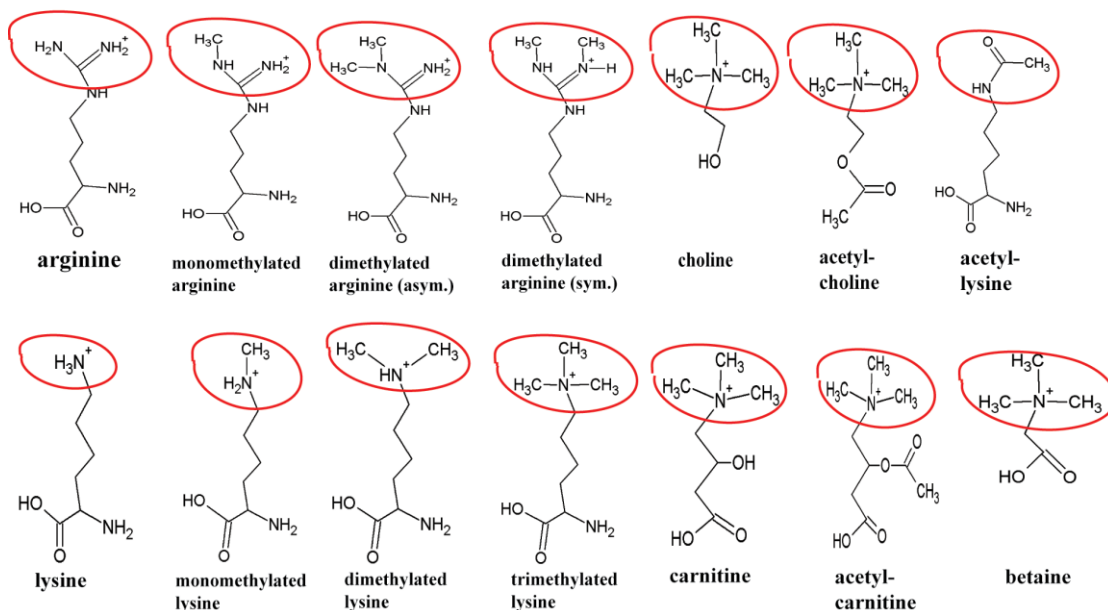


Figure 3.4.2.2. Structural formulas of ligands bound by aromatic cages showing similar architecture to the binding pocket of the CULT domain. The groups, which were expected to bind the CULT domain, are circled in red.

Many of such post-translational modifications are found in histones, representing an important part of the so called “histone code”, which carries epigenetic information extending the genetic message beyond DNA sequences. Histone readers are proteins that decode these messages by recognizing distinct modifications and subsequently initiating downstream signaling processes. Notably, methylation states of lysine residues K4, K9, and K27 of histone 3 (H3K4, H3K9, H3K27) regulate chromatin activity and are coupled with the function of E3 ubiquitin ligase activity. Histone ubiquitination helps in unwinding the nucleosome and allows enzymes to access their substrates (Higa et al., 2006; Lewis et al., 2010; Kuscü et al., 2014; Xu et al., 2010; Zhao et al., 2010).

Baek and co-workers reported methylation-dependent ubiquitination by the Cul4/DDB1/DCAF1 E3 ubiquitin ligase complex (Lee et al., 2012). Based on available structural and functional data, it is conceivable that the recognition of post-translationally modified histone or non-histone proteins by cereblon represents a possible scenario.

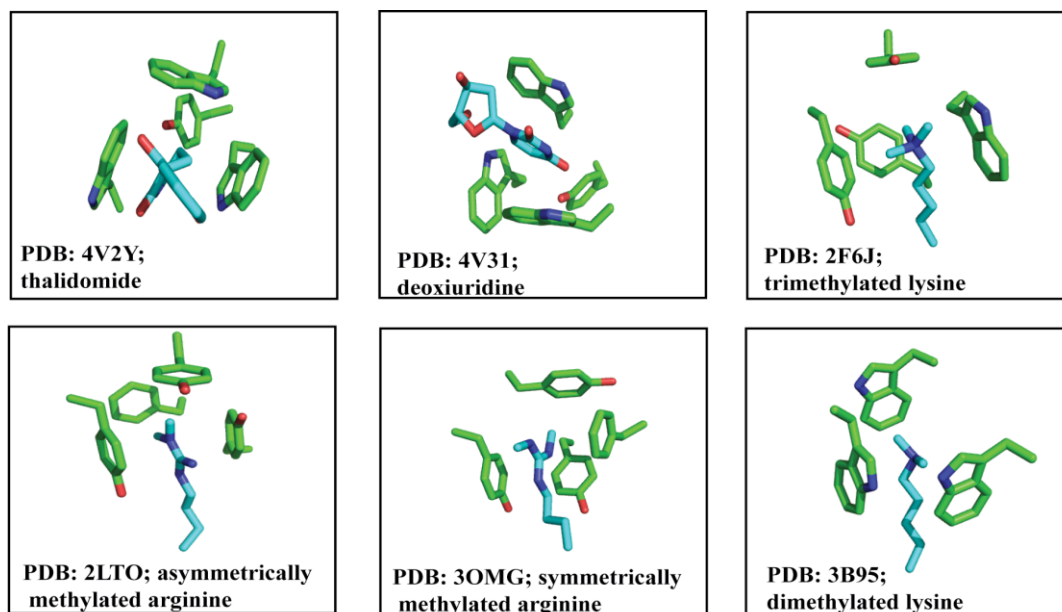


Figure 3.4.2.3. A comparison of the aromatic cages of MscI4 and histone readers. Pocket-forming residues are marked in green and ligands are marked in blue (summarized from Hartmann et al., 2014; Taverna et al., 2007; Yun et al., 2011).

Using the FRET-based assay (Boichenko et al., 2016), we tested the binding of the bacterial and the human CULT domain to the three forms of methylated lysine and arginine. The unmethylated forms served as controls. Since no interaction was detected, we assumed that the context of a linear motif, in which methylation is found, might play an auxiliary role in the recognition. Therefore, we further examined several H3 peptides with different methylation states of K4, K9 or K27 flanked by 4-6 amino acid residues on both sites. The methylation state of these residues was previously shown to be dependent on the DDB1/Cul4A E3 ubiquitin ligase activity (Higa et al., 2006). Nevertheless, no binding of the mentioned histone peptides to the CULT domain could be detected.

Readers of posttranslational modification in proteins often do not only identify one modification, they rather read combinations of multiple modifications, which might also be of different type. Considering this, we expanded our search and started a screen of histone peptide microarrays containing the histone code library with most of today known post-translational modifications of histones and their possible combinations spotted on glass slides. The experiment was comparatively conducted for MscI4, the *C. elegans* variant *CeCRBN_{sec}Δ1-15*, and *hCRBNΔ1-315*, in dependence of thalidomide presence. Despite

knowing that histones do not represent physiological substrates for the secreted *C. elegans* homolog as well as the bacterial MsCI4, we included them in the study. The given conservation of the binding pocket of the CULT domain, which was used as bait in the screen, raised the expectation of binding of a potential ligand to all three proteins. Unfortunately, these analyses remained unsuccessful for several reasons. Data analysis turned out to be difficult for the eukaryotic proteins as the background noise signal appeared to be high, which was apparently caused by unspecific sticking of the proteins to the glass material. This prevented us from drawing final conclusions. In the case of the bacterial protein, we obtained a lot of false positive results, which made the data inconclusive and difficult to interpret. Nonetheless, aliphatic chains with one or more branched methyl groups, such as N ϵ -trimethyllysine, N ϵ -butyryllysine, N ϵ -propionyllysine, N-methylarginine, and asymmetric dimethylarginine, appeared to be generally preferred.

MEIS2, the only endogenous substrate known so far, which binding to cereblon is ablated by thalidomide, contains an N-terminal highly conserved region (Fisher et al., 2014, Lupas et al., 2015). The arginine, histidine, and tyrosine residues in this region appear to be likely candidates for recognition by the CULT domain. However, a test of this hypothesis in our *in vitro* binding assay using a peptide encoding the N-terminal 16 amino acid residues of MEIS, optionally with different methylation states of the arginine residue in position 4, remained unsuccessful.

Does the CULT domain recognize cationic ligands?

In silico analysis further suggested cationic ligands like carnitine, betaine, choline, and their derivatives to bind to the CULT domain (Lupas et al., 2015) (Fig. 3.4.2.1). These compounds, quaternary amines and additionally all proteinogenic amino acids were tested in the FRET-based assay for their possible binding to the CULT domain without any positive results.

The presence of the CULT domain in proteins that are localized inside as well as outside of the cell indicates the availability of the candidate ligand in both environments. Continuing our search for such ligands, we have started collaboration with the screening unit at the Institute of Molecular Pharmacology (AG Kries, FMP, Berlin). Here, the FRET-based assay is applied in a high-throughput screen of low-molecular weight ligand libraries.

3.4.3 Cereblon-mediated teratogenicity of pharmacological compounds

The molecular mechanism of the thalidomide-cereblon interaction became clear after obtaining the crystal structure of their complex. As we previously showed, two hydrogen bonds between the drug and the binding pocket of cereblon were shown to be crucial for this interaction (Hartmann et al., 2014). Further, we assumed that other pharmacological compounds resembling the structure of the glutarimide moiety of thalidomide might also bind cereblon in a similar fashion and, therefore, could be potential teratogens. An extensive search of possible ligands of cereblon was performed based on their structural similarity to glutarimide. We found a set of different pharmaceutical compounds that share common structural features with thalidomide, including succinimide, ethosuximide, hydantoin, 2,4-thiazolidinedione, and barbiturates (Fig 3.4.3.1).

Importantly, these compounds represent basic building blocks of pharmaceuticals used in treatment of various disorders. For example, ethosuximide is the first remedy in absence seizures medication, thiazolidinedione derivatives are prescribed to treat type II diabetes, and barbiturates have sedative effects on the central nervous system. Interestingly, a possible correlation between the structural similarity of some of these compounds to thalidomide and their potential teratogenic properties has already been reported by Gary Shull in 1984, long before cereblon was identified as a molecular target of thalidomide in the cell.

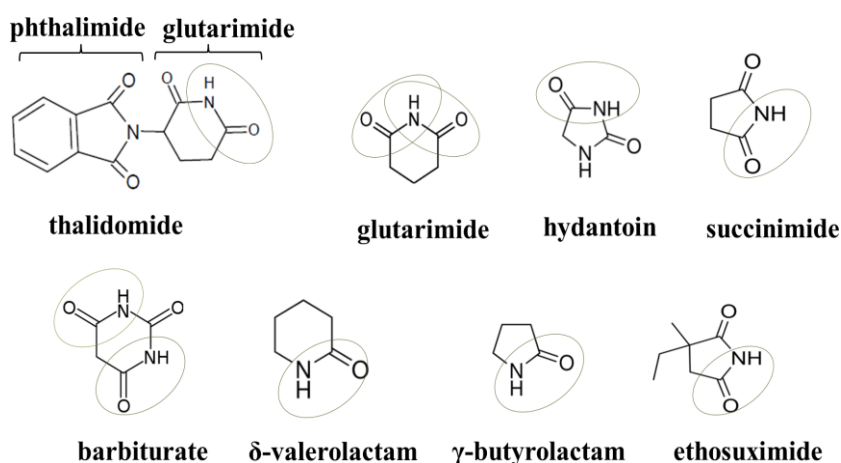


Figure 3.4.3.1. Potential cereblon ligands identified based on their structural similarity to thalidomide.

Upon binding of these compounds to cereblon, E3 ubiquitin ligase function might be affected. We employed our FRET-assay to test the interaction of the CULT domain with the identified compounds. Indeed, with exception of barbiturate, they were shown to bind MsCI4 with affinities in the micromolar range (Table 3.4.3.1). In the structure of the MsCI4-thalidomide complex, only the NH- and one of the neighboring keto groups directly interact with the protein. Based on this, we surmised that the minimal binding moiety in the ring system also interacts with the binding pocket of cereblon. Accordingly, we determined binding constants for γ -butyrolactam and δ -valerolactam composed of 5 and 6 membered rings, respectively (Fig. 3.4.3.1), to be in the micromolar range.

Performing a comparative analysis of the binding affinities of MsCI4 with the human CULT domain, we observed higher affinities of the bacterial protein to its ligands. As shown by further studies using the mutant MsCI4^{Y101F}, this is a result of an additional hydrogen bond that is formed between Y101 of the bacterial protein and the compounds. The human protein does not make this auxiliary contact as it carries phenylalanine in the corresponding position. However, considering that only the CULT domain was analyzed in the assay, the N-terminal domains of human CRBN, which are absent in our construct, are expected to stabilize protein-ligand binding *in vivo* (Hartmann et al., 2015).

Table 3.4.3.1 Binding affinities of the CULT domains from M. gryphiswaldense and H. sapiens to different compounds.

Ligand	MsCI4 ^{WW/FF} IC ₅₀ [μ M]	MsCI4 ^{WW/FF} K _i [μ M]	<i>h</i> CRBN Δ 1-315 IC ₅₀ [μ M]	<i>h</i> CRBN Δ 1-315 K _i [μ M]
1-(N-2-MANT-imidoethyl)-uracil	-	3.3 \pm 0.3 (K _D)	-	7.9 \pm 0.7 (K _D)
Succinimide	7.6 \pm 1.0	4.3	34.5 \pm 4.8	26.2
Glutarimide	48.9 \pm 8.9	27.8	135.8 \pm 19.2	103.1
Hydantoin	76.3 \pm 14.8	43.4	464.3 \pm 124.7	352.7
Ethosuximide	239.1 \pm 20.4	136.0	2616 \pm 455	1987
γ -Butyrolactam	100.6 \pm 19.0	57.2	2700	\approx 2000
δ -Valerolactam	257 \pm 35.7	146.2	no binding	-
Barbiturate	no binding	-	no binding	-

Based on the observed binding of the majority of the tested ligands to cereblon, we assumed that they might have teratogenic effects. To verify this hypothesis, we used an *in vivo* assay in zebrafish embryos, which was previously established by Handa and co-workers to demonstrate cereblon-dependent thalidomide teratogenicity. Dechorinated embryos were grown in the medium containing the appropriate drug for 75 hours post fertilization, and pectoral fin length was subsequently measured and used as a diagnostic criterion of impaired development. As shown in Fig. 3.4.3.2, succinimide, ethosuximide, hydantoin, and glutarimide cause fins malformation in zebrafish. Consistently, the phenotype was rescued by injection of *zcrbn*^{YW/AA} mRNA encoding a binding-deficient mutant of CRBN, but not by injection of wild type *zcrbn* mRNA, which additionally indicates that the teratogenic effect is indeed mediated *via* cereblon.

As expected, barbiturate does not cause fins malformation in zebrafish (Fig. 3.4.3.3), which is consistent with no observed binding to the CULT domain in our *in vitro* test (Table 3.4.3.1). Despite the interaction of the lactams with MsCI4 *in vitro*, these compounds do not impair fins development (Fig. 3.4.3.3), apparently due to their fast decay in the fish-living medium or in the embryos themselves. Notably, in the *in vitro* assay *hCRBN* Δ 1-315 does not show affinities for lactams comparable to MsCI4. The additional hydrogen bond in the bacterial protein might represent a possible explanation for this observation as described above.

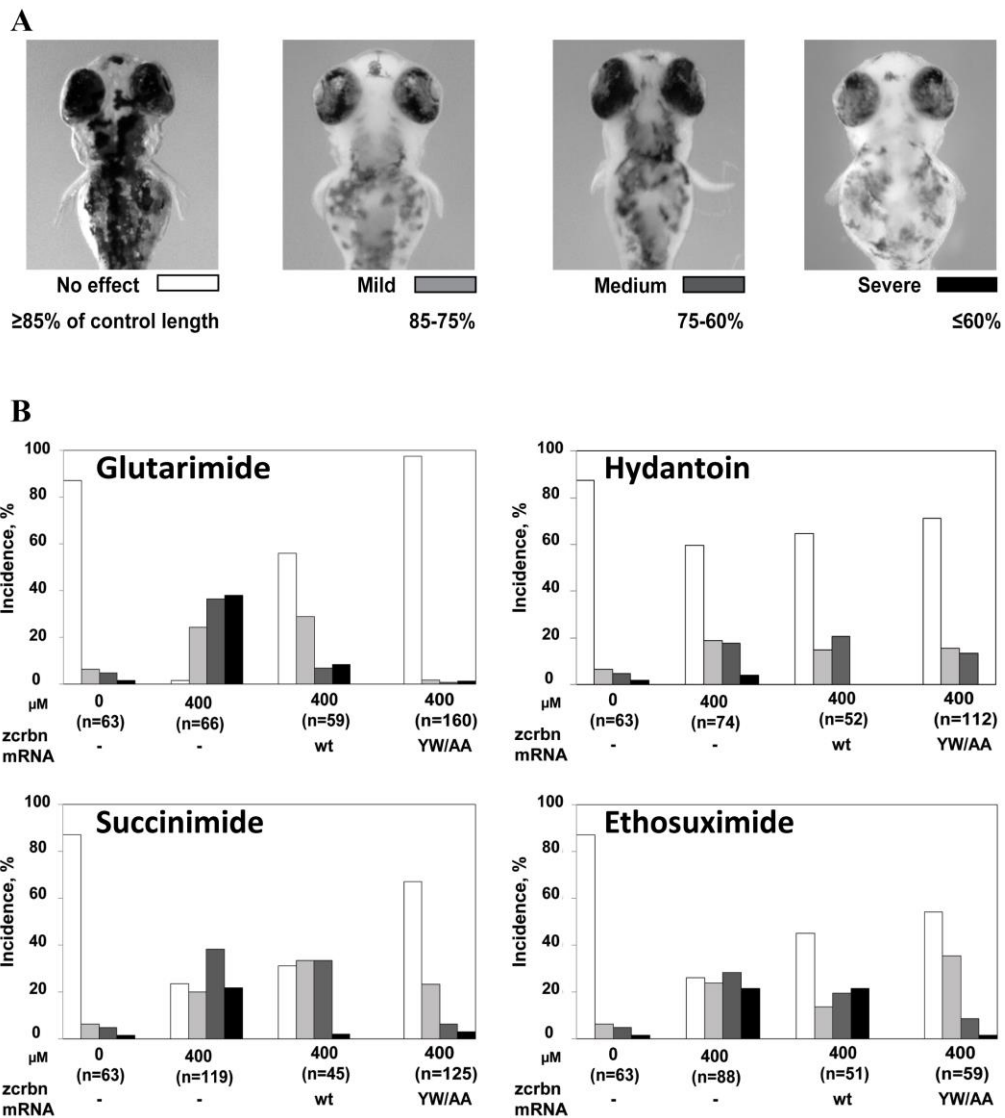


Figure 3.4.3.2. Glutarimide, hydantoin, ethosuximide and succinimide cause developmental defects in zebrafish. A. Typical examples of zebrafish embryos belonging to four phenotypic groups showing different grades of fin malformation, as evaluated from the length of the pectoral fin. B. Incidence of fin deformities under drug treatment. The color of the bars corresponds to A.

Taken together, our findings indicate possible off-target effects of various, clinically relevant pharmaceuticals broadly used in medicine, which become visible as cereblon-mediated developmental defects. The separation of teratogenic and therapeutic properties of ligands mediated *via* their binding to cereblon is an important requirement to guarantee for the safety of these pharmaceuticals. In this context, our *in vitro* assay represents an appropriate

tool for testing the teratogenic potential of such drugs. However, beside these detrimental effects, the specific binding of ligands to cereblon offers a useful tool for triggering targeted protein degradation mediated by the CRBN-DDB1-E3 ubiquitin ligase complex.

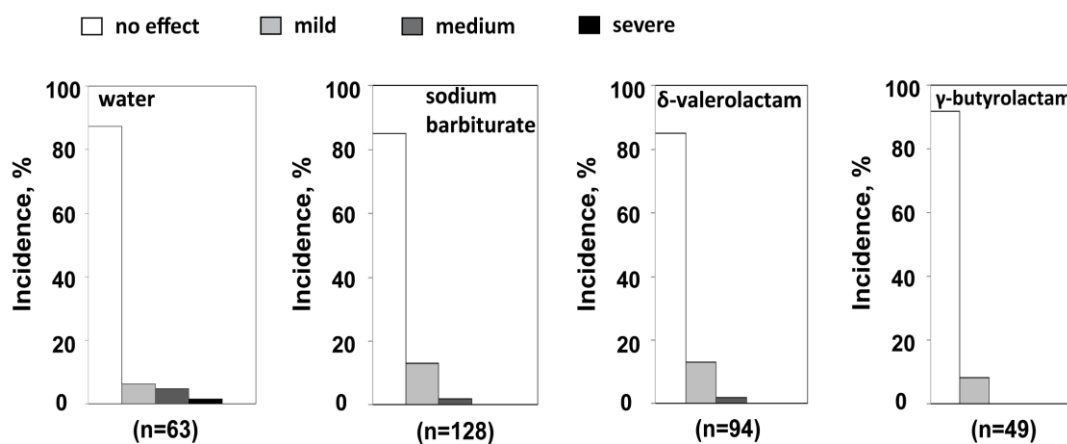


Figure 3.4.3.3. Incidence of fins deformity in zebrafish after treatment with 400 μ M sodium barbiturate, δ -valerolactam or γ -butyrolactam.

3.4.4 Interaction of MsCI4 with transcription termination factor Rho

The previous studies revealing cereblon binding partners were focused on gain of function effects upon drug-treatment. As IMiDs are artificial ligands, the obtained data may reflect an incomplete picture of cereblon action under physiological conditions (Krönke et al., 2014; Lu et al., 2014; Gandhi et al., 2013; Krönke et al., 2015, Fisher et al., 2014).

Since the question about the natural ligand(s) recognized by the CULT domain does not have an explicit answer so far, another approach was pursued in order to uncover cellular processes that CULT domain participates in. We were interested in binding partners of the domain with particular focus on interactions requiring the complicity of the aromatic pocket. Therefore, a candidate approach was used to reveal potential binding partners from a pull down experiment with bait protein and cell lysate. The experiment was performed with the CULT domain of MsCI4 and cell extract of *M. gryphiswaldense*. Mass spectrometry was applied to identify proteins that bound to the bait MsCI4, which was immobilized on magnetic beads. FLAG-tagged MsCI4 and MsCI4^{YW/AA}, a binding-deficient mutant, were

used as baits for comparative analysis. Given its small size, high specificity and minimal interference with protein function, the FLAG tag is frequently used for this purpose. FLAG-MsCI4 was purified to homogeneity (Fig. 3.4.4.1) and bound to the anti-FLAG magnetic beads.

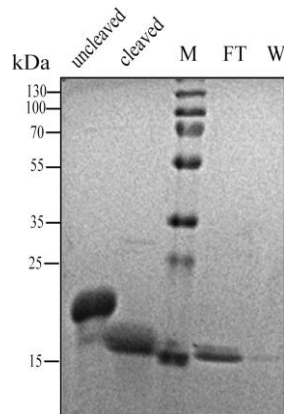


Figure 3.4.4.1. Purification of FLAG-MsCI4 by Ni-affinity chromatography after cleavage of the His-tag by TEV protease. FLAG-MsCI4 was found in the FT (flow through). M – marker.

Subsequently, we incubated the beads with *M. gryphiswaldense* cell extract, which was prepared as a natural source of potential protein binders. The mass spectrometric analysis was conducted to perform quantitative proteomics on the basis of stable labeling of peptides by dimethylation (Boersema et al., 2009). The according analyses were performed at the Proteome Center (Universität Tübingen). In Fig. 3.4.4.2, the practical approach is schematically depicted. In addition to the baits FLAG-MsCI4 and FLAG-MsCI4^{YW/AA}, we used FLAG-MsCI4 in the presence of thalidomide to block the binding pocket. Data comparison of the samples FLAG-MsCI4 and FLAG-MsCI4^{YW/AA} were expected to reveal proteins binding to the aromatic cage of the CULT domain. Mass spectrometric analysis detected numerous proteins, which were pulled down by MsCI4 for each given condition (Fig. 3.4.4.3).

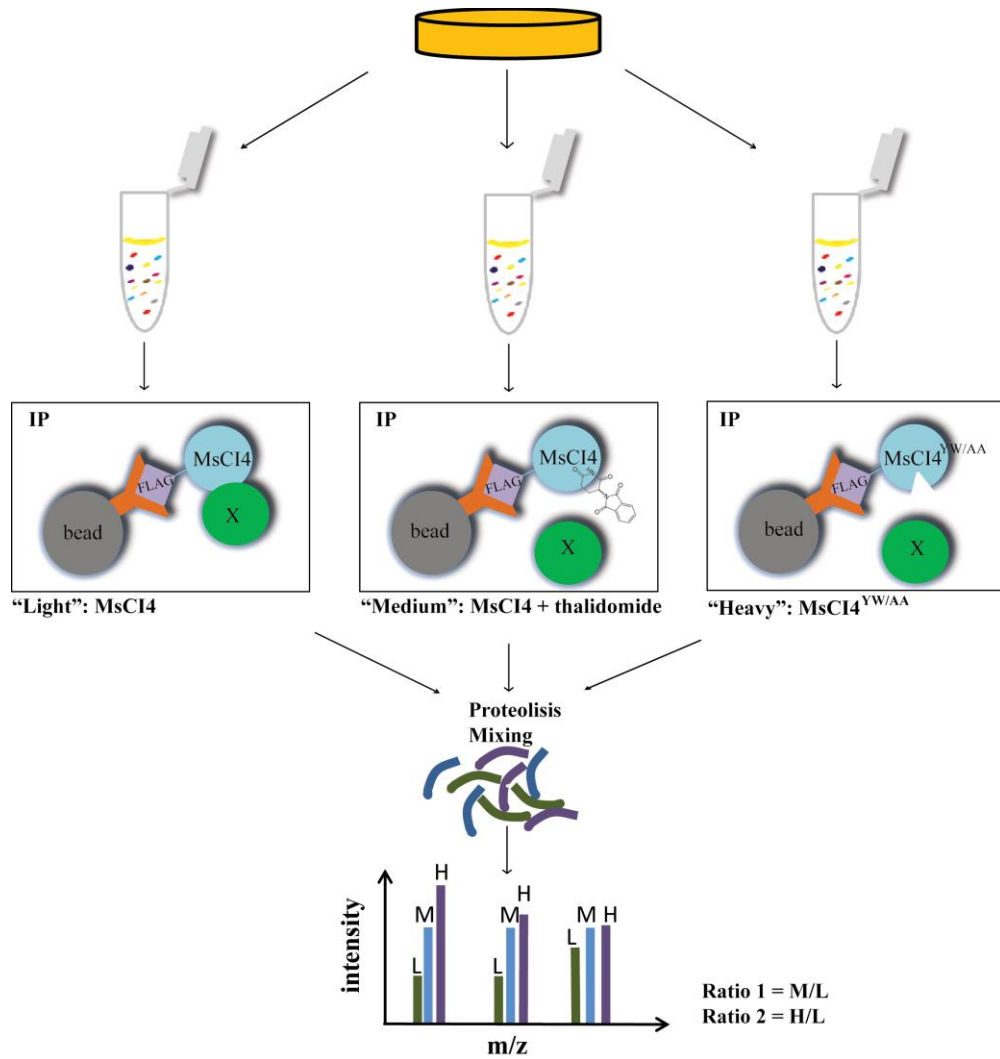


Figure 3.4.4.2 Schematic overview of the immunoprecipitation (IP) experiment to identify binding partners of MsCI4 from *M. gryphiswaldense* cell lysate and following proteomic analysis. H - “heavy” labeled samples, M – “medium” labeled samples, L – “light” labeled samples.

During data analysis, we concentrated our attention on proteins interacting with the pocket. These proteins are expected to be present in the samples with FLAG-MsCI4 as bait, but to be absent in samples containing thalidomide or using FLAG-MsCI4^{YW/AA} as bait. Firstly, we selected candidate binders based on the difference between FLAG-MsCI4 wild type protein and (I) FLAG-MsCI4^{YW/AA}; (II) FLAG-MsCI4 wild type protein sample additionally containing thalidomide. The difference was set to be at least two fold in intensity signal. Thalidomide treatment prevented 9 proteins to interact with FLAG-MsCI4, whereas

usage of FLAG-MsCI4^{YW/AA} mutant obstructed the interaction with 30 proteins in comparison to the wild type protein. Secondly, we excluded from these lists proteins with low sequence coverage as an indicator of low confidence. Thirdly, as the biological function of the pocket was violated in both cases, when thalidomide was present as well as when the bait MsCI4^{YW/AA} was used, we searched for candidate proteins found to be less abundant in both samples. Correspondingly, we short-listed 5 candidates for following studies, which interaction with FLAG-MsCI4 was abolished upon thalidomide treatment or FLAG-MsCI4^{YW/AA} application (Fig. 3.4.4.3 C and D).

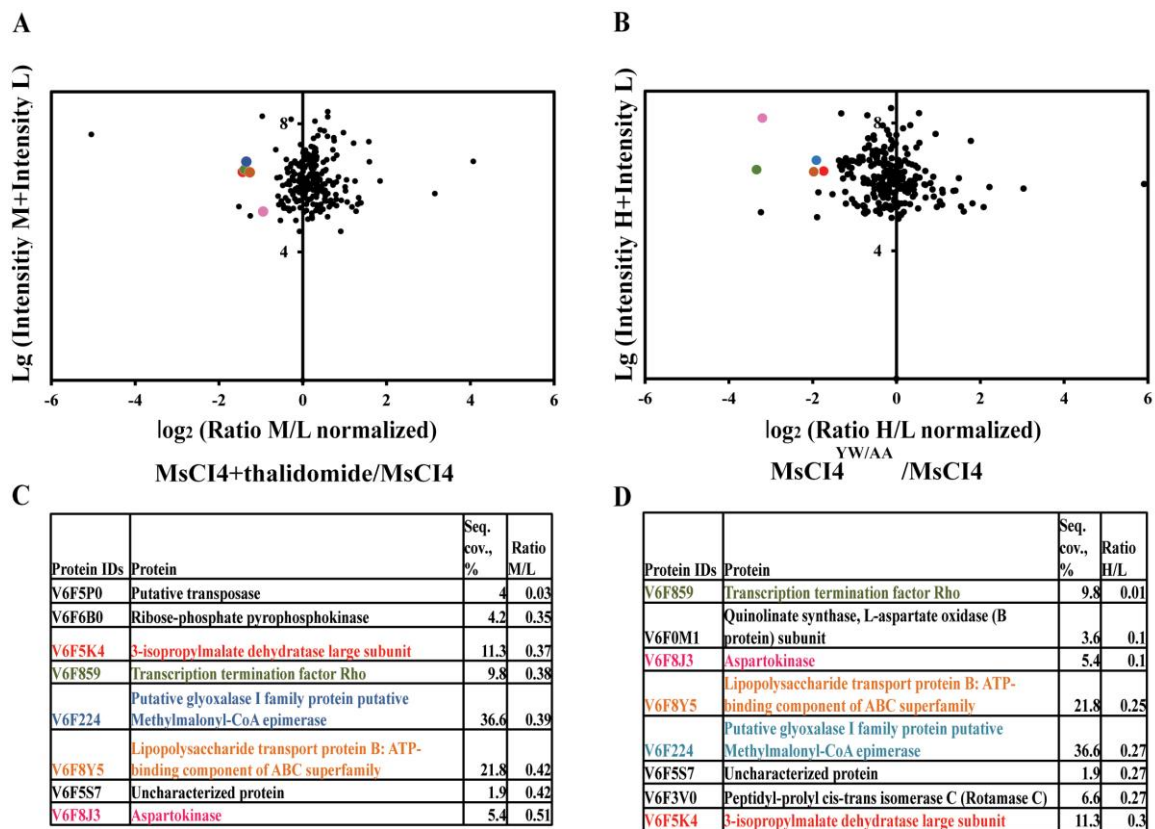


Figure 3.4.4.3. Mass spectrometric analysis revealing potential binding partners of the CULT domain. A, C. Comparison of samples using FLAG-MsCI4 as bait in the presence and in the absence of thalidomide. B, D. Comparison of samples using either FLAG-MsCI4 or mutant FLAG-MsCI4^{YW/AA} as bait. Selected candidates are coloured. Seq. cov. – sequence coverage, H – “heavy”, M – “medium”, L – “light”.

Several proteins, for example putative GGDEF: diguanylate cyclase (GGDEF; V6EVZ1) and putative NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12 (V6EY28), bound to FLAG-MsCI4 in the presence of thalidomide according to mass spectrometric analysis. Surprisingly, they also interact with FLAG-MsCI4^{YW/AA}, which binding pocket is disrupted and, therefore, is not expected to bind its ligands. Considering this discrepancy, we did not select these proteins for further validation in the present study.

The candidate proteins are involved in various cellular processes, including carbohydrate metabolism and amino acid biosynthesis, fatty acid and purine metabolism, regulation of transcription and protection against reactive oxygen species. Comparison of domain composition of the candidate proteins did not reveal any similarities.

Since pull down experiments with cell lysates result in the isolation of large multi-protein complexes, in which many proteins only indirectly interact with the bait (Köcher et al., 2007), additional *in vitro* validation of their binding to the CULT domain was required for all the selected candidates. Therefore, we performed co-immunoprecipitation experiments to confirm possible direct binders.

The selected candidate genes were amplified from *M. gryphiswaldense* DNA and cloned in pETHis1a vector for expression of N-terminally His-tagged proteins in *E. coli* C41 (DE3). Expression of the genes showed that only lipopolysaccharide transport protein B and transcription termination factor Rho were found in the soluble fractions. The expression of the other candidates was further optimized, using the *E. coli* Arctic Express strain, which possesses plasmids encoding chaperones. The chaperones are produced upon exposure to low temperature in order to promote folding and increase the solubility of the overexpressed protein. Nevertheless, the solubility of some of the candidates remained rather poor and required detection by Western blotting. Performed co-immunoprecipitation experiments showed that only the His-tagged transcription termination factor Rho expressed in *E. coli* specifically interacts with FLAG-MsCI4 (Fig. 3.4.4.4). For the other candidates no binding to the bait could be confirmed, suggesting that they were indirectly interacting with the CULT domain, possibly as parts of macromolecular complexes. Alternatively, the absence of a detectable interaction could result from improper folding or low binding affinities under given experimental conditions.

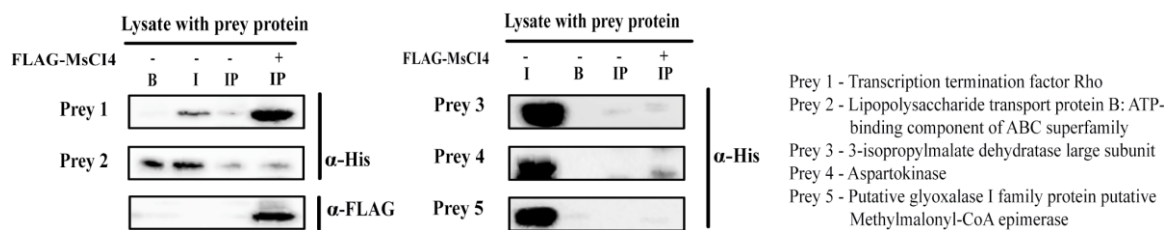


Figure 3.4.4.4. Western blot analysis of potential binding partners of the CULT domain co-immunoprecipitated with FLAG-MsCI4. FLAG-MsCI4 was immunoprecipitated (IP) with anti-FLAG antibodies from the *E. coli* cell extract expressing prey proteins (I). Immunocomplexes were then immunoblotted with anti-FLAG and anti-His antibodies. B - control for unspecific binding of prey protein to the beads.

Next, we tested the interaction between the transcription termination factor Rho and FLAG-MsCI4 in the presence of thalidomide or by using FLAG-CULT^{YW/AA} mutant as bait. Surprisingly, we could not detect any effect as it was expected from the mass spectrometric data that showed decreased abundance of Rho in the presence of thalidomide (Fig. 3.4.4.5). One explanation for this observation might be that Rho additionally binds to amino acid residues located outside of the aromatic cage. Given that both proteins are present in a high excess, even low-affinity binding could be detected by Western blotting. Alternatively, it might be possible, that another unknown molecule, except thalidomide, which is not present in the heterologous *E. coli* expression system in sufficient amount or at all, is required for blocking the MsCI4 interaction with Rho.

Transcription termination factor Rho is a ring-shaped hexameric RNA/DNA helicase, which regulates gene expression in bacteria by unwinding RNA/DNA heteroduplex, using the energy from ATP hydrolysis. The Rho-mediated termination is dependent on the recognition of a specific pyrimidine-rich sequence in the nascent RNA called *rut* (rho ut^{ilization}). Initial RNA contact with Rho promotes a closure of the hexameric ring with RNA in its central channel.

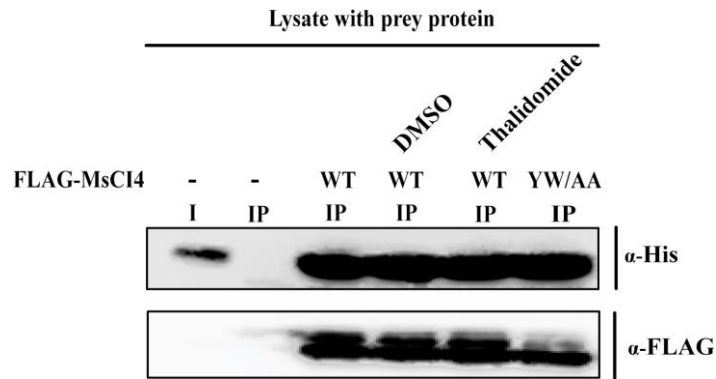


Figure 3.4.4.5. Western blot analysis of transcription termination factor Rho co-immunoprecipitation with FLAG-MsCI4. The FLAG-MsCI4 was immunoprecipitated (IP) with anti-FLAG antibodies from the *E. coli* cell extract (I). Immunocomplexes were immunoblotted with anti-FLAG and anti-His antibodies. WT – wild type FLAG-MsCI, YW/AA - FLAG-MsCI4^{YW/AA} mutant.

This further activates an ATP binding and hydrolysis, allowing translocation of the complex along the RNA (for review see Boudvillain et al., 2013). After reaching the RNA-polymerase Rho stimulates its release from the DNA matrix and dissociation of the mRNA. Additionally, Rho's activity is modulated by riboswitches, leading to a premature termination of the transcripts as a negative regulatory mechanism of their expression (Hollands et al., 2012). Our results suggest a possible involvement of the CULT domain in transcriptional regulation in bacteria. However, further studies have to confirm the specificity of this interaction and to elucidate the functional role of the interplay between Rho and the CULT domain.

3.4.5 Proteomic analysis of CULT-mediated hCRBN binding partners

The effect of thalidomide treatment was studied in embryonic stem cells on the transcriptomics level (Meganathan et al., 2012) and revealed disturbed expression of transcription factors involved in limb, heart, and embryonic development after 14 days of differentiation. The expression pattern of transcription factors required for neuronal development and axon outgrowth was altered, stressing the importance of cereblon in the brain. Additional studies investigating proteomic alterations in response to thalidomide

treatment were conducted on myeloma cells, showing gain of function effects after drug treatment (Krönke et al., 2014). These experiments shed light on the function of cereblon in the specific cell lineages, but provide a limited insight into a general role of cereblon in the cell, which is ubiquitously expressed in all tissues. Although cereblon was identified as a primary target of thalidomide, there is limited knowledge about the consequences of this interaction for the cells and affected biological processes. Therefore, we performed an analysis of proteins co-immunoprecipitating with FLAG-tagged cereblon (FLAG-CRBN) from HEK293T cell extracts in the presence of thalidomide or appropriate DMSO concentrations. The experiment was performed as a SILAC-based comparative proteomic analysis in HEK293T cells representing a powerful tool for the detection of protein-protein interactions. The accurate quantification of protein abundance in both drug-treated and control samples allowed us to distinguish true interactions from experimental contaminants. As CRBN is present in the cytoplasm as well as in the nucleus (Lopez-Girona et al., 2012), whole cellular extract expressing FLAG-CRBN was used for performing a pull down with anti-FLAG magnetic beads (Fig. 3.4.5.1).

The mass spectrometric analysis identified several proteins, the binding of which was gained upon drug treatment. The scatter plot demonstrates \log_2 Heavy/Light ratios for two replicates containing proteins bound to cereblon in the absence and presence of thalidomide, respectively (Fig. 3.4.5.2). Candidate binding proteins, which are color-marked in Fig. 3.4.5.2, were chosen by comparative analysis of the obtained data. Binding of these proteins to cereblon was acquired upon thalidomide treatment. Many of them contain Zn finger motifs, a conspicuous feature of DNA-binding proteins, particularly transcription factors, and are known to be part of multi-protein complexes. Moreover, most of them are involved in transcriptional repression and gene silencing.

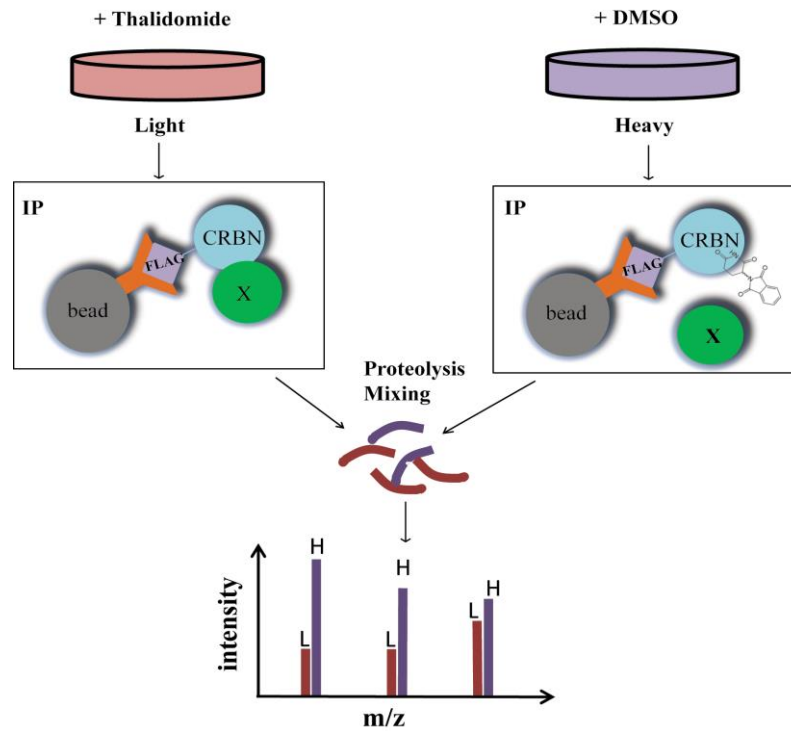


Figure 3.4.5.1. Schematic overview of the immunoprecipitation experiment and following proteomic analysis to identify cereblon binding partners from HEK293T cells treated or non-treated with thalidomide. H – “heavy”, L – “light”.

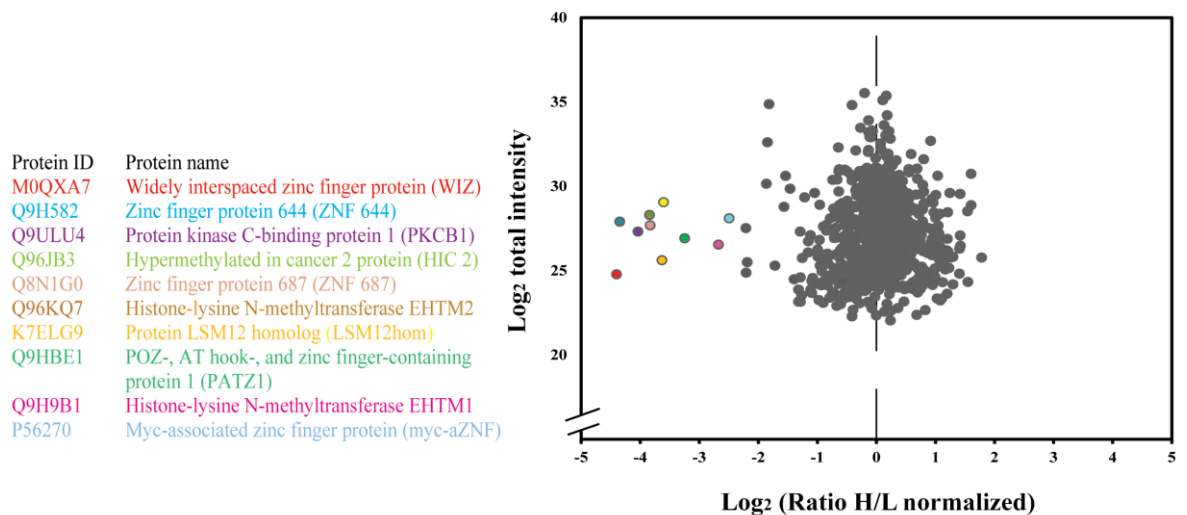


Figure 3.4.5.2. Results of SILAC-based mass spectrometric analysis revealed potential binding partners of cereblon. H – “heavy”, L – “light”.

For validation of binding to cereblon, selected candidates were analyzed further in single co-immunoprecipitation experiments with CRBN as bait. Cereblon was expressed with an N-terminal HA-tag and coupled to anti-HA magnetic beads. Prey constructs (candidates and lacZ as a control) were expressed containing an N-terminal Myc- or FLAG-tag. Due to difficulties in gene amplification on the human cDNA considering their high GC content, not all of the selected candidates could be cloned. The expression of successfully cloned potential binding partners, namely WIZ, PKCB1, HIC2, PATZ, LSM12 homolog, and EHTM1, was tested. With the exception of PKCB1, all the proteins demonstrated expression sufficient to perform the co-immunoprecipitation experiment. The binding experiments showed that none of these candidates directly interacts with cereblon (Fig.3.4.5.3). We assume that these proteins are parts of large multi-protein complexes that contain a direct cereblon target, which still remained unknown to us.

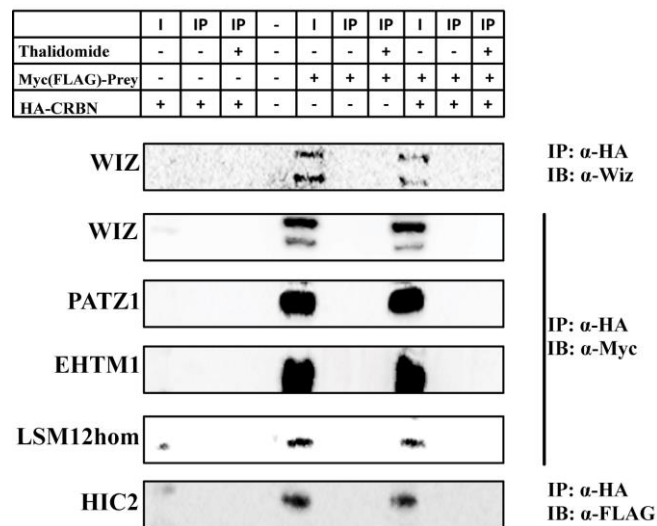


Figure 3.4.5.3. Western blot analysis of co-immunoprecipitation experiments analysing interaction between potential binding partners and cereblon. HA-CRBN coupled to beads served as bait to co-immunoprecipitate Myc- or FLAG-tagged prey proteins from the whole cell extract of transiently transfected HEK293T. (I). Immunocomplexes were then immunoblotted (IB) with anti-FLAG, anti-Myc or anti-WIZ antibodies.

The binding of WIZ protein to cereblon was affected by thalidomide at most as detected by mass spectrometric analysis. WIZ associates with the heteromeric EHTM1/EHTM2 lysine methyltransferases protein complex, which is responsible for mono- and di-methylation of

H3K9, stabilizes and links this complex to CtBP corepressor machinery (Ueda et al., 2006; Kuppuswamy et al., 2008) (Fig. 3.4.5.4). Therefore, we assume that all three proteins were pulled down as a part of macromolecular complex, in which they are only indirectly bound to cereblon.

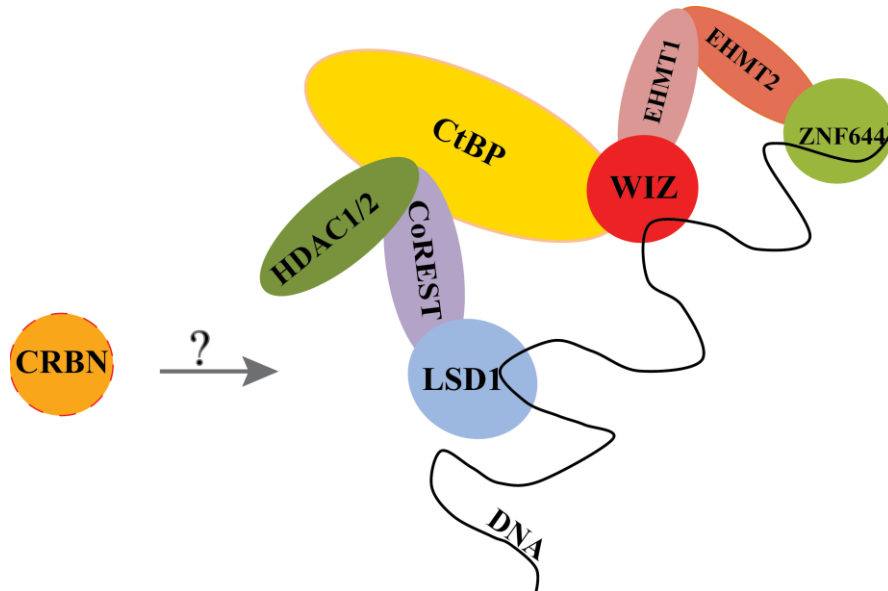


Figure 3.4.5.4. Representation of CtBP corepressor machinery (modified from Kuppuswamy et al., 2007 and Bian et al., 2015)

Two isoforms of the 13 known isoforms of WIZ were identified in HEK293T cells, which migrate as 150 kDa and 125kDa proteins in the gel. Co-immunoprecipitation experiments signified that neither of them binds cereblon directly (Fig. 3.4.5.3.). Additionally, we monitored Myc-WIZ expression in HEK293T cells upon thalidomide treatment and HA-CRBN co-expression. Neither the presence of thalidomide nor HA-CRBN affects Myc-WIZ expression in the cells even upon usage of proteasome inhibitor MG132 (Fig. 3.4.5.5).

In the 2015 study, Yu and colleagues identified that ZNF644 together with WIZ targets the EHTM1/EHTM2 complex to specific gene loci for H3K9 methylation (Bian et al., 2015). Both ZNF644 and WIZ have many zinc finger motifs, which recognize DNA sequences and mediate the recruitment of histone modifying enzymes. The post-translation modifications of histones lead to changes in transcriptional regulation of the gene.

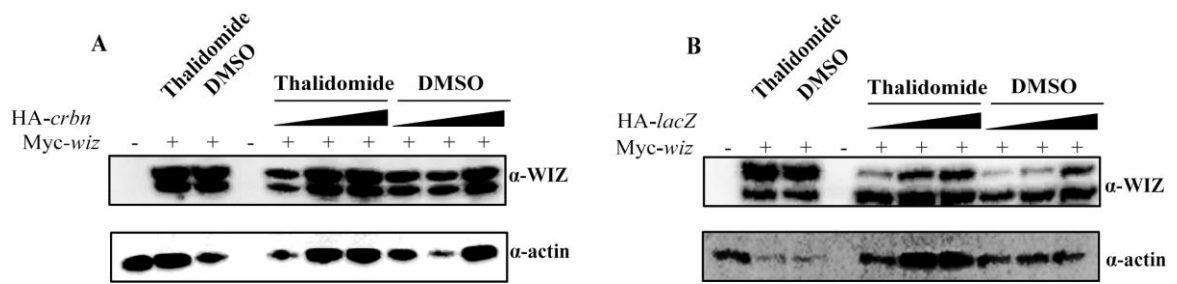


Figure 3.4.5.5. WIZ expression in HEK293T cells is not affected by cereblon co-expression and thalidomide treatment. A. Myc-WIZ expression with gradual increase of HA-crbn_pcDNA3.1 added. B. The same, but HA-lacZ_pcDNA3.1 was used as a control. 50 μM final concentration of thalidomide or 0.05 % DMSO was used. Cells were transfected with 4 μg of Myc-wiz_pcDNA3.1 and 2, 4, or 8 μg of HA-crbn_pcDNA3.1 (or HA-lacZ_pcDNA3.1) in 6-well plates (Greiner Bio-One).

The Cul4A/DDB1 complex associates with several DCAFs, that are components of histone methyltransferase complexes (Higa and Zhang, 2007; Lee and Zhou, 2007; Lewis et. al, 2010; Kuscu et. al., 2014). In addition to that, *ddb1* and *cul4A* gene expression in mammalian cells is essential for trimethylation of H3K4, H3K9, H3K27, thus triggering chromatin transcriptional activity (Higa et. al., 2006). Histone ubiquitination is required for disassembly of the nucleosome particle and the accessibility of its components for the enzymes, which methylate histones or DNA.

The binding of EHTM1/EHTM2 lysine methyltransferases protein complex, which specifically mono- and dimethylates H3K9, a hallmark of inactive chromatin, was also abrogated upon exposure to the drug in our experiments. Taken together, all these observations point at the fact that cereblon is implicated in chromatin organization. Cul4A/DDB1 constitutes numerous E3 ubiquitin ligase complexes modulating different cellular processes, and cereblon, as a part of such a molecular assembly, might play a role in the regulation of transcription. Given the fact that the CULT domain of cereblon possesses the aromatic cage resembling the one from histone readers, it could trigger E3 ubiquitin ligase activity upon recognition of post-translational modification(s) of yet unknown proteins.

In 2010, evidence about the methylation-dependent ubiquitination raised the idea of the existence of a so called “methyl degron” (Lee et. al., 2012). Considering that the histone methyltransferases in addition to the histones also methylate non-histone proteins, CRBN

could alternatively recognize non-histone methylated substrates for their further ubiquitination and degradation.

3.4.6 Cereblon co-immunoprecipitates demethylase LSD1

Detailed studies of the interactome of human methylases and demethylases using yeast two-hybrid system identified cereblon as a binding partner of SUV39H1 and LSD1 (KDM1A) (Weimann et al., 2013). The histone methyltransferase SUV39H1 specifically trimethylates H3K9, using monomethylated H3K9 as a substrate. This process initiates heterochromatin formation and triggers epigenetic transcriptional repression. Histone demethylase LSD1 demethylates mono- and dimethylated H3K4 and H3K9, thereby, depending on the context, acting either as a coactivator or a corepressor. The corepressor function is mediated by demethylation of H3K4me, a specific tag for epigenetic transcriptional activation, whereas H3K9me demethylation leads to an activation of transcription. This attracted our attention since the CULT domain of cereblon has a very peculiar aromatic cage, which resembles the ones from histone readers that recognize different kinds of arginine or lysine methylation states (mono-, di-, trimethylation) (Fig. 3.4.2.3). Given that the cereblon interacting protein DDB1 is an important participant in DNA repair as well as hetero- and euchromatin transition, cereblon might be involved in histone recognition. Additionally, LSD1 was found as cereblon binding partner in our pilot mass spectrometric analysis with a weak intensity signal, which might result from a low abundance of the LSD1 in the cell.

We performed co-immunoprecipitation experiments in HEK293T cells to demonstrate that cereblon binds LSD1 and SUV39H1. Indeed, FLAG-LSD1, but not SUV39H1, was confirmed to interact with HA-fused cereblon (Fig. 3.4.6.1). This binding is invulnerable to thalidomide treatment, suggesting that the LON domain or N-terminal unstructured region of cereblon mediate this interaction.

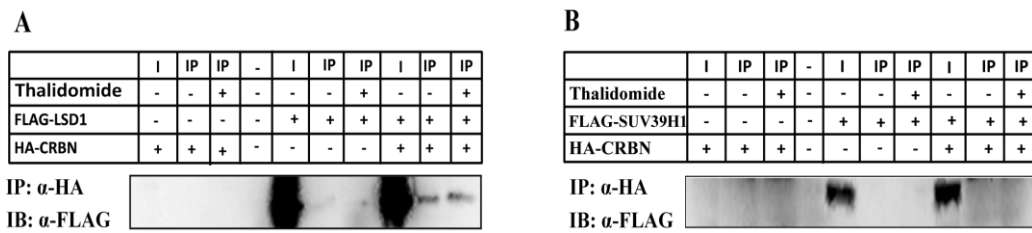


Figure 3.4.6.1. Binding of cereblon to LSD1 and SUV39H1. HA-CRBN co-immunoprecipitates FLAG-LSD1 (A) from HEK293T cells, but not FLAG-SUV39H1 (B). Cereblon was immunoprecipitated (IP) with anti-HA antibodies from the whole cell extract (I). Immunocomplexes were then immunoblotted (IB) with anti-FLAG antibodies.

To identify the regions involved in the binding of these two molecules, further co-immunoprecipitation experiments were designed, using differently truncated constructs of CRBN and LSD1 (Fig. 3.4.6.2). Overexpression trials of truncated cereblon constructs in HEK293T and HELA cells failed. Consistently, we observed generally lower expression levels of full length cereblon in comparison to other proteins, which we expressed in human cells. This might be the result of improper folding and following degradation under the condition of overexpression, when the capacity of the cell to cope with high amounts of the protein is limited. LSD1 comprises an N-terminal unstructured region, the SWIRM domain, a common module for protein-protein and protein-DNA interaction, and an Amine Oxidase domain, which possesses the FAD- and substrate binding regions in the catalytic center of the enzyme. Additionally, the Amine Oxidase domain contains an insertion of about 100 amino acid residues building the Tower domain that interacts with the transcriptional repressor CoREST and links the demethylase activity to gene silencing. The co-immunoprecipitation experiment with cereblon as bait was conducted for differently truncated LSD1 constructs and revealed that the SWIRM and Amine Oxidase domains are required for interaction.

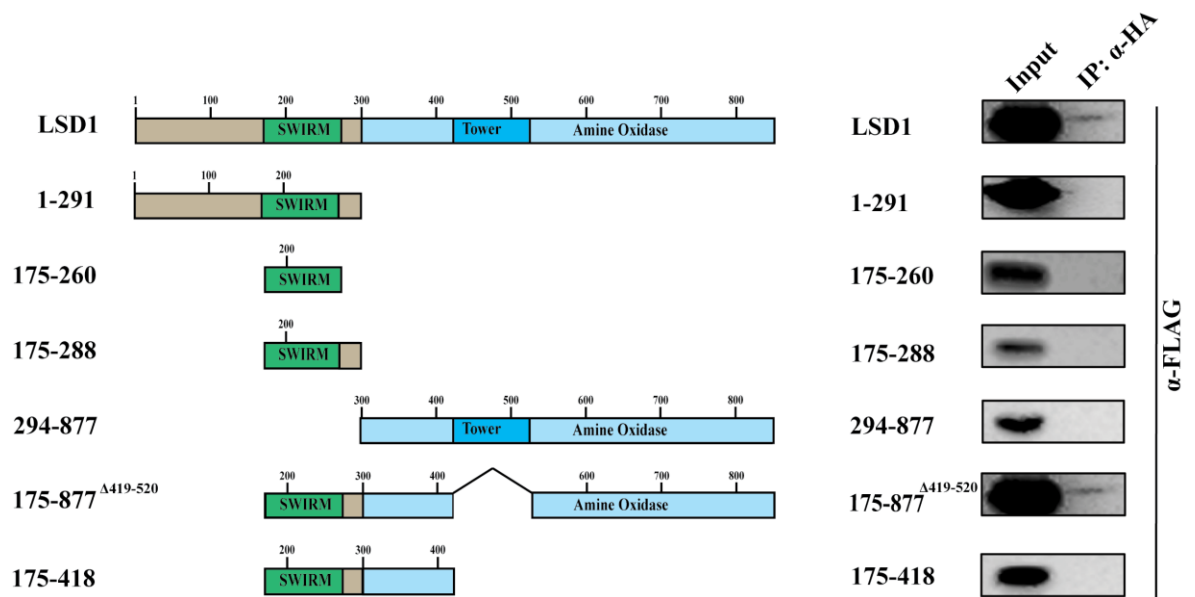


Figure 3.4.6.2. Mapping the region of LSD1 required for the interaction with cereblon. Cereblon was immunoprecipitated (IP) with anti-HA antibodies from whole extract of HEK293T cells mixed with equal amount of the whole extract of HEK293T expressing different FLAG-tagged LSD1 constructs. Immunocomplexes were then analyzed by western blotting with anti-FLAG antibodies.

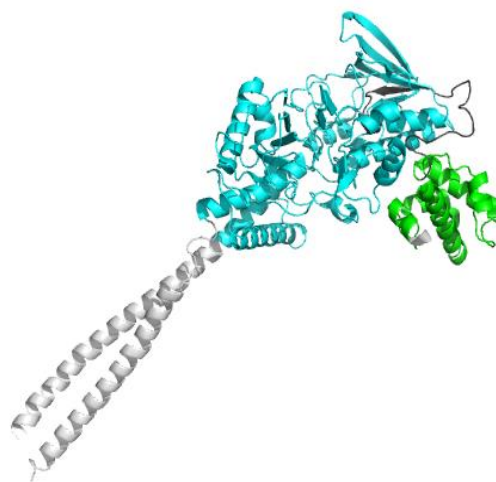


Figure 3.4.6.3. LSD1 crystal structure with the SWIRM marked in green, the Amine Oxidase domain marked in cyan, and the Tower domain marked in light grey. The Tower domain is not necessary for the interaction with cereblon (modified from Chen et al., 2006)

The SWIRM domain alone is not sufficient to preserve the interaction, even if it is followed by the N-terminal part of the Amino Oxidase domain. The Amino Oxidase domain has a large insertion, the Tower domain, which is indispensable for the binding to the corepressor and catalytic activity, but not for interaction with cereblon. After building this module the C-terminal part of the molecule flips back to complement the functional catalytic domain (Fig. 3.4.6.3). Apparently, the C-terminal part of LSD1 either has an interaction interface with cereblon or is required for proper folding of other regions. Thus, we conclude that the binding interface of LSD1 is built by the SWIRM and Amino Oxidase domains, excluding the embedded Tower domain.

Additionally, we monitored FLAG-LSD1 expression in HEK293T cells upon thalidomide treatment and / or HA-CRBN co-expression. Neither thalidomide nor HA-CRBN co-production affected the FLAG-LSD1 expression in the cells even upon usage of proteasome inhibitor MG132 (Fig. 3.4.6.4). Cereblon expression slightly impairs the production of LSD1. Since we do not observe a reduction of the anti-FLAG signal in dependence on the increasing amounts of HA-*crbn* DNA added, this effect appears likely to result from the increased amounts of plasmid transfected into the cells rather than the cereblon production itself (Fig. 3.4.6.4A). The absence of any effect in the presence of thalidomide let us suppose that the LON domain and / or the N-terminal unstructured region of cereblon, but not the thalidomide-binding domain, interact with LSD1.

The role of LSD1-CRBN binding still remains poorly understood and we envisage several possible scenarios of the biological function of this protein tandem. Firstly, LSD1 demethylates not only histones, but also other proteins. Therefore, it could work on cereblon as a substrate that is demethylated. However, post-translational modification of cereblon by methylation has not been reported yet. Secondly, cereblon might target the demethylase for ubiquitination and subsequent degradation. Thirdly, given that cereblon has an aromatic pocket similar to readers of lysine and arginine methylation, it might share its substrate proteins with LSD1. It is conceivable that both proteins are part of a bigger protein complex facilitating ubiquitination-dependent protein degradation triggered by the methylation state of its substrates.

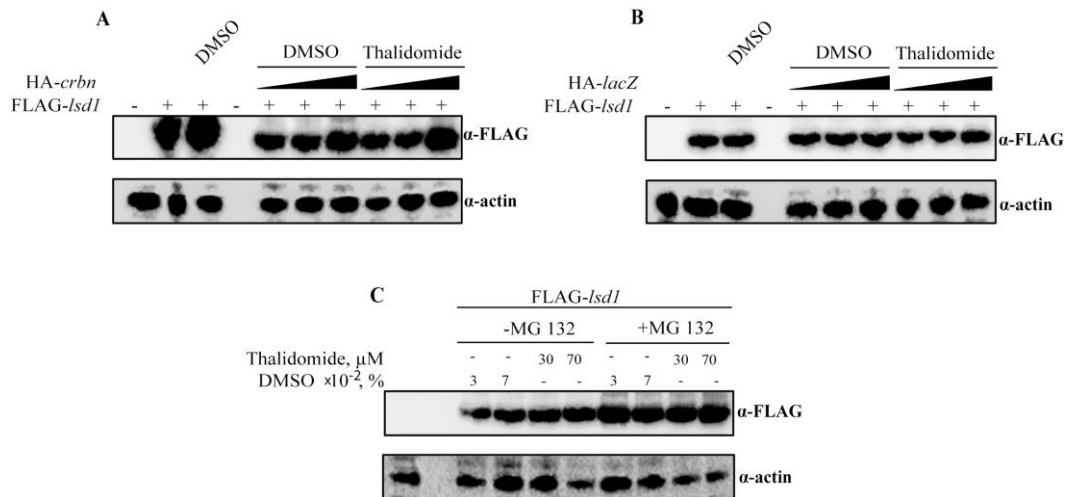


Figure 3.4.6.4. Expression of LSD1 in HEK293T cells is not affected by cereblon co-expression or thalidomide treatment. A. FLAG-LSD1 expression with gradual increase of HA-*crbn*_pcDNA3.1 co-transfected. B. The same, but HA-*lacZ*_pcDNA3.1 was co-transfected as a control. C. FLAG-LSD1 expression in the presence of thalidomide with and without proteasome inhibitor MG 132. 50 μM final concentration of thalidomide or 0.05 % DMSO was used unless otherwise specified. Cells were transfected with 4 μg of FLAG-*lsd1*_pcDNA3.1 and 2, 4 or 8 μg of HA-*crbn*_pcDNA3.1 (or HA-*lacZ*_pcDNA3.1) in 6-well plates (Greiner Bio-One).

Thalidomide and its derivatives increase the level of p21^{WAF-1} expression in myeloma cells. Verhelle and co-workers showed that p21^{WAF-1} upregulation is associated with the alteration from methylated to acetylated histone H3K9 on the p21^{WAF-1} promoter (Escoubet-Lozach et al., 2009). Interestingly, *lsd1* knock down reduces lenalidomide- and pomalidomide-dependent upregulation of p21^{WAF-1}, suggesting that this histone demethylase is involved in the regulation of p21^{WAF-1} promoter. Lenalidomide and pomalidomide treatment was hypothesized to activate LSD1, which removes the H3K9me2 chromatin silencing mark, subsequently allowing the acetylation of this residue and activation of transcription. The post-translational modifications of histones require the E3 ubiquitin ligase machinery for unwinding the nucleosome and permitting entry of enzymes to their substrates (Higa et al., 2006; Zhao et al., 2010; Xu et al., 2010; Kuscus et al., 2014). Therefore, we surmise a possible function of cereblon in the recognition of protein substrates, the methylation state of which is modulated by LSD1. LSD1 is involved in many different

processes having pro- or anti-proliferative effects. As a potential oncogene it demethylates and inactivates p53 (Huang et al., 2007). In contrast, it also inhibits TGF- β 1 expression, leading to tumor suppression in breast cancer (Wang et al., 2009). Furthermore, being a part of the LSD1/CoREST/HDAC complex, LSD1 regulates expression of key regulators (including limb growth controllers), during early embryonic development (Foster et al., 2010; Wang J. et al., 2007; Wang Y. et al., 2009). Together with corepressor (CoREST) and histone deacetylase enzyme (HDAC), LSD1 forms a large multi-protein complex, which also includes the carboxy-terminal binding protein (CtBP). CtBP works as an assembly factor for this complex, to which the sequence-specific DNA-binding repressors, LSD1, histone deacetylases (HDAC1/2), histone lysine methyltransferases (EHTM1 and EHTM2), and corepressors bind (Kuppuswamy et al., 2008) (Fig. 3.4.5.4). Initial deacetylation of histones by HDAC1/2 provides access for CoREST and LSD1. Subsequently, LSD1 demethylates H3K4me2, which further allows EHTM1/EHTM2 to methylate H3K9, a stable repression mark. Such a complex is recruited by a number of transcription factors to operate gene silencing of various genes (Foster et al., 2010; Yokoyama et al., 2014). As a binding partner of LSD1, cereblon might, depending on the context of the multi-protein complex, bridge either repressor or activator and the E3 ubiquitin ligase machinery. The possibility of cereblon function in transcriptional regulation is also supported by our data obtained in the SILAC-based proteome analysis of cereblon binding partners, which identified EHTM1, EHTM2 and WIZ to co-immunoprecipitate with cereblon. Different tissues have different expression pattern of their transcription factors at any particular time point, which require methylase and demethylase activities for their proper functioning. In this context, the linkage of gene regulation and ubiquitination might also contribute to the explanation of tissue-specific teratogenic effects of thalidomide, which binds to cereblon and inhibits ubiquitin ligase activity.

3.4.7 Investigation on the sub-cellular localization of the secreted *C. elegans* cereblon protein

Resolving the sub-cellular localization of a protein is an essential step towards understanding its function. Animal proteins comprising CULT domain preceded by N-terminal secretion signal sequence have not been studied yet, despite growing interest in their

role outside of the cell. Initial bioinformatic analysis showed that these proteins are predicted to function extracellularly (Lupas et al., 2015). Indeed, the homolog from the sand fly *Phlebotomus arabicus* has been found in salivary gland, using MS (Hostomska et. al., 2009). In this chapter, I briefly describe our efforts towards revealing the sub-cellular localization of *C. elegans* secreted cereblon protein, which we expected to help us to shed light on the biological role of CULT domain outside as well as inside of the cell.

In *C. elegans*, two main approaches are generally used to determine the sub-cellular localization of a protein: expression of GFP or other fusion constructs or immunocytochemistry, both of which have their prerogatives and limitations (Duerr., 2006; Wormbook, The *C. elegans* Research Community, www.wormbook.org). Gene expression pattern in *C. elegans* is frequently determined, using a fusion of the promoter of the studied gene to a reporter gene such as *lacZ* or *gfp*. This provides information about spatial and temporal expression of the gene in the whole organism. Beyond that, further developments of this method allow even the determination of the sub-cellular localization of a protein, if the reporter gene is fused directly to the full coding sequence of the gene of interest. Therefore, we employed the reporter gene fusion method to find out a sub-cellular localization of *CeCRBN_{sec}*. Initially, a construct containing *Cecrbn_{sec}* gene, annotated as R08B4.3 in Wormbase (<https://www.wormbase.org>), with its own 5'- and 3'-UTR regions and the *gfp* gene inserted in the reading frame was generated by fusion PCR. The fusion product was validated by sequencing and further injected into gonads of the adult worms. The transgenes were successfully generated as we observed the phenotype caused by marker gene expression restraining worm movement to a circle trajectory (Fig. 3.4.7.1). Next, transgenic worms were crossed to obtain the second generation of transgenic animals to increase the stability of foreign DNA expression. However, fluorescent microscopy did not yield any GFP signal in the second as well as the in first progeny. We hypothesize that R08B4.3 gene expression might require special conditions for its initiation, which are not known to us at present.

Considering that any information about the spatial and temporal expression of *Cecrbn_{sec}* within the organism won't be achieved using this method, we continued with constitutive promoters for fusion constructs. Two different constitutively expressed genes were chosen: *ges-1* (intestinal) and *myo-3* (muscle). The DNA constructs were designed similarly to the previous one with the exception of 5'-UTR region, which was substituted by *ges-1* or *myo-3* promoter regions to make *Cecrbn_{sec}* constitutively expressed in the intestine or muscles, respectively (Fig. 3.4.7.1). Transgenes were prepared as described above. Again, microscopy

analysis did not reveal any GFP signal in positive transgenes, which indicated that *Cecrbn_{sec}* is not expressed at the required level or even at all. Previously, abnormal regulation of expression from multi-copy transgenic arrays was reported as one of the disadvantages of the technique (Mello and Fire, 1995), and we surmise that it explains the absence of the signal in our experiments.

The use of antibodies to visualize the distribution and sub-cellular localization of gene products powerfully complements molecular analysis of a protein function. Thus, we further decided to employ immunohistochemistry as an alternative technique for determination of the sub-cellular localization of the protein. To produce specific antibodies against our target protein, we used the *CeCRBN_{sec}* expressed and purified from *E. coli* (Fig. 3.4.7.2) as an antigen for rabbit immunization. The antibodies were purified from the rabbit serum, using antigen covalently coupled to an NHS-activated column. The working concentration of the antibodies was determined to be 1:1,000.

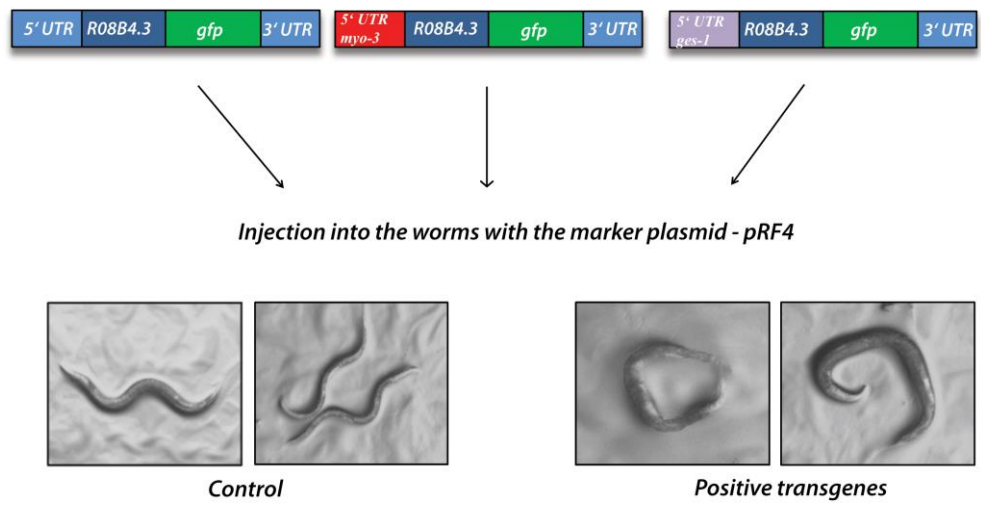


Figure 3.4.7.1. Reporter gene constructs injected into worms with marker plasmid generate positive transgenes. UTR – untranslated region.

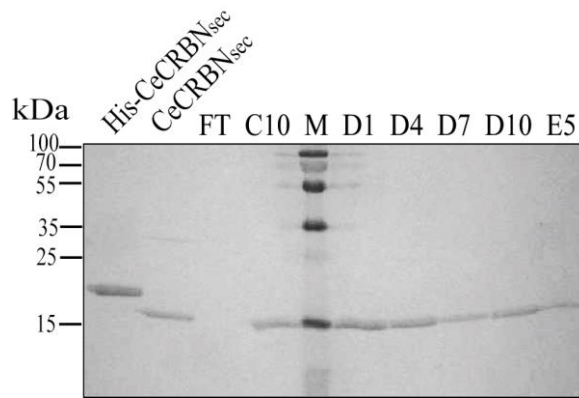


Figure 3.4.7.2. Purification of *CeCRBN_{sec}* by Ni-affinity chromatography after cleavage of the His-tag by TEV protease. Fractions D4-E5 were pooled and used for immunization. FT (flow through), M – marker.

As any information concerning either topological or stage-dependent expression of *CeCRBN_{sec}* is missing, we first aimed to determine, at which developmental stage it is expressed in worms (Fig. 3.4.7.3A and B). Worms of different stages were separately collected after synchronization. Total protein concentration was measured in each worm extract (Pierce BSA kit). 30µg of each sample was loaded on the SDS-gel and further analyzed by Western blot using generated antibodies (Fig. 3.4.7.3 B). Fig. 3.4.7.3 D represents a sensitivity test, which demonstrated that as little as 10 ng of *CeCRBN_{sec}* can be detected under given experimental conditions. Since the *CeCRBN_{sec}* expression was not observed in the protein extracts under normal physiological conditions, we surmise that it is dependent on some unknown environmental queues, which are, considering the missing knowledge about the protein function, difficult to predict. Alternatively, *CeCRBN_{sec}* might be produced in a very low amount in the organism and, therefore, its detection is limited for many methods.

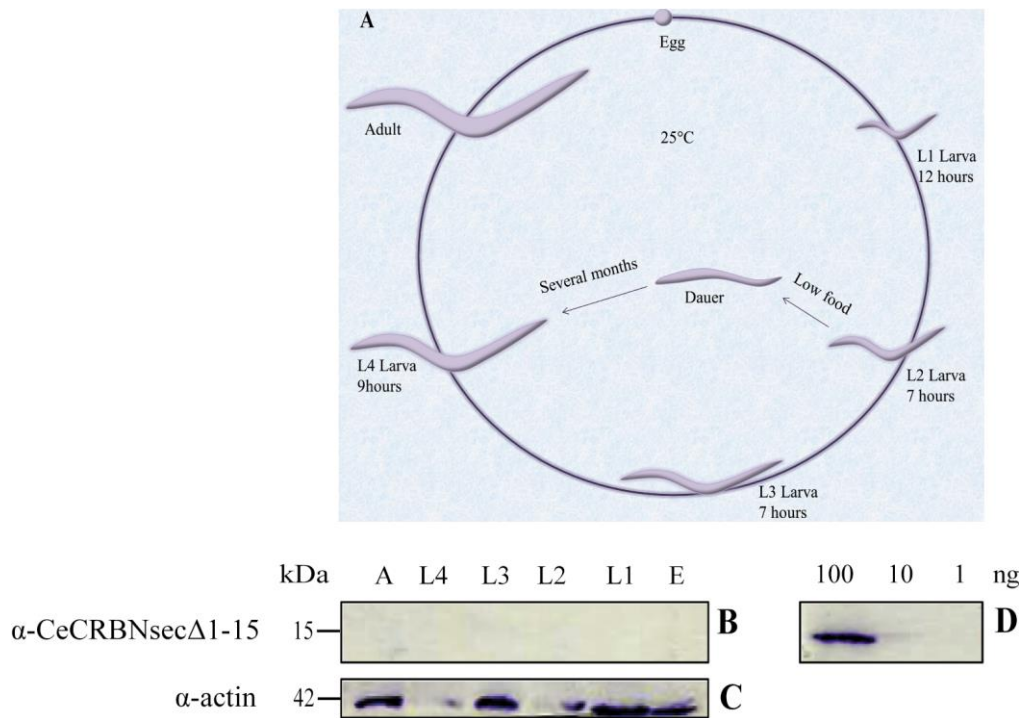


Figure 3.4.7.3. *CeCRBN_{sec}* expression in *C. elegans*. A. Worm life cycle. B. Western blot analysis of different worm stages using the anti *CeCRBN_{sec}* antibodies. C. Loading control. D. Western blot demonstrating the sensitivity of the produced anti *CeCRBN_{sec}* antibodies against recombinant *CeCRBN_{sec}*.

3.4.8 Contributions

This PhD thesis was conducted in the department of Prof. Andrei Lupas, who initiated the project, performed bioinformatic analysis, supervised the study, and continuously discussed it with our team members. I performed all the experiments under intermediate supervision of Birte Hernandez Alvarez, who designed the experimental work together with Marcus Hartmann unless otherwise specified. In this chapter, my contributions to the experiments are listed, which have not been published yet. My contributions to the publications, presented in this thesis, are described before (3.1.2, 3.2.2, 3.3.2).

3.4.1 CULT domain: structural insight

As mentioned in 3.1.2, I performed cloning, expression, and purification trials for all presented constructs, except the bacterial ones. The three construct of N-terminally truncated

human cereblon were designed by Birte Hernandez Alvarez and cloned by Silvia Deiss into pFAST-Bac vector. I performed expression studies in insect cells and purified the proteins.

3.4.2 What are the physiological ligands of CULT?

The DNA / RNA binding studies using SEC were initiated by Birte Hernandez Alvarez and continued by myself after joining the department. I further performed all EMSA and pull down experiments with biotinylated DNA, biotinylated peptides and microarrays. Klaus Kopec helped me with the microarray analysis. Silvia Deiss provided technical assistance in the conduction of pull downs with biotinylated peptides. The FRET-based ligand binding assay was established by Birte Hernandez Alvarez and some of the measurements were performed by Kerstin Bär.

3.4.3 Cereblon-mediated teratogenicity of pharmacological compounds

Marcus Hartmann initiated the study of cereblon binding to pharmaceuticals sharing structural similarity with thalidomide. The FRET-based ligand binding assay was established by Birte Hernandez Alvarez and some of the measurements were performed by Kerstin Bär. I established the *in vivo* assay in zebrafish, performed all experiments and analyzed the data. Silvia Deiss helped me to mount the embryos in the agarose before imaging the fish.

3.4.4 The bacterial CULT domain interacts with transcription termination factor Rho

Birte Hernandez Alvarez initiated and designed the study. I contributed to the experimental design of the experiments for validation of identified binding proteins. I performed all experiments described in this chapter. Mass spectrometric analysis was conducted at the Proteome Center (Universität Tübingen).

3.4.5 Proteomic analysis of CULT-mediated *h*CRBN binding partners

Birte Hernandez initiated and designed the study, performed pull down for SILAC-based proteomic analysis, and analyzed the data. I cloned and expressed binding candidates and performed co-immunoprecipitation experiments. Mass spectrometric analysis was conducted at the Proteome Center (Universität Tübingen).

3.4.6 Cereblon co-immunoprecipitates demethylase LSD1

I initiated the study; together with Birte Hernandez Alvarez I designed the experiments. All the experimental work was conducted by me.

3.4.7 On the sub-cellular localization of *C. elegans* secreted cereblon

All experiments with worms were performed by me under the supervision of Igor Iatsenko from the Department of Integrative Evolutionary Biology. The experimental setup was designed by Adrian Streit, Igor Iatsenko and me. Martin Schüchel supported me in antibody production by injecting animals and collecting the blood samples.

4 CONCLUSIONS

In summary, I present here the study of cereblon proteins with particular focus on their thalidomide-binding domain, also referred to as CULT. By determining the crystal structure of the bacterial homolog of the CULT domain we demonstrated how cereblon recognizes its ligand. The CULT domain harbors a binding pocket formed by three conserved tryptophan residues that build an aromatic cage for ligand docking. Using bacterial MsCI4 as a model system, we show that a part of the CULT domain folds upon ligand binding, accompanied by an increase of the thermal stability of the protein. Searching for natural ligands of the CULT domain, we identify uridine mimicking thalidomide binding, which is demonstrated by structural studies as well as *in vivo*. Up to now, uridine represents the only known physiological ligand of cereblon proteins that is universally present in bacteria and eukaryotes. Additionally, we note that the binding pocket of the CULT domain resembles the aromatic cage of lysine or arginine methylation readers. Although no interaction between the protein and these amino acids and their methylated derivatives could be shown *in vitro*, we consider that in the context of a specific flanking sequence, such residues are conceivable as ligands of cereblon.

Further, we developed a FRET-based *in vitro* assay for testing binding and affinities of cereblon to its ligands. Comparative analysis of the thalidomide-binding domains from *H. sapiens*, *M. gryphiswaldense*, and *C. elegans* displayed similar affinities to the same ligands on a relative scale, suggesting that all these proteins bind the same natural ligand. Moreover, this study demonstrated the expedience of the bacterial protein as a robust model system for testing the substrate specificity of the CULT domain to its ligands, such as newly developed pharmacologically relevant compounds. Using a FRET-based assay, we showed that various pharmaceuticals broadly used in medicine interact with the CULT domain in the micromolar range, displaying their off-target effect. *In vivo* test further validated their cereblon-dependent teratogenicity in zebrafish embryos. In time, the assay provides a useful tool for the characterization of the teratogenic potential of new compounds mediated *via* cereblon.

Searching for interacting proteins, we showed that the bacterial CULT domain binds to transcription termination factor Rho, suggesting its role in transcriptional regulation. Similarly, proteins involved in transcriptional repression and gene silencing were shown to co-immunoprecipitate with human cereblon *via* its CULT domain from HEK293T cells. This is further supported by the fact that *hCRBN* interacts with demethylase LSD1, which is

essential for transcriptional regulation. Development of multicellular organism is controlled by temporal and spatial expression of specific genes, which is regulated by transcription repression and activation complexes. Taken all together, our results provide a promising starting point for following investigations to unravel cereblon function that possibly includes transcriptional regulation, particularly during limb formation.

5 BIBLIOGRAPHY

Aizawa, M., Abe, Y., Ito, T., Handa, H., and Nawa, H. (2011). mRNA distribution of the thalidomide binding protein cereblon in adult mouse brain. *Neurosci. Res.* *69*, 343–347.

Ambros, V. (1991). Efficient maintenance. *EMBO J.* *10*, 3959–3970.

Bedford, L., Lowe, J., Dick, L.R., Mayer, R.J., and Brownell, J.E. (2011). Ubiquitin-like protein conjugation and the ubiquitin–proteasome system as drug targets. *Nat. Rev. Drug Discov.* *10*, 29–46.

Bian, C., Chen, Q., and Yu, X. (2015). The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *Elife* *4*, 1–17.

Bina M. (2006). Gene mapping, discovery, and expression: methods and protocols. Humana press.

Boichenko, I., Deiss, S., Bär, K., Hartmann, M.D., and Hernandez Alvarez, B. (2016). A FRET-Based Assay for the Identification and Characterization of Cereblon Ligands. *J. Med. Chem.* *59* (2), 770–774.

Bradford M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* *72*, 248–54.

Brent R.L. (1964). Drug testing in animals for teratogenic effects. Thalidomide in the pregnant rat. *J. Pediatr.* *64*, 762–770.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J.R. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* *4*, 484–494.

Boudvillain, M., Figueroa-Bossi, N., and Bossi, L. (2013). Terminator still moving forward: expanding roles for Rho factor. *Curr. Opin. Microbiol.* *16*, 118–124.

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., et al. (2014). Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat. Struct. Mol. Biol.* *21*, 803–809.

Chang, X., Zhu, Y., Shi, C., and Stewart, A. K. (2013). Mechanism of immunomodulatory drugs ' action in the treatment of multiple myeloma Development of IMiDs Effects of IMiDs in the Treatment of MM. *Acta Biochim. Biophys.* *46*, 1–14.

Chen, Y.-A., Peng, Y.-J., Hu, M.-C., Huang, J.-J., Chien, Y.-C., Wu, J.-T., Chen, T.-Y., and Tang, C.-Y. (2015). The Cullin 4A/B-DDB1-Cereblon E3 Ubiquitin Ligase Complex Mediates the Degradation of CLC-1 Chloride Channels. *Sci. Rep.* *5*, 1–13.

Chung, F., Lu, J., Palmer, B.D., Kestell, P., Browett, P., Baguley, B.C., Tingle, M., and Ching, L.M. (2004). Thalidomide pharmacokinetics and metabolite formation in mice, rabbits, and multiple myeloma patients. *Clin. Cancer Res.* *10*, 5949–5956.

Cohen, P., and Tcherpakov, M. (2010). Will the Ubiquitin System Furnish as Many Drug Targets as Protein Kinases? *Cell* *143*, 686–693.

Corral, L.G., Albert, M., Kaplan, G., Sarno, E.N., Galilly, R., Cohn, Z. A, Kaplan, G., Mertins, P., Svinkina, T., Carr, S. A, et al. (2014). The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. *Science* *343*, 305–309.

D'Amato, R.J., Loughnan, M.S., Flynn, E., and Folkman, J. (1994). Thalidomide is an inhibitor of angiogenesis. *Proc. Nat. Acad. Sci. U. S. A.* *91*, 4082–4085.

Dredge, K., Horsfall, R., Robinson, S.P., Zhang, L.H., Lu, L., Tang, Y., Shirley, M. A., Muller, G., Schafer, P., Stirling, D., et al. (2005). Orally administered lenalidomide (CC-5013) is anti-angiogenic in vivo and inhibits endothelial cell migration and Akt phosphorylation in vitro. *Microvasc. Res.* *69*, 56–63.

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H., and Calos, M.P. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell Biol.* *7*, 379–387.

Duerr J.S. (2006). Immunohistochemistry. The *C. elegans* Research Community (ed), WormBook. <http://www.wormbook.org>.

Escoubet-Lozach, L., Lin, I.L., Jensen-Pergakes, K., Brady, H. a., Gandhi, A.K., Schafer, P.H., Muller, G.W., Worland, P.J., Chan, K.W.H., and Verhelle, D. (2009). Pomalidomide and lenalidomide induce p21(WAF-1) expression in both lymphoma and multiple myeloma through a LSD1-mediated epigenetic mechanism. *Cancer Res.* *69*, 7347–7356.

Field, H. A., Dong, P.D.S., Beis, D., and Stainier, D.Y.R. (2003). Formation of the digestive system in zebrafish. *Dev. Biol.* *261*, 197–208.

Fischer, E.S., Böhm, K., Lydeard, J.R., Yang, H., Stadler, M.B., Cavadini, S., Nagel, J., Serluca, F., Acker, V., Lingaraju, G.M., et al. (2014). Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* *512*, 49–53.

Foster, C.T., Dovey, O.M., Lezina, L., Luo, J.L., Gant, T.W., Barlev, N., Bradley, A., and Cowley, S.M. (2010). Lysine-Specific Demethylase 1 Regulates the Embryonic Transcriptome and CoREST Stability. *Mol. Cell Biol.* *30*, 4851–4863.

Franks, M.E., Macpherson, G.R., and Figg, W.D. (2004). Thalidomide. *363*, 1802–1811.

Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of Centromere for CENP-A Recruitment by Human hMis18 α , hMis18 β , and M18BP1. *Dev. Cell* *12*, 17–30.

- Gandhi, A.K., Kang, J., Havens, C.G., Conklin, T., Ning, Y., Wu, L., Ito, T., Ando, H., Waldman, M.F., Thakurta, A., et al. (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.
- Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J.I., Saijo, M., Drapkin, R., Kisselev, A.F., Tanaka, K., and Nakatani, Y. (2003). The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* *113*, 357–367.
- Grotewold, L., and Rüther, U. (2002). Bmp, Fgf and Wnt signalling in programmed cell death and chondrogenesis during vertebrate limb development: The role of Dickkopf-1. *Int. J. Dev. Biol.* *46*, 943–947.
- Gülich, S., Uhlén, M., and Hober, S. (2000). Protein engineering of an IgG-binding domain allows milder elution conditions during affinity chromatography. *J. Biotechnol.* *76*, 233–243.
- Hansen, J.M., and Harris, C. (2004). A novel hypothesis for thalidomide-induced limb teratogenesis: redox misregulation of the NF-kappaB pathway. *Antioxid. Redox Signal* *6*, 1–14.
- Hansen, J.M., Gong, S.-G., Philbert, M., and Harris, C. (2002). Misregulation of gene expression in the redox-sensitive NF-kappab-dependent limb outgrowth pathway by thalidomide. *Dev. Dyn.* *225*, 186–194.
- Hansen, J.M., Harris, K.K., Philbert, M. A, and Harris, C. (2002). Thalidomide modulates nuclear redox status and preferentially depletes glutathione in rabbit limb versus rat limb. *J. Pharmacol. Exp. Ther.* *300*, 768–776.
- Hartmann, M.D., Boichenko, I., Coles, M., Zanini, F., Lupas, A.N., and Hernandez Alvarez, B. (2014). Thalidomide mimics uridine binding to an aromatic cage in cereblon. *J. Struct. Biol.* *188*, 225–232.
- Hartmann, M.D., Boichenko, I., Coles, M., Lupas, A.N., and Hernandez Alvarez, B. (2015). Structural Dynamics of the Cereblon Ligand Binding Domain. *PLoS One* *10* (5), 1-16.
- Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* *118*, 715–729.
- Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.* *2*, 1849–1861.
- Higa, L.A., and Zhang, H. (2007). Stealing the spotlight: CUL4-DDB1 ubiquitin ligase docks WD40-repeat proteins to destroy. *Cell Div.* *2*:5, 1-9.

- Higa, L.A., Wu, M., Ye, T., Kobayashi, R., Sun, H., and Zhang, H. (2006). CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat. Cell Biol.* 8, 1277–1283.
- Higgins, J.J., Pucilowska, J., Lombardi, R. Q., Rooney, J.P. (2004). A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology.* 63, 1967-1931
- Hobert O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques.* 4, 728-30.
- Hohberger, B., and Enz, R. (2009). Cereblon is expressed in the retina and binds to voltage-gated chloride channels. *FEBS Lett.* 583, 633–637.
- Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E. A. (2012). Riboswitch control of Rho-dependent transcription termination. *Proc. Nat. Acad. Sci.* 109, 5376–5381.
- Hostomská, J., Volfová, V., Mu, J., Garfield, M., Rohousová, I., Volf, P., Valenzuela, J.G., and Jochim, R.C. (2009). Analysis of salivary transcripts and antigens of the sand fly *Phlebotomus arabicus*. *BMC Genomics* 10, 282.
- Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J. A., Richter, M., Opravil, S., Shiekhatar, R., Bedford, M.T., Jenuwein, T., et al. (2007). p53 is regulated by the lysine demethylase LSD1. *Nature* 449, 105–108.
- Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., and Handa, H. (2010). Identification of a primary target of thalidomide teratogenicity. *Science* 327, 1345–1350.
- Janer, G., Slob, W., Hakkert, B., Vermeire, T., and Piersma, A. (2008). A retrospective analysis of developmental toxicity studies in rat and rabbit: what is the added value of the rabbit as an additional test species? *Regul. Toxicol. Pharmacol.* 50, 206–217.
- Jo, S., Lee, K.-H., Song, S., Jung, Y.-K., and Park, C.-S. (2005). Identification and functional characterization of cereblon as a binding protein for large-conductance calcium-activated potassium channel in rat brain. *J. Neurochem.* 94, 1212–1224.
- Käll, L., Storey, J.D., MacCoss, M.J., and Noble, W.S. (2008). Posterior error probabilities and false discovery rates: Two sides of the same coin. *J. Proteome Res.* 7, 40–44.
- Kelley, K.D., Miller, K.R., Todd, A., Kelley, A.R., Tuttle, R., and Berberich, S.J. (2010). YPEL3, a p53-regulated gene that induces cellular senescence. *Cancer Res.* 70, 3566–3575.
- Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* 1751, 119–139.
- Knobloch, J., and Ruther, U. (2008). Shedding light on an old mystery: thalidomide suppresses survival pathways to induce limb defects. *Cell Cycle* 7, 1121–1127.

Knobloch, J., Shaughnessy, J.D., and Rüther, U. (2007). Thalidomide induces limb deformities by perturbing the Bmp/Dkk1/Wnt signaling pathway. *FASEB J.* 21, 1410–1421.

Knobloch, J., Schmitz, I., Götz, K., Schulze-Osthoff, K., and Rüther, U. (2008). Thalidomide induces limb anomalies by PTEN stabilization, Akt suppression, and stimulation of caspase-dependent cell death. *Mol. Cell. Biol.* 28, 529–538.

Köcher, T., and Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* 4, 807–815.

Krönke, J., Udeshi, N.D., Narla, A., Grauman, P., Hurst, S.N., Mcconkey, M., Svinkina, T., Heckl, D., Comer, E., Li, X., et al. (2014). Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma cells. *Science* 343, 301–306.

Krönke, J., Fink, E.C., Hollenbach, P.W., MacBeth, K.J., Hurst, S.N., Udeshi, N.D., Chamberlain, P.P., Mani, D.R., Man, H.W., Gandhi, A.K., et al. (2015). Lenalidomide induces ubiquitination and degradation of CK1 α in del(5q) MDS. *Nature* 523, 183–188.

Kuppuswamy, M., Vijayalingam, S., Zhao, L.-J., Zhou, Y., Subramanian, T., Ryerse, J., and Chinnadurai, G. (2008). Role of the PLDLS-binding cleft region of CtBP1 in recruitment of core and auxiliary components of the corepressor complex. *Mol. Cell. Biol.* 28, 269–281.

Lacy, M.Q., Hayman, S.R., Gertz, M. A., Dispenzieri, A., Buadi, F., Kumar, S., Greipp, P.R., Lust, J. A., Russell, S.J., Dingli, D., et al. (2009). Pomalidomide (CC4047) plus low-dose dexamethasone as therapy for relapsed multiple myeloma. *J. Clin. Oncol.* 27, 5008–5014.

Lacy, M.Q., Allred, J.B., Gertz, M. A., Hayman, S.R., Short, K.D., Buadi, F., Dispenzieri, A., Kumar, S., Greipp, P.R., Lust, J. A., et al. (2011). Pomalidomide plus low-dose dexamethasone in myeloma refractory to both bortezomib and lenalidomide: comparison of 2 dosing strategies in dual-refractory disease. *Blood* 118, 2970–2975.

Laemmli U.K (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227(5259),680-5.

Lee, J., and Zhou, P. (2007). DCAFs, the Missing Link of the CUL4-DDB1 Ubiquitin Ligase. *Mol. Cell* 26, 775–780.

Lee, B.C., Dikiy, A., Kim, H.-Y., and Gladyshev, V.N. (2009). Functions and evolution of selenoprotein methionine sulfoxide reductases. *Biochim. Biophys. Acta* 1790, 1471–1477.

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., et al. (2012). EZH2 generates a methyl degon that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. *Mol. Cell* 48, 572–586.

Lee, K.J., Lee, K.M., Jo, S., Kang, K.W., and Park, C.S. (2010). Induction of cereblon by NF-E2-related factor 2 in neuroblastoma cells exposed to hypoxia-reoxygenation. *Biochem. Biophys. Res. Commun.* 399, 711–715.

- Lee, K.M., Jo, S., Kim, H., Lee, J., and Park, C.-S. (2011). Functional modulation of AMP-activated protein kinase by cereblon. *Biochim. Biophys. Acta* 1813, 448–455.
- Lee, K.M., Lee, J., and Park, C.S. (2012). Cereblon inhibits proteasome activity by binding to the 20S core proteasome subunit beta type 4. *Biochem. Biophys. Res. Commun.* 427, 618–622.
- Lee, K.M., Yang, S.-J., Kim, Y.D., Choi, Y.D., Nam, J.H., Choi, C.S., Choi, H.-S., and Park, C.-S. (2013). Disruption of the Cereblon Gene Enhances Hepatic AMPK Activity and Prevents High-Fat Diet-Induced Obesity and Insulin Resistance in Mice. *Diabetes* 62, 1855–1864.
- Lenz, W. (1988). A short history of thalidomide embryopathy. *Teratology* 38, 203-215.
- Leung, D.W., and Amarasinghe, G.K. (2012). Structural insights into RNA recognition and activation of RIG-I-like receptors. *Curr. Opin. Struct. Biol.* 22, 297–303.
- Lewis, Z. A, Adhvaryu, K.K., Honda, S., Shiver, A.L., Knip, M., Sack, R., and Selker, E.U. (2010). DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC. *PLoS Genet.* 6, 1-11.
- Liu, J., Ye, J., Zou, X., Xu, Z., Feng, Y., Zou, X., Chen, Z., Li, Y., and Cang, Y. (2014). CRL4A(CRBN) E3 ubiquitin ligase restricts BK channel activity and prevents epileptogenesis. *Nat. Commun.* 5, 1-9.
- Lopez-Girona, a, Mendy, D., Ito, T., Miller, K., Gandhi, a K., Kang, J., Karasawa, S., Carmel, G., Jackson, P., Abbasian, M., et al. (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., and Korycinski, M. (2015). The Thalidomide-Binding Domain of Cereblon Defines the CULT Domain Family and Is a New Member of the β -Tent Fold. *PLoS Comput. Biol.* 11, 1-11.
- Mahony, C., Erskine, L., Niven, J., Greig, N.H., Figg, W.D., and Vargesson, N. (2013). Pomalidomide is nonteratogenic in chicken and zebrafish embryos and nonneurotoxic in vitro. *Proc. Nat. Acad. Sci. U. S. A.* 110, 12703–12708.
- Malkova, N. V, Gallagher, J.J., Yu, C.Z., Jacobs, R.E., and Patterson, P.H. (2014). Manganese-enhanced magnetic resonance imaging reveals increased DOI-induced brain activity in a mouse model of schizophrenia. *Proc. Nat. Acad. Sci. U. S. A.* 111, 2492–2500.
- Martiniani, R., Di Loreto, V., Di Sano, C., Lombardo, A., and Liberati, A.M. (2012). Biological activity of lenalidomide and its underlying therapeutic effects in multiple myeloma. *Adv. Hematol.* 2012, 842945.
- Meganathan, K., Jagtap, S., Wagh, V., Winkler, J., Gaspar, J.A., Hildebrand, D., Trusch, M., Lehmann, K., Hescheler, J., Schlüter, H., et al. (2012). Identification of thalidomide-specific

transcriptomics and proteomics signatures during differentiation of human embryonic stem cells. *PLoS One* 7, 1-15.

Melchert, M., and List, A. (2007). The thalidomide saga. *Int. J. Biochem. Cell Biol.* 39, 1489–1499.

Miller, M.T., and Strömblad, K. (1999). Teratogen update: Thalidomide: A review, with a focus on ocular findings and new potential uses. *Teratology* 60, 306–321.

Mello C., Fire A. (1995). DNA transformation. *Methods Cell Biol.* 48,451-82.

Newman C.G. (1986). The thalidomide syndrome: risks of exposure and spectrum of malformations. *Clin. Perinatol.* 13, 555–573.

Nüsslein-Volhard C., Dahm R. (2002) *Zebrafish: a practical approach*. Oxford: Oxford University Press

Pajni-Underwood, S., Wilson, C.P., Elder, C., Mishina, Y., and Lewandoski, M. (2007). BMP signals control limb bud interdigital programmed cell death by regulating FGF signaling. *Development* 134, 2359–2368.

Parkie M., Webb M. (1983). Embryotoxicity and teratogenicity of thalidomide in rats. *Teratology* 3, 327-32.

Parman, T., Wiley, M.J., and Wells, P.G. (1999). Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. *Nat. Med.* 5, 582–585.

Rajadhyaksha, A.M., Ra, S., Kishinevsky, S., Lee, A.S., Romanienko, P., Duboff, M., Yang, C., Zupan, B., Byrne, M., Daruwalla, Z.R., et al. (2012). Behavioral characterization of cereblon forebrain-specific conditional null mice: A model for human non-syndromic intellectual disability. *Behav. Brain Res.* 226, 428–434.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

Sambrook J., Fritsch E. F., Maniatis T. (1989). *Molecular cloning: A laboratory manual*: 2nd ed. Cold Spring Harbor Laboratory Press.

Scrima, A., Koníčková, R., Czyzewski, B.K., Kawasaki, Y., Jeffrey, P.D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N.P., and Thomä, N.H. (2008). Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2 Complex. *Cell* 135, 1213–1223.

Shortt, J., Hsu, A. K., and Johnstone, R.W. (2013). Thalidomide-analogue biology: immunological, molecular and epigenetic targets in cancer therapy. *Oncogene* 32, 4191–4202.

Stephens, T.D., Bunde, C.J., and Fillmore, B.J. (2000). Mechanism of action in thalidomide teratogenesis. *Biochem. Pharmacol.* 59, 1489–1499.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Nat. Acad. Sci. U. S. A.* *101*, 6062–6067.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct.* *38*; *Mol. Biol.* *14*, 1025–1040.

Tenyi T., Csabi G., Hamvas E., Varga E., Herold R. (2008) The decoding of the flouting of the Gricean relevance maxim is impaired in mental retardation caused by perinatal hypoxia. *Neuropsychopharmacol.* *10*, 271–274.

Therapontos, C., Erskine, L., Gardner, E.R., Figg, W.D., and Vargesson, N. (2009). Thalidomide induces limb defects by preventing angiogenic outgrowth during early limb formation. *Proc. Nat. Acad. Sci. U. S. A.* *106*, 8573–8578.

Ueda, J., Tachibana, M., Ikura, T., and Shinkai, Y. (2006). Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. *J. Biol. Chem.* *281*, 20120–20128.

Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K. A., Zhu, P., Garcia-Bassets, I., et al. (2007). Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* *446*, 882–887.

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., et al. (2009). LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell* *138*, 660–672.

Weber, D.M., Chen, C., Niesvizky, R., Wang, M., Belch, A., Stadtmauer, E. A., Siegel, D., Borrello, I., Rajkumar, S.V., Chanan-Khan, A.A., et al. (2007). Lenalidomide plus dexamethasone for relapsed multiple myeloma in North America. *N. Engl. J. Med.* *357*, 2133–2142.

Weimann, M., Grossmann, A., Woodsmith, J., Özkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E., et al. (2013). A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat. Methods* *10*, 339–342.

WHO (World Health Organization). Use of thalidomide in leprosy. <http://www.who.int/lep/research/thalidomide/en/>

Wittschieben, B.O., and Wood, R.D. (2003). DDB complexities. *DNA Repair.* *2*, 1065–1069.

Wu K.K. (2006). Analysis of protein-DNA binding by streptavidin–agarose pull down. *Methods in Molecular Biology* *338*, 281–290

Xu, G., Jiang, X., and Jaffrey, S.R. (2013). A mental retardation-linked nonsense mutation in cereblon is rescued by proteasome inhibition. *J. Biol. Chem.* *288*, 29573–29585.

Xu, H., Wang, J., Hu, Q., Quan, Y., Chen, H., Cao, Y., Li, C., Wang, Y., and He, Q. (2010). DCAF26, an adaptor protein of Cul4-based E3, is essential for DNA methylation in *Neurospora crassa*. *PLoS Genet.* 6, 1-13.

Yokoyama, A., Igarashi, K., Sato, T., Takagi, K., Otsuka, M.I., Shishido, Y., Baba, T., Ito, R., Kanno, J., Ohkawa, Y., et al. (2014). Identification of myelin transcription factor 1 (MyT1) as a subunit of the neural cell type-specific lysine-specific demethylase 1 (LSD1) complex. *J. Biol. Chem.* 289, 18152–18162.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. *Cell Res.* 21, 564–578.

Zeldis J.B., Williams B.A., Thomas S.D., Elsayed M.E (1999) S.T.E.P.: A Comprehensive Program for Controlling and Monitoring Access to Thalidomide. *Clin. Ther.* 21, 319-330

Zhao, Y., Shen, Y., Yang, S., Wang, J., Hu, Q., Wang, Y., and He, Q. (2009). Ubiquitin ligase components Cullin4 and DDB1 are essential for DNA methylation in *Neurospora crassa*. *J. Biol. Chem.* 285, 4355-4365.

Zhu, Y.X., Braggio, E., Shi, C.X., Bruins, L. A., Schmidt, J.E., Van Wier, S., Chang, X.B., Bjorklund, C.C., Fonseca, R., Bergsagel, P.L., et al. (2011). Cereblon expression is required for the antimyeloma activity of lenalidomide and pomalidomide. *Blood* 118, 4771–4779.

6 ACKNOWLEDGEMENTS

It is my pleasure to thank people, who have supported me during my doctorate.

I would like to express my gratitude to my advisor Prof. Dr. Andrei Lupas for believing in my personal skills and giving me the opportunity to work on this project. I thank him for the scientific guidance, his continuing support, and discussions.

I would also like to express my special appreciation to my intermediate supervisor Birte Hernandez Alvarez for her patience, trust, support, teaching skills, and our countless discussions about the project.

I am grateful to the members of my thesis advisory committee for helpful comments concerning the project, Dr. Birte Höcker and Prof. Dr. Alfred Nordheim, who also supervised this thesis at the Eberhard Karls University, Tübingen, and to Prof. Dr. Volkmar Braun and Prof. Dr. Karl Forchhammer for being in the committee at my PhD examination.

Further, I would like to thank Marcus Hartmann for discussions and the ideas he contributed to the project. Next, I want to express appreciation to Murray Coles and Manish Chaubey for performing NMR studies; Boris Macek, Mirita Franz, and Johannes Madlung for their help with planning and performing mass spectrometric analysis at the Proteome Centre (University Tübingen); Mateusz Korycinski and Hongbo Zhu for help in bioinformatic analysis; Uwe Irion, Christian Söllner, Horst Geiger for introducing me to the fish facilities; Igor Iatsenko and Adrian Streit for their advice during my work with *C. elegans*; Martin Schückel for his help in animal care during antibody production; Reinhard Albrecht and Kerstin Bär for setting up the crystallization screens; Fabio Zanini for writing a software for the analysis of zebrafish images; and Klaus Kopec for his help in microarray data analysis. Next, I would like to thank Silvia Deiss and Kerstin Bär for their technical assistance on the cereblon project.

I am grateful to all past and present members of the Department of Protein Evolution for creating an enjoyable working atmosphere. Specially, I would like to thank (in alphabetical order): Harshul Arora, Emilija Basina, Silvia Deiss, Anitha Jegantham, Ioanna Karamichali, Amit Kumar, Franka Scharfenberg, Martin Schückel, Edgardo Sepulveda, Bogdan Sieriebriennikov, Fabio Zanini for their support beyond the lab.

I also deeply appreciate constant support from all my Ukrainian friends.

I would like to send my deepest words of gratitude to my husband Igor for his constant care, encouragement, help, and every day support. Finally, I thank my parents and my brother for their unexhausted support and motivation throughout all the years.

Appendix I

	Origin	Vector	Cloning sites	Tag	Expression system	Expression level	Remarks
MsCI4	<i>Magnetospirillum gryphiswaldense</i>	pETHis1a	NcoI, BamHI	N-terminal-His-FLAG	<i>E. coli</i>	+++	partly soluble
MsCI4 ^{YW/AA}	<i>Magnetospirillum gryphiswaldense</i>	pETHis1a	NcoI BamHI	N-terminal-His-FLAG	<i>E. coli</i>	+++	partly soluble
CRBN	<i>Homo sapiens</i>	pETHis1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	++	insoluble
CRBN Δ 1-79	<i>Homo sapiens</i>	pET His1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	+	insoluble
CRBN Δ 1-186	<i>Homo sapiens</i>	pETHis1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	++	insoluble
CRBN Δ 1-318	<i>Homo sapiens</i>	pETHis1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	+++	insoluble
CRBN 80-318	<i>Homo sapiens</i>	pETHis1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	++	insoluble
CRBN 32-247	<i>Homo sapiens</i>	pETHis1a	NcoI HindIII	N-terminal-His	<i>E. coli</i>	+++	insoluble

CeCRBN ^{secΔ1-15}	<i>Caenorhabditis elegans</i>	pETHis1a	NcoI BamHI	N-terminal-His	<i>E. coli</i>	+++	insoluble
CRBN	<i>Alvinella pompejana</i>	pETHis1a_NdeI	NdeI BamHI	N-terminal-His	<i>E. coli</i>	+++	insoluble
CRBN	<i>Alvinella pompejana</i>	pET28b	AgeI	N-terminal-His-SUMO	<i>E. coli</i>	+++	insoluble
CRBN	<i>Homo sapiens</i>	pFAST-Bac/CT TOPO	Blunt-end cloning	C-terminal-His	<i>Sf9</i>	+++	degradation
CeCRBN ^{secΔ1-15}	<i>Caenorhabditis elegans</i>	pFAST-Bac/CT TOPO	Blunt-end cloning	C-terminal-His	<i>Sf9</i>	+++	-
CRBN	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI XhoI	N-terminal-His-ZZ	<i>Sf9</i>	+++	degradation
CRBN ^{Δ1-108}	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI XhoI	N-terminal-His-ZZ	<i>Sf9</i>	++	degradation
CRBN ^{Δ1-193}	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI XhoI	N-terminal-His-ZZ	<i>Sf9</i>	++	degradation
CRBN ^{Δ1-315}	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI XhoI	N-terminal-His-ZZ	<i>Sf9</i>	+++	-
CRBN 76-186	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI	N-terminal-His-	<i>Sf9</i>	+	-

			HindIII	ZZ			
CRBN 76-249	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI HindIII	N-terminal-His- ZZ	<i>Sf9</i>	+	degradation
CRBN 76-316	<i>Homo sapiens</i>	pFastBac_NcoI	NcoI HindIII	N-terminal-His- ZZ	<i>Sf9</i>	+	degradation
DDB1	<i>Homo sapiens</i>	pFAST-Bac/CT TOPO	Blunt-end cloning	C-terminal-His	<i>Sf9</i>	+++	-
DDB1	<i>Homo sapiens</i>	pFAST-NdeI	AgeI XbaI	N-terminal-His	<i>Sf9</i>	+++	-

Appendix II

	PCR product	Restriction Enzymes	Tag used in petHis 1a	Expression strain	Expression level	Solubility
Transcription termination factor Rho	+	NdeI BamHI	N-terminal-His	<i>E. coli C41</i>	+++	++
Aspartokinase	+	NdeI BamHI	N-terminal-His	<i>E. coli Arctic Express</i>	+++	+
Lipopolysaccharide transport protein B: ATP-binding component of ABC superfamily	+	NdeI BamHI	N-terminal-His	<i>E. coli C41</i>	+++	++
Putative glyoxalase I family protein putative Methylmalonyl-CoA epimerase	+	NdeI BamHI	N-terminal-His	<i>E. coli Arctic Express</i>	+++	++
3-isopropylmalate dehydratase large subunit	+	NdeI BamHI	N-terminal-His	<i>E. coli Arctic Express</i>	+++	+

Appendix III

Name	Oligonucleotides
bio C+	5`-AAGTCCTGAATGAATCAAGCAGGCGTTGAAG-3`
bio U+	5`-AAGTCCTGAATGAATUAAGCAGGCGTTGAAG-3`
bio G-	5`-CTTCAACGCCTGCTTGATTCATTCAGGACTT-3`
bio C+/bio G-	5`-AAGTCCTGAATGAATCAAGCAGGCGTTGAAG-3` 3`-TTCAGGACTTACTTAGTTCGTCGCAACTTC-5`
bio U+/bio G-	5`-AAGTCCTGAATGAATUAAGCAGGCGTTGAAG-3` 3`-TTCAGGACTTACTTAGTTCGTCGCAACTTC-5`
bio CC+	5`-GATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGGAA CGAAAACACGTTAAGGGATTTGGTCATG-3`
bio UU+	5`-GATCCTTTGATCTTTUACGGGGTCTGACGCTCAGTGGAAUGAAAACACGTTAAGGGATTTGGTCATG-3`
bio GG-	5`-CATGACCAAATCCCTTAACGTGAGTTTTCTTCCACTGAGCGTCAGACCCGTA GAAAAGATCAAAGGATC-3`
bio UU+/bio GG-	5`-GATCCTTTGATCTTTUACGGGGTCTGACGCTCAGTGGAAUGAAAACACGTTAAGGGATTTGGTCATG-3` 3`-CTAGGAACTAGAAA GATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGAATTCCTAAAACAGTAC-5`
bio CC+/bio GG-	5`-GATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGGAA CGAAAACACGTTAAGGGATTTGGTCATG-3` 3`-CTAGGAACTAGAAA GATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGAATTCCTAAAACAGTAC-5`

«+» and «-» show complementary strands. All DNA sequences, except «-» strands in the double stranded DNA, carried biotin (bio) at their 5`.

Appendix IV

Antibodies	Dilution
Mouse anti-His (Dianova; DIA-900)	1/1,000
Rabbit anti-FLAG (Dianova; ABR-01098)	1/1,000
Mouse anti-HA (Covance; MMS-101R200)	1/1,000
Rabbit anti-HA (Sigma; H6908)	1/1,000
Rabbit anti-Myc (Abcam; ab9106)	1/1,000
Goat anti Wiz (Acris GmbH; AP16897PU-N)	1/500
Rabbit anti-actin (Sigma; A2066)	1/1,000
Rabbit anti-CeCRBN _{sec} (in-house made)	1/1,000
Anti-mouse HRP (Sigma; A4416-1ML)	1/3,500
Anti-rabbit (Sigma; A0545-1ML)	1/3,500
Anti-goat (Sigma; A5420-1ML)	1/7,500
Anti-mouse (Sigma; SAB3701116)	1/1,000

Appendix V

	Matrix	PCR product	Restriction Enzymes	Tag used in pcDNA3.1
CRBN	plasmid	+	HindIII XhoI	N-terminal-His
CRBN	plasmid	+	HindIII XhoI	N-terminal-FLAG
CRBN	plasmid	+	HindIII XhoI	N-terminal-Myc
CRBN	plasmid	+	HindIII XhoI	N-terminal-HA
DDB1	plasmid	+	KpnI AgeI	N-terminal-His
DDB1	plasmid	+	KpnI AgeI	N-terminal-FLAG
DDB1	plasmid	+	KpnI AgeI	N-terminal-Myc
DDB1	plasmid	+	KpnI AgeI	N-terminal-HA

WIZ	human cDNA	+	NheI AgeI	N-terminal-Myc
HIC2	human cDNA	+	NheI AgeI	N-terminal-Myc
EHTM1	gene synthesis	-	NheI AgeI	N-terminal-Myc_pCMV6*
EHTM2	human cDNA	-	NheI AgeI	N-terminal-Myc
PATZ1	human cDNA	+	NheI AgeI	N-terminal-Myc
zinc finger protein 644	human cDNA	+	NheI AgeI	N-terminal-Myc
zinc finger protein 687	human cDNA	+	BamI AgeI	N-terminal-Myc
PKCBP	human cDNA	+	NheI NotI	N-terminal-Myc
LSM12	human cDNA	+	NheI AgeI	N-terminal-Myc

myc-associated zinc finger protein	human cDNA	-	Nhe I Age I	N-terminal-Myc
LSD1	human cDNA	+	Nhe I Xho I	N-terminal-FLAG
SUV39H1	human cDNA	+	Nhe I Bam I	N-terminal-FLAG
LSD1	FLAG-LSD1 pcDNA3.1	+	Kpn I Xho I	C-terminal-HA*
Δ1-108CRBN	HA-CRBN pcDNA3.1	+	Nhe I Xho I	N-terminal-HA
Δ1-194CRBN	HA-CRBN pcDNA3.1	+	Nhe I Xho I	N-terminal-HA
Δ1-314CRBN	HA-CRBN pcDNA3.1	+	Nhe I Xho I	N-terminal-HA
175-877 ^{Δ419-520} LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG
175-260 LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG

175-288 LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG
1-291 LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG
294-877 LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG
175-418 LSD1	FLAG-LSD1 pcDNA3.1	+	Nhe I Xho I	N-terminal-FLAG

*ordered

CURRICULUM VITAE

Name: Iuliia Boichenko
Date of birth: 20.06.1988
Nationality: Ukrainian
Address: Albrecht Dürer str. 3, 72076 Tübingen

Education and research experience

- 07.2011-present - Max Planck Institute for Developmental Biology, Department of Protein Evolution. Thesis title “The CULT domain of cereblon: a pharmacological target and teratogenicity gateway”.
- 09.2009-06.2011 - Taras Shevchenko Kyiv National University, Biological department. Master Degree in Biochemistry with honours.
- 09.2005–06.2009 - Taras Shevchenko Kyiv National University, Biological department. Bachelor Degree in Biology with honours
- 09.1994-06.2005 - Secondary school №1, Shevchenkove, Kiliya district, Odesa region, full general secondary education with honours

PUBLICATIONS

1. Hartmann, M.D., **Boichenko, I.**, Coles, M., Zanini, F., Lupas, A.N., and Hernandez Alvarez, B. (2014). Thalidomide mimics uridine binding to an aromatic cage in cereblon. J. Struct. Biol. *188*, 225–232.
2. Hartmann, M.D., **Boichenko, I.**, Coles, M., Lupas, A.N., and Hernandez Alvarez, B. (2015). Structural Dynamics of the Cereblon Ligand Binding Domain. PLoS One *10*, e0128342.
3. **Boichenko, I.**, Deiss, S., Bär, K., Hartmann, M.D., and Hernandez Alvarez, B. (2016). A FRET-Based Assay for the Identification and Characterization of Cereblon Ligands. J. Med. Chem. *acs.jmedchem.5b01735*.
4. **Boichenko I.**, Lupas A.N., Hernandez Alvarez B., Hartmann M.D. Cereblon is responsive to a wide range of common small molecules and pharmaceuticals. Manuscript in preparation.



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.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

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).

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 MsCl4 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 MsCl4^{YW/AA} and MsCl4^{Y101F} were constructed using QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene) with wild type MsCl4 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 MgCl₂, 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. MsCl4 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 MsCl4 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 MsCl4 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 μ M protein samples both alone and in the presence of 10–500 μ M 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 μ M. 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 μ M. 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 MsCl4^{Y101F} mutant (thalidomide, uracil, uridine and deoxyuridine) and for the MsCl4^{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 μ l 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, MsCl4-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 MsCl4-pomalidomide, MsCl4-lenalidomide, MsCl4-deoxyuridine and MsCl4^{Y101F}-thalidomide were subsequently solved based on the MsCl4-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 (MsCl4-thalidomide), 4V2Z (MsCl4-pomalidomide), 4V30 (MsCl4-lenalidomide), 4V31 (MsCl4-deoxyuridine) and 4V32 (MsCl4^{Y101F}-thalidomide).

Table 1
Crystallization conditions and cryo protection.

Structure	Protein solution additives	Reservoir solution (RS)	Cryo solution
MsCl4-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
MsCl4-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
MsCl4-lenalidomide	3 mM lenalidomide 3% (v/v) DMSO	15 %(w/v) PEG 6000 5 %(w/v) glycerol	RS + 10% (v/v) PEG 300
MsCl4-deoxyuridine	3 mM deoxyuridine	0.1 M citric acid pH 3.5 25 %(w/v) PEG 3350	–
MsCl4 ^{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.

	MsCl4-thalidomide	MsCl4-pomalidomide	MsCl4-lenalidomide	MsCl4-deoxyuridine	MsCl4 ^{Y101F} .thalidomide
<i>Data collection</i>					
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (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
<i>R</i> _{merge}	7.4 (73.2)	3.8 (44.9)	7.3 (41.6)	8.5 (50.5)	8.6 (77.3)
<i>I</i> / <i>σ</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)
<i>Refinement</i>					
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
<i>R</i> _{work} / <i>R</i> _{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 (MsCl4; 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 MsCl4. The double mutant MsCl4^{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 MsCl4^{Y101F}, with a substitution in the binding pocket as described later on. To confirm that MsCl4 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 MsCl4 and MsCl4^{Y101F} both have binding constants for thalidomide in the low micromolar range, whereas the binding constant of the double mutant MsCl4^{YW/AA} was shifted to the millimolar range. Crystallization trials were performed for MsCl4 and MsCl4^{YW/AA}, both in apo form and in presence of thalidomide. Only the trial with MsCl4 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 *P*2₁2₁2₁. 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 MsCl4 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 F770 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 MsCl4-lenalidomide and the MsCl4-pomalidomide complex in the same crystal form as MsCl4-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 MsCl4, as Y101 corresponds to a phenylalanine (F402) in hCrbn. This led us to confirm the validity of our results in the point mutant MsCl4^{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 MsCl4. This is further underpinned by a crystal structure obtained for the MsCl4^{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 MsCl4 and MsCl4^{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 MsCl4-deoxyuridine complex by co-crystallization, again with the same crystal form as for MsCl4-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.

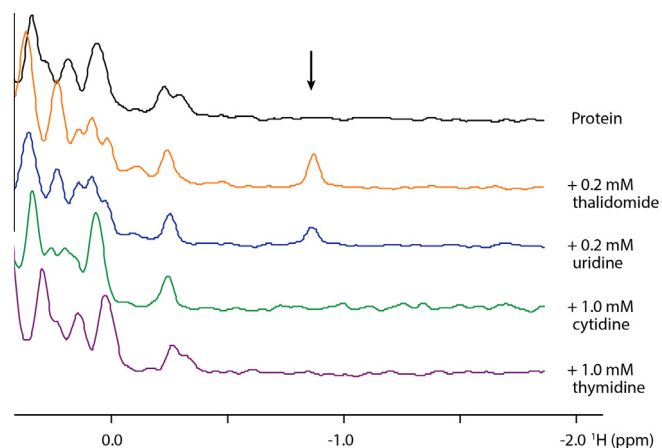


Fig. 2. Thalidomide and nucleosides in the NMR-based ligand binding assay. The traces show 1D proton spectra on 50 μ M MsCl4 at 600 MHz. A detail is shown of the region containing methyl groups that are up-field shifted by close contacts with aromatic groups, and are thus presumably within the protein 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.)

(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).

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

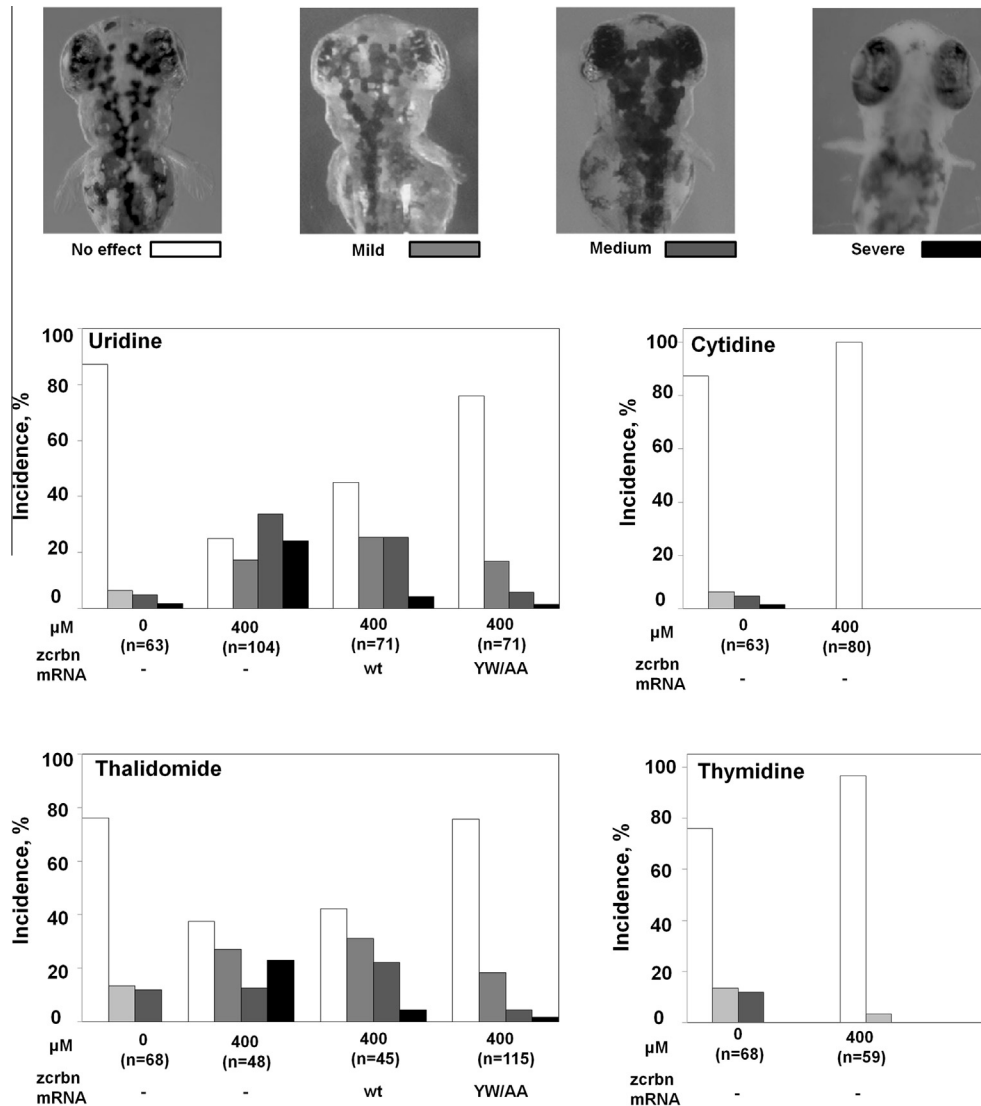


Fig. 3. Zebrafish pectoral fin deformity resulting from thalidomide and nucleoside treatment. The images illustrate the diagnostic criteria of the pectoral fins of embryos at 75 h post fertilization. Uridine treatment causes similar developmental defects as evoked by thalidomide, whereas cytidine and 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, MsCl4 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 DCAF1s 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 MsCl4:thalidomide complex, all with RMSD values <1 Å and a virtually identical geometry of the aromatic cage. Together with our *in vivo* experiments, verifying

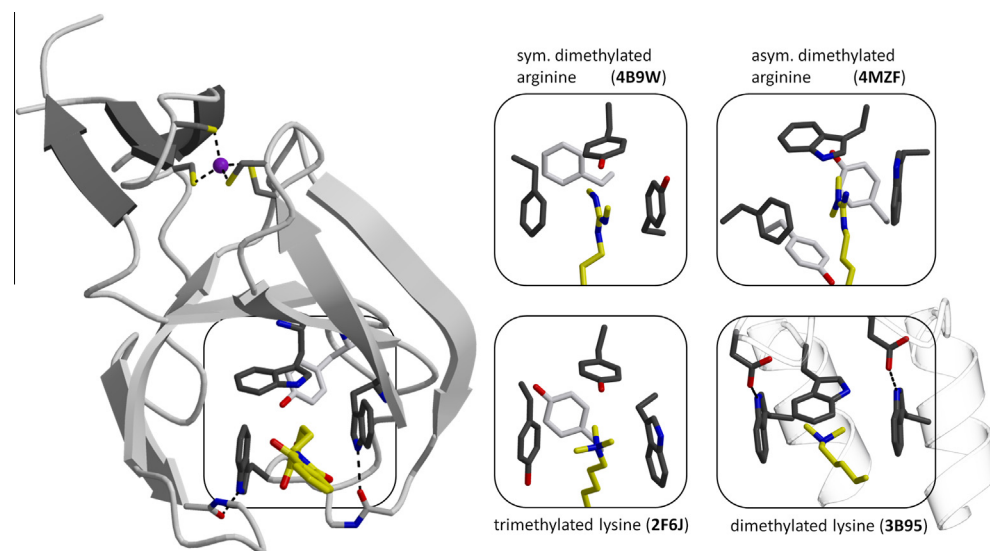


Fig. 4. The aromatic cage of the thalidomide binding pocket compared to the aromatic cages for the readout of specific methylation states of lysines and arginines with the bound substrates symmetrical dimethylated arginine, asymmetrical dimethylated arginine, trimethylated lysine and dimethylated lysine. 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 MsCl4 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 is 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.

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., Dalglish, A.G., 2004. The evolution of thalidomide and its IMiD derivatives as anticancer agents. *Nat. Rev. Cancer* 4, 314–322.

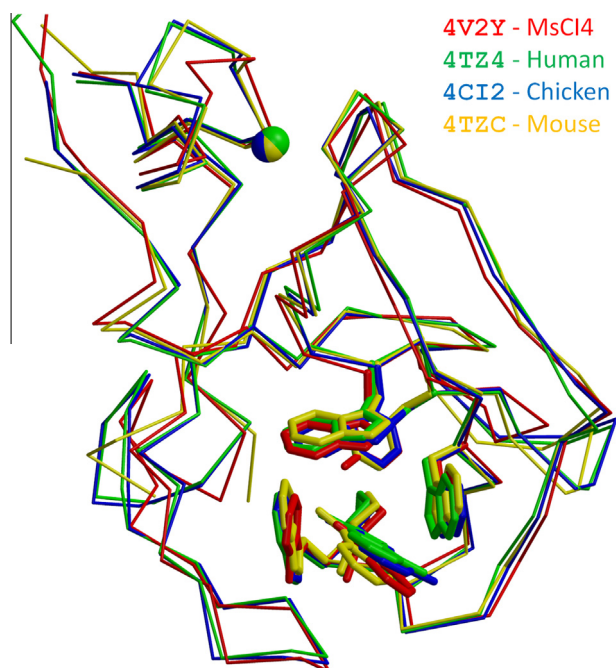


Fig. 5. A comparison of MsCl4-thalidomide to IMiD-bound thalidomide binding domains from human, chicken and mouse, which were published shortly before submission (chicken) or during peer review (human, mouse) of this work. All structures superimpose with an RMSD < 1 Å.

- 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- π 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 deon 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.

RESEARCH ARTICLE

Structural Dynamics of the Cereblon Ligand Binding Domain

Marcus D. Hartmann*, Iuliia Boichenko, Murray Coles, Andrei N. Lupas, Birte Hernandez Alvarez*

Department of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen, Germany

* marcus.hartmann@tuebingen.mpg.de (MDH); birte.hernandez@tuebingen.mpg.de (BHA)

Abstract

Cereblon, a primary target of thalidomide and its derivatives, has been characterized structurally from both bacteria and animals. Especially well studied is the thalidomide binding domain, CULT, which shows an invariable structure across different organisms and in complex with different ligands. Here, based on a series of crystal structures of a bacterial representative, we reveal the conformational flexibility and structural dynamics of this domain. In particular, we follow the unfolding of large fractions of the domain upon release of thalidomide in the crystalline state. Our results imply that a third of the domain, including the thalidomide binding pocket, only folds upon ligand binding. We further characterize the structural effect of the C-terminal truncation resulting from the mental-retardation linked R419X nonsense mutation *in vitro* and offer a mechanistic hypothesis for its irresponsiveness to thalidomide. At 1.2Å resolution, our data provide a view of thalidomide binding at atomic resolution.



OPEN ACCESS

Citation: Hartmann MD, Boichenko I, Coles M, Lupas AN, Hernandez Alvarez B (2015) Structural Dynamics of the Cereblon Ligand Binding Domain. PLoS ONE 10(5): e0128342. doi:10.1371/journal.pone.0128342

Academic Editor: Andreas Hofmann, Griffith University, AUSTRALIA

Received: March 1, 2015

Accepted: April 25, 2015

Published: May 29, 2015

Copyright: © 2015 Hartmann et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The structures were deposited in the Protein Data Bank (PDB) under accession codes 5AMH (trigonal), 5AMI (orthorhombic, Wash I), 5AMJ (orthorhombic, Wash II), 5AMK (hexagonal).

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

In 1957, thalidomide was introduced as a potent sedative and anti-nausea drug, and was widely used by pregnant women to alleviate morning sickness. At the end of 1961 it was banned from the market as its catastrophic teratogenic side effects had led to horrific birth defects in more than 10,000 newborns. However, over the decades after its withdrawal, thalidomide was rediscovered as a promising anti-inflammatory, antiangiogenic and immunomodulatory agent with high potential in the treatment of a diverse spectrum of diseases including leprosy, AIDS and cancer. Especially its potential in cancer therapy sparked the search for derivatives with improved properties, which led to the establishment of a new class of immunomodulatory drugs (IMiDs). The two most relevant derivatives are lenalidomide and pomalidomide, which are both approved for the treatment of multiple myeloma.

In 2010, the protein cereblon was identified as a target of thalidomide by Ito et al. [1]. Cereblon was originally found in a genetic screen for mutations linked to mild mental retardation [2], and has been implicated in the modulation of ion channels [3], the regulation of AMP-activated protein kinase [4], and in general neural development [5]. Ito et al. showed that cereblon associates with damaged DNA binding protein 1 (DDB1) to act as a substrate receptor for the

DDB1/cullin4 E3 ubiquitin ligase complex. The ubiquitin ligase activity of this complex is altered by thalidomide binding to cereblon [1], and also lenalidomide and pomalidomide were shown to bind to cereblon, competing for the same binding site [6]. In mammals, cereblon has a LON domain and a C-terminal thalidomide binding domain. However, homologs of the latter, also termed the CULT domain, are further found in animals and bacteria as single-domain proteins, that have all essential residues of the thalidomide binding site conserved [7].

The identification of cereblon as a target for thalidomide has sparked a number of structural studies. Last year, crystal structures of full-length human [8] and chicken [9] cereblon in complex with DDB1 were reported, as well as the structures of the thalidomide binding domain of mouse [8] and that of a bacterial representative, *Magnetospirillum gryphiswaldense* cereblon isoform 4 (MsCI4) [10]. These studies have shown an exceptional conservation of the domain across phyla, with a root-mean-square deviation of less than 1 Å over 100 Cα positions between the bacterial and human proteins. Architecturally, the thalidomide binding domain is a member of the β-tent fold [7], which consists of two antiparallel beta sheets that are oriented at an approximately right angle and pinned together at the top via a structural zinc ion. The thalidomide binding site is formed within the larger, C-terminal β-sheet. In most of the structures it is occupied either by thalidomide, lenalidomide or pomalidomide. These three IMiDs have a glutarimide moiety in common (see also inset in Fig 1). This moiety is bound within an aromatic cage that is mainly formed by three invariant tryptophan residues.

Now, although the thalidomide binding mode is known and invariant among animals and bacteria, evidence for natural ligands of cereblon are rather sparse. With the transcription factors Ikaros and Aiolos, the first target proteins of the cereblon/DDB1/cullin4 E3 ligase have been identified [11, 12]. The degradation of the two proteins is stimulated by thalidomide, lenalidomide or pomalidomide in myeloma cells. However, it remains unclear what exactly cereblon recognizes in absence of these agents. Given the invariant nature of the thalidomide binding site between animals and bacteria, a natural ligand universal to all domains of life seems plausible. Two kinds of such ligands have been proposed so far. One of them was strongly suggested by the nature of the binding pocket. Aromatic cages of similar architecture are typical binding sites of cationic groups like in the quaternary ammonium compounds betaine or choline [13] and are especially widespread in proteins reading posttranslationally modified lysine and arginine residues [14, 15]. However, *in vitro* binding studies of such groups to cereblon were inconclusive [8, 10]. The other potential natural ligand is uridine. Based on the binding mode of the thalidomide glutarimide ring we proposed and verified that cereblon binds uridine in the same way as thalidomide and with comparable affinity; in a co-crystal structure with MsCI4, deoxyuridine mimics the binding mode of thalidomide, and *in vivo*, in a zebrafish model system, uridine caused the same teratogenic effects as thalidomide [10]. However, although it is evidently a universal feature of cereblon, we could not consolidate a biological context for uridine recognition.

In this study, we take another step towards understanding cereblon substrate recognition by analyzing the behavior of the thalidomide binding domain in the unliganded state, based on MsCI4 as a model protein. Via X-ray crystallography, we show the domain in different conformations, and follow the unfolding of a third of the domain upon release of thalidomide in the crystalline state. Underpinned by biophysical data, our results imply that the domain as a whole only becomes structured upon ligand binding. They additionally point to the possibility that the binding site could also assume a different architecture, suitable for the binding of ligands of yet unknown nature. Further, we study the effects of premature chain termination, as caused by the mental-retardation linked R419X nonsense mutation, and find that it does not affect ligand binding. All results, mapped onto the structure of the multi-domain human cereblon, yield an intelligible picture of the structural dynamics of cereblon substrate recognition.

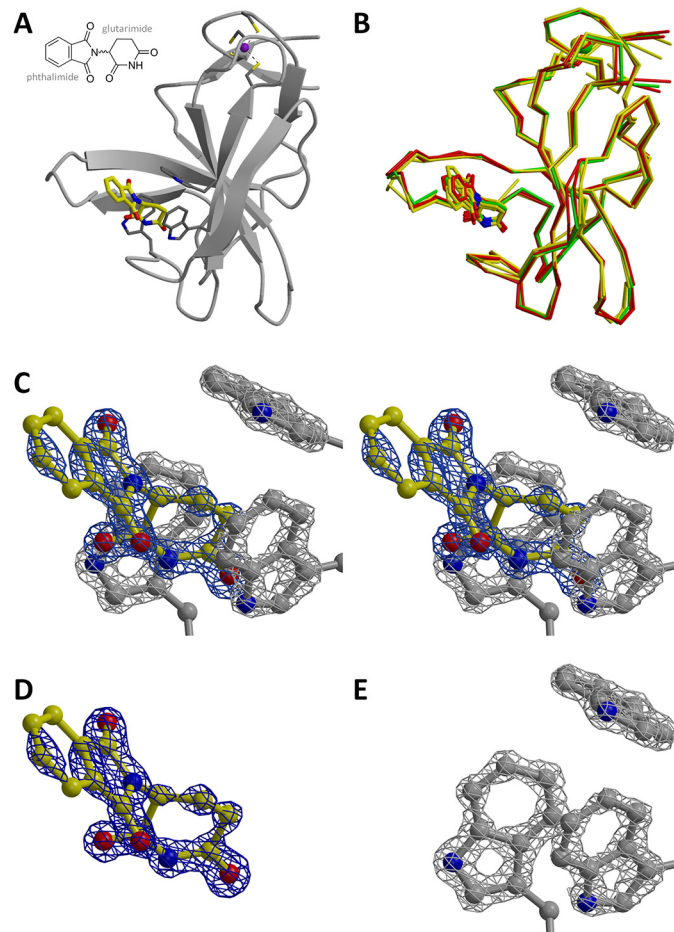


Fig 1. Atomic resolution structure of MsCI4•thalidomide. (A) The overall structure of MsCI4•thalidomide from the trigonal crystal form. (B) Superposition of the latter (in green) with the three MsCI4•thalidomide monomers from the orthorhombic crystal form (4V2Y, in yellow) and the two MsCI4•thalidomide monomers from the hexagonal crystal form (this work, in red). While the conformation of the protein and the glutarimide moiety is virtually identical, the orientation of the phthaloyl moieties differ by up to 13° between the structures. (C) Stereo close-up of the thalidomide molecule and the three tryptophan residues of the aromatic cage of the atomic resolution structure, together with an omit electron density map at 1.2 Å resolution. The omit map was calculated for thalidomide and the three indoles and is contoured at 6.5 sigma. Whereas individual atoms of the indoles and of the glutarimide ring are sharply localized in the density, the phthaloyl moiety is poorly resolved due to thermal disorder. (D) Only thalidomide with omit map. (E) Only the tryptophans with omit map.

doi:10.1371/journal.pone.0128342.g001

Materials and Methods

Cloning, Expression and Purification

Wild-type MsCI4 and mutant MsCI4^{YW/AA} were prepared as described [10]. The mutant MsCI4^{WW/FF} with the substitutions W36F and W59F, and mutant MsCI4^{WWK/FFX} with the additional substitution K115X were cloned with mutagenic primers using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) on the basis of wild-type MsCI4 in pETHis-1a. All proteins were expressed and purified as wild-type MsCI4.

CD spectroscopy

Circular dichroism (CD) measurements were performed using a JASCO J810 spectropolarimeter equipped with a temperature controller. Protein concentrations were 20 μM in 20 mM

Tris, pH 7.6, 50 mM KF. Spectra and melting curves were recorded using a 0.1 cm path length cuvette, a bandwidth of 1 nm, and a response of 1 s. For far-UV CD spectra, a scanning speed of 100 nm/min and a data pitch of 1 nm were used. Thermal denaturation was monitored at 222 nm with a temperature ramp of 1°C per minute and a data pitch of 0.5°C. JASCO software was used for subtraction of buffer baselines, smoothing and determination of melting points.

Fluorescence spectroscopy

Fluorescence was measured using a 2 ml cuvette in a JASCO FP-6500 spectrofluorometer at protein concentrations of 20 μM in 20 mM Tris, pH 7.5, 50 mM KF and optionally ligand concentrations of 200 μM. Samples were excited at 295 nm and emission spectra were recorded using the following settings: data pitch 0.5, excitation bandwidth 1 nm, emission bandwidth 3 nm, response 0.5 s and scan speed 200 nm/min. For data analysis, the JASCO software was used.

Thermal shift assay (Differential scanning fluorimetry)

Ligation dependent thermal stability changes were assayed via differential scanning fluorimetry for wild-type MsCI4 and the binding deficient mutant MsCI4^{YW/AA}. Samples containing 100 μM protein and 400 μM ligand in buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM 2-Mercaptoethanol) supplemented with 50x SYPRO Orange dye (Sigma-Aldrich) were set up in 96-well real-time PCR plates (Thermo Scientific). The sample temperature was continuously ramped up from 25°C to 95°C at a rate of 1.25°C/min in a CFX96TM Real-Time System (Bio-Rad) monitoring the fluorescence of the dye with a temperature resolution of 0.5°C. The experiments were performed in three repeats for MsCI4 and MsCI4^{YW/AA}, and five repeats for MsCI4^{WWK/FFX}. The experiments for thalidomide and thymidine—including the controls without ligand—were performed in presence of 0.4% DMSO. The statistical significance of the shifts caused by the ligands was assessed with a two sample equal variance t-test.

Crystallization, Data collection and Structure determination

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 contained 3 mM thalidomide and 3% (v/v) DMSO in addition to 17 mg/ml of MsCI4 in a buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM 2-Mercaptoethanol. In addition to the described orthorhombic crystal form [10], the screen yielded two further crystal forms, one trigonal and one hexagonal, with the crystallization conditions detailed in Table 1. Crystals of the trigonal form were loop mounted directly from the plate and flash-cooled in liquid nitrogen. For the hexagonal crystal form, crystals were transferred

Table 1. Crystallization conditions and cryo protection / washing procedure.

Crystal form	Reservoir solution (RS)	Cryo/washing solution
Trigonal	100 mM sodium acetate pH 4.6, 2.2 M calcium chloride	-
Orthorhombic (Wash I)	100 mM sodium acetate pH 4.6, 20% (w/v) PEG 6000	RS + 10% (v/v) PEG 300
Orthorhombic (Wash II)	100 mM tri-Sodium citrate pH 5.6, 1.0 M Ammonium phosphate	RS + 20% (v/v) PEG 300
Hexagonal	64 mM Sodium citrate pH 7.0, 100 mM HEPES pH 7.0, 10% (w/v) PEG 5000 MME	RS + 20% (v/v) PEG 300

doi:10.1371/journal.pone.0128342.t001

Table 2. Data collection and refinement statistics.

Crystal form (treatment)	Trigonal	Orthorhombic		Hexagonal
	-	Wash I	Wash II	-
Data collection				
Space group	P3 ₂ 21	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P6 ₁ 22
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	51.7, 51.7, 84.7	57.3, 59.6, 86.7	56.7, 59.7, 88.1	137.9, 137.9, 154.4
Resolution (Å)	30.8–1.20 (1.27–1.20)*	37.3–1.75 (1.86–1.75)*	37.2–1.75 (1.86–1.75)*	39.0–2.90 (3.07–2.90)*
<i>R</i> _{merge}	5.7 (74.4)	4.8 (68.2)	6.6 (47.6)	9.9 (53.2)
<i>I</i> / <i>σ</i>	11.85 (1.78)	16.98 (1.90)	9.79 (2.16)	10.53 (2.22)
Completeness (%)	99.7 (99.3)	99.7 (98.7)	97.7 (96.5)	99.1 (98.3)
Redundancy	3.91 (3.69)	4.23 (4.23)	2.86 (2.79)	2.97 (2.95)
Refinement				
Resolution (Å)	30.8–1.20	37.3–1.75	37.2–1.75	39.0–2.90
No. reflections	39531	29087	28637	18776
<i>R</i> _{work} / <i>R</i> _{free}	0.13 / 0.16	0.17 / 0.21	0.15 / 0.19	0.17 / 0.22
# chains / AU	1	3	3	4
PDB code	5AMH	5AMI	5AMJ	5AMK

*Highest resolution shell is shown in parenthesis.

doi:10.1371/journal.pone.0128342.t002

into a separate drop of cryo-solution prior to flash-cooling, and crystals of the orthorhombic form were washed for 40 h in a cryo-solution as indicated in Table 1. All data were collected at 100 K and a wavelength of 1 Å on a PILATUS 6M detector at beamline PXII of the Swiss Light Source (PSI, Villigen, Switzerland). Diffraction images were indexed, integrated and scaled using XDS [16]. The structures of the washed orthorhombic crystals, which contain 3 monomers of MsCI4 in the asymmetric unit (ASU), were solved on the basis of the MsCI4•thalidomide coordinates 4V2Y; the other two structures were solved by molecular replacement with MOLREP [17], locating one monomer of MsCI4 in ASU of the trigonal crystal form and 4 monomers in the ASU of the hexagonal crystal form. The structures were finalized by cyclic manual modeling with Coot [18] and refinement with REFMAC5 [19]. Data collection and refinement statistics are summarized in Table 2. All molecular depictions were prepared using MolScript [20], BobScript [21], and Raster3D [22]. The structures were deposited in the Protein Data Bank (PDB) under accession codes 5AMH (trigonal), 5AMI (orthorhombic, Wash I), 5AMJ (orthorhombic, Wash II), 5AMK (hexagonal).

NMR Spectroscopy

NMR experiments were performed as described [10]. Significant, characteristic chemical shift changes were associated with binding of 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. Like thalidomide, uridine induced these characteristic chemical shift changes at the lowest concentrations tested. Cytosine and thymidine did not, with concentrations of at least 500 μM. Here, this assays was repeated for the MsCI4^{WVK/FFX} mutant with thalidomide, uridine, thymidine and cytosine, yielding affinities in the same range as for wild-type MsCI4.

Results and Discussion

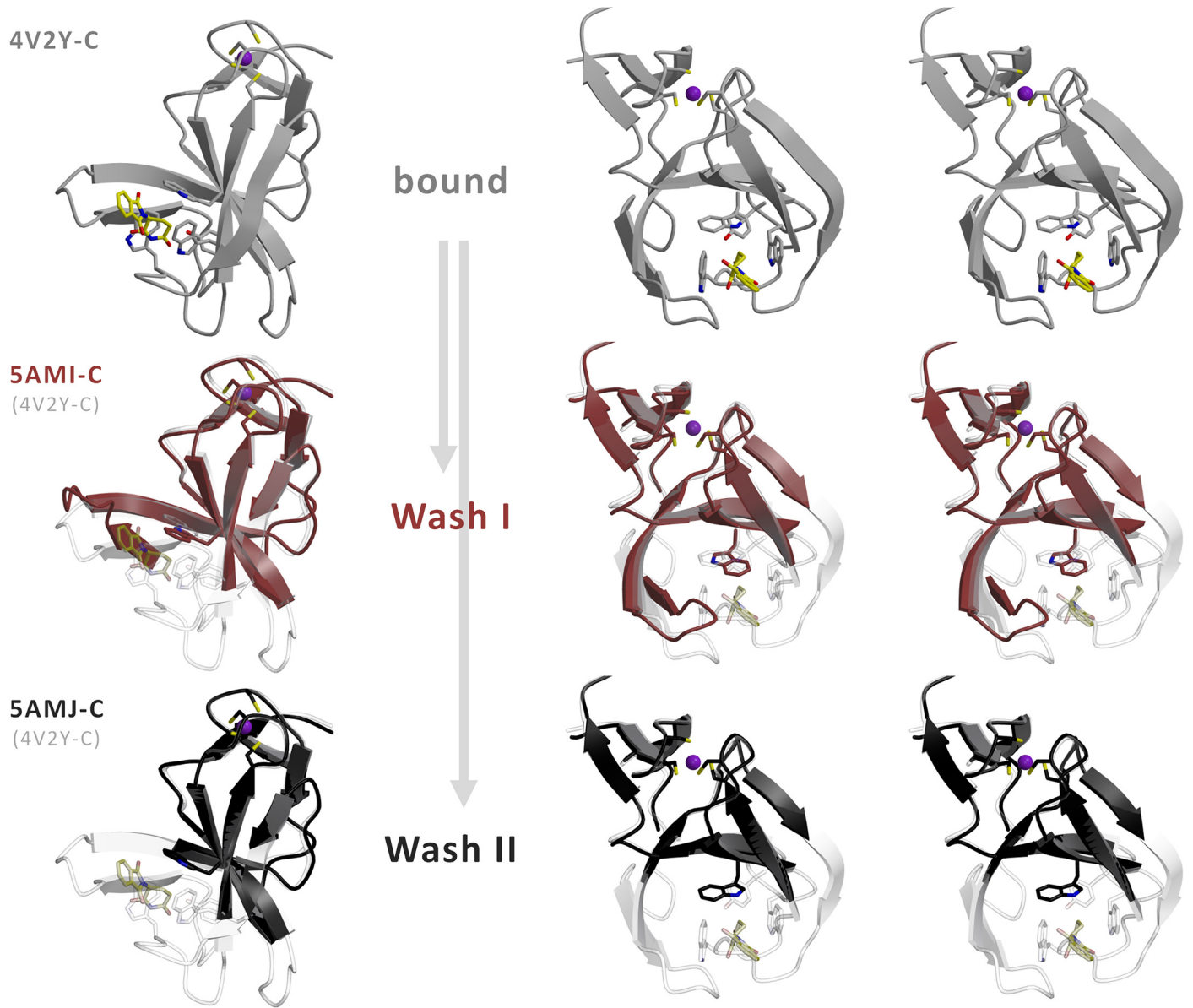
A view of thalidomide binding at atomic resolution

We previously reported the structure of MsCI4 bound to thalidomide from an orthorhombic crystal form [10]. We now obtained a new, trigonal crystal form of MsCI4•thalidomide that yielded a dataset with a resolution of 1.2 Å. The crystals contain one monomer in the asymmetric unit (ASU) in space group P3₂21 (Fig 1A), which has the same conformation as in all other known co-crystal structures of MsCI4. This structure superimposes closely with the 3 other monomers of MsCI4•thalidomide from the orthorhombic crystal form and the 2 further new MsCI4•thalidomide monomers described below. However, between all structures, there is a significant difference in the conformation of the thalidomide molecule (Fig 1B): while the glutarimide moieties superimpose tightly, the phthaloyl moieties protruding from the binding pocket differ in their orientation by up to 13°. This angular freedom is especially apparent in the atomic resolution electron density. In an omit map calculated with missing thalidomide and tryptophan side chains in Fig 1C–1E, the individual atoms of the tryptophan side chains and of the glutarimide moiety are sharply resolved, while the density of the phthaloyl moiety shows much less detail and the most distal part shows no density at the depicted sigma level. This is a consequence of conformational disorder of the phthaloyl ring system, which is also apparent from its anisotropically refined thermal ellipsoids (not shown), which are widened in the plane of the ring system. An obvious rationale for this conformational freedom lies in the lack of specific interactions of the phthaloyl moiety with the protein. A similar degree of freedom can be derived from a superposition of the individual monomers of MsCI4 in complex with pomalidomide, lenalidomide or deoxyuridine. This underlines that there is no selective recognition for any part of the ligand apart from the glutarimide (or uracil) moiety.

Unfolding upon ligand release in the crystalline state

To gain deeper insight into substrate binding, it was highly desirable to know the apo state of the thalidomide binding domain. As we have so far not been able to crystallize MsCI4 in the absence of ligands [10], we tried the following prolonged soaking experiment: We soaked crystals of the original orthorhombic crystal form of MsCI4•thalidomide for many hours in the cryo-protectant solution in order to “wash out” thalidomide from the binding sites. Diffraction experiments were performed with crystals from two different crystallization conditions after 40 h of washing. In the two resulting structures, thalidomide was retained in two of the three monomers in the ASU, but released in the third monomer. Strikingly, the loss of the ligand was accompanied by the unfolding of a large portion of the protein around the binding site. In one of the structures, “Wash II”, three regions totaling to 37% of the previously structured part of the main chain were no longer traceable in the electron density. The largest is the most N-terminal unfolded region, which starts between β 2 and β 3 and comprises the hairpin formed by β 3 and β 4, leaving only a C-terminal remainder of β 4. The second region starts in the middle of β 5, comprises the first tryptophan of the aromatic cage, and ends at the cage’s second tryptophan at the start of β 6, rendering its indole side chain disordered. The third and smallest region is the loop connecting β 7 and β 8. In the other structure, “Wash I”, the second and third region were unfolded to the same extent, whereas the β -hairpin in the first region was still mostly structured, albeit shifted in position. The unfolding process is detailed in Fig 2.

Of the three unfolded regions, the first one does not form specific interactions with thalidomide but stabilizes the geometry of the aromatic cage and thus of the second region. In particular, it fixes the first tryptophan (W79) of the cage in its conformation by forming a hydrogen bond with the indole-side chain [10]. It further contains a highly conserved NPxG motif at the



Mg (17) GASIFRCRQCQGT-ISRRLWLLPMGGD-HEHVVENPAGMIFRVCVFLAQGLRLIGAPSGEFSWFKGYDWTIALCGQCGSHLGWHYEG---GSQPQTFFGLIKDRLAEGPAD
 Hs (317) CTSLCCRQCQETEITTKNEIFSLSLCGPMAYVNPFGYVHETLTVYKACNLNLIGRSTESHSWFPGYAWTVACCKICASHIGWKEFATKKDMSPOKFWGLTRRSALLPTI (16)

Fig 2. Partial unfolding of Msc14 upon loss of thalidomide. Crystals of the orthorhombic crystal from two crystallization conditions were washed for 40 h in a solution without thalidomide. For the initial structure (grey) and both experiments, “Wash I” (brown) and “Wash II” (black), the monomer that had the ligand washed out (chain C) is depicted in two perspectives. Loss of ligand was accompanied by the unfolding of three regions. The structures after washing are overlaid with the initial structure in transparent. On the bottom, the sequence alignment of Msc14 with human cereblon details the unfolded regions and indicates secondary structure elements. The unfolded regions are shaded. In “Wash I”, the unfolding of the first region is incomplete, so the hairpin formed by the 3rd and 4th β -strand is still folded, albeit shifted in position. The C-terminal segment deleted in the R419X mutant of human cereblon and the corresponding segment in Msc14 are underlined. Key-residues are highlighted red.

doi:10.1371/journal.pone.0128342.g002

tip of the β 3- β 4 hairpin [7]. The second region constitutes the largest part of the thalidomide binding site, contributing two of the three tryptophans of the aromatic cage (W79 and W85) and forming hydrogen bonds to the glutarimide moiety [10]. Moreover, it contains the tyrosine

residue (Y83) that was found to have an impact on cereblon function in the thalidomide-binding deficient mutant hCrbn^{Y384A} [1]. This conserved tyrosine forms hydrogen bonds with the third region and is part of a common hydrophobic core of the second and third region. It is therefore conceivable that the two regions fold and unfold in a concerted manner and that the third region, which does not directly contribute to thalidomide binding, serves to stabilize the binding site in the bound conformation. Hypothetically, an MsCI4^{Y83A} substitution corresponding to hCrbn^{Y384A} would disrupt the common hydrophobic core and therewith the mutual stabilizing interactions.

We note that in this experiment the unfolding happened within a given crystal packing, which per se restrains the conformational freedom. It would therefore be conceivable, that the extent of unfolding was limited by these restraints. However, the regions found flexible here are in perfect agreement with those identified in the following in further crystal forms of MsCI4 and mouse cereblon.

A crystal structure with intertwined MsCI4 monomers in different conformations

In addition to the orthorhombic and the high-resolution trigonal crystal form, we obtained a hexagonal crystal form in a co-crystallization screen with thalidomide; we solved the structure of this third crystal form by molecular replacement, locating four monomers in the ASU. Surprisingly, only two of them are in the known thalidomide-bound conformation—the other two form an intertwined dimer and have no thalidomide bound. The two monomers of this dimer, depicted blue and pink in Fig 3, are in grossly different conformations: The blue monomer adopts an overall conformation comparable to the thalidomide-bound state but has the β 3- β 4 hairpin shifted in position. Therefore, the stabilizing hydrogen bond between the hairpin and the first tryptophan (W79) of the cage is broken, and the latter is found flipped out of the aromatic cage. In contrast, the conformation of the pink monomer is very different. The binding pocket is not formed and essentially all regions previously identified to be flexible are dramatically different from the thalidomide-bound conformation: In the first flexible region, the loop connecting β 2 and β 3 is disordered to a similar extent as in the “Wash I” structure, while the β 3- β 4 hairpin is shifted in position and even has another strand register. The whole second flexible region is found in an entirely different conformation—the amplitude of its largest displacement, at W79, is 30Å. Consequently, as also Y83 is not in place to anchor the second to the third region, the third flexible region is actually found in two alternative conformations, both different from the thalidomide-bound state.

Intriguingly, the W79 side chain of the pink monomer reaches over to the binding site of the blue monomer to complete the aromatic cage (Fig 3B). At the later stages of refinement, electron density of unknown origin became apparent within this cage, which was convincingly modeled as a DMSO molecule from the thalidomide stock solution. Sterically, the flipping-out of the W79 side chain from the aromatic cage of the blue monomer was a prerequisite to provide enough volume for the accommodation of DMSO. Although this particular dimeric conformation and the DMSO ligand are seemingly highly artificial, this brings to our attention that the natural ligand might be structurally quite different from thalidomide, for which the binding pocket might also adopt a conformation that is unexpectedly different from the thalidomide-bound one.

When we superimpose all available conformations of MsCI4, the thalidomide-bound one, the ones after the washing experiment, and the ones of the two monomers in the intertwined dimer, we obtain the ensemble depicted in Fig 4. Therein we find a minimal invariant consensus structure that exactly coincides with the structured part of the “Wash II” structure: While

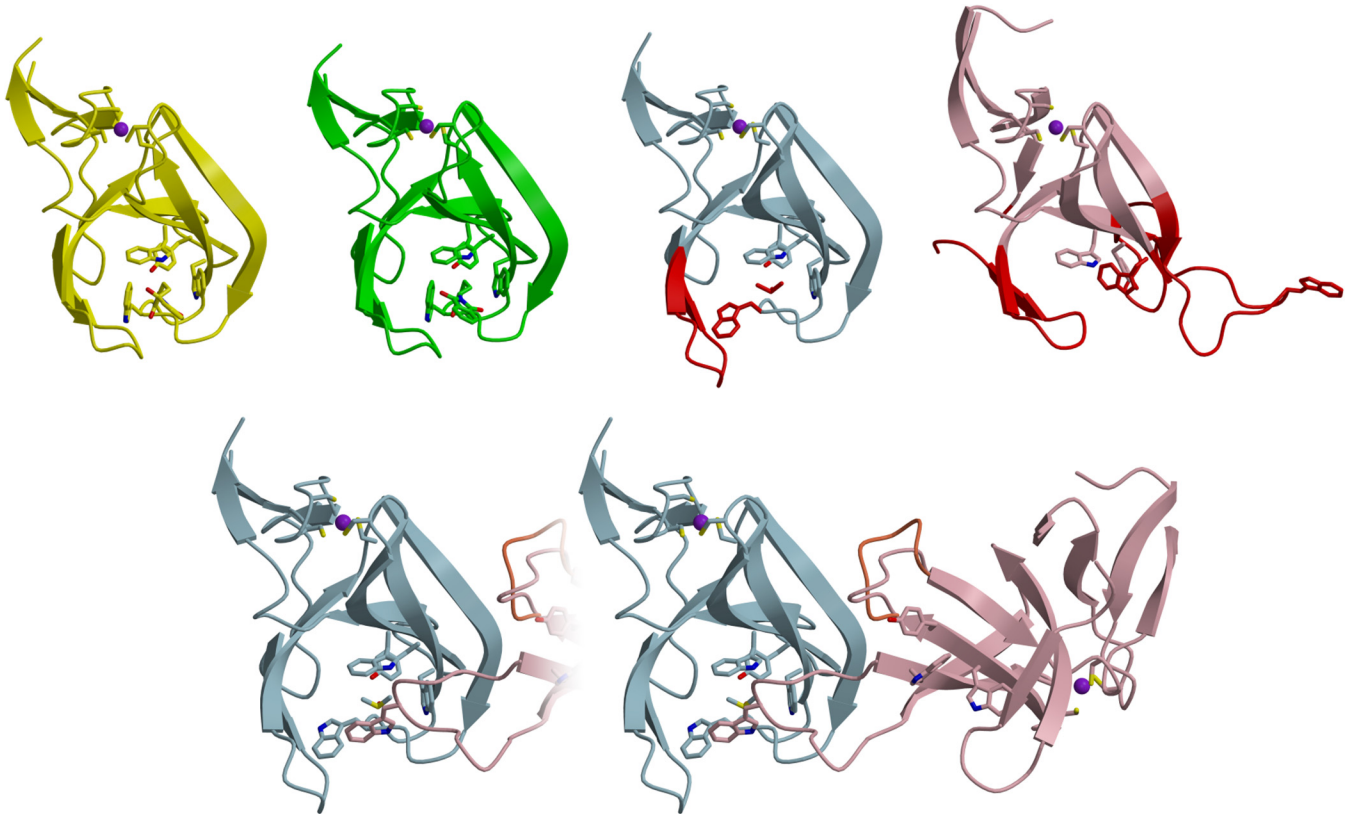


Fig 3. MsCI4 in different conformations in the hexagonal crystal form. Top: The 4 monomers in the ASU are found in three different conformations. The yellow and green monomer are in the known thalidomide-bound conformation. The blue monomer is in a conformation similar to the thalidomide-bound one, but its binding site is occupied by a DMSO molecule, which is accompanied with a displacement of one tryptophan of the aromatic cage (red). The pink monomer is in an overall distorted conformation and does not have the aromatic cage formed: the β 3- β 4 hairpin has another strand register, one of the tryptophans is flipped, and another tryptophan is displaced by 30Å (all in red). A second conformation of the 3rd flexible region is displayed in another shade of red. Bottom: The blue and the pink monomers form an intertwined dimer, in which the displaced tryptophan of the pink monomer completes the aromatic cage of the blue one, trapping a DMSO molecule in a binding site with a modified architecture. Here, the second conformation of the 3rd flexible region of the pink monomer is in orange.

doi:10.1371/journal.pone.0128342.g003

the three flexible regions derived from the washing experiment are found in a multitude of conformations, the remainder of the protein is essentially invariant in all structures.

Folding upon ligand binding in solution

After consolidating structural evidence for immanent flexibility of the thalidomide binding domain, we set out to study the effects of ligand binding in solution. Previously, we had established an NMR-based ligand binding assay that relies on specific chemical shift changes [10]. Comparison of spectra of wild-type MsCI4 alone and in presence of thalidomide revealed significant 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. Chemical shift changes of this extent can be explained by close contact with aromatic moieties. However, examination of the MsCI4 binding site revealed no candidate methyl groups in contact with the aromatic rings of thalidomide itself. This indicates that ligand binding is associated considerable conformational rearrangements in residues not directly in contact with the ligand. This prompted us to further investigate the effects of ligand binding by further spectroscopic approaches.

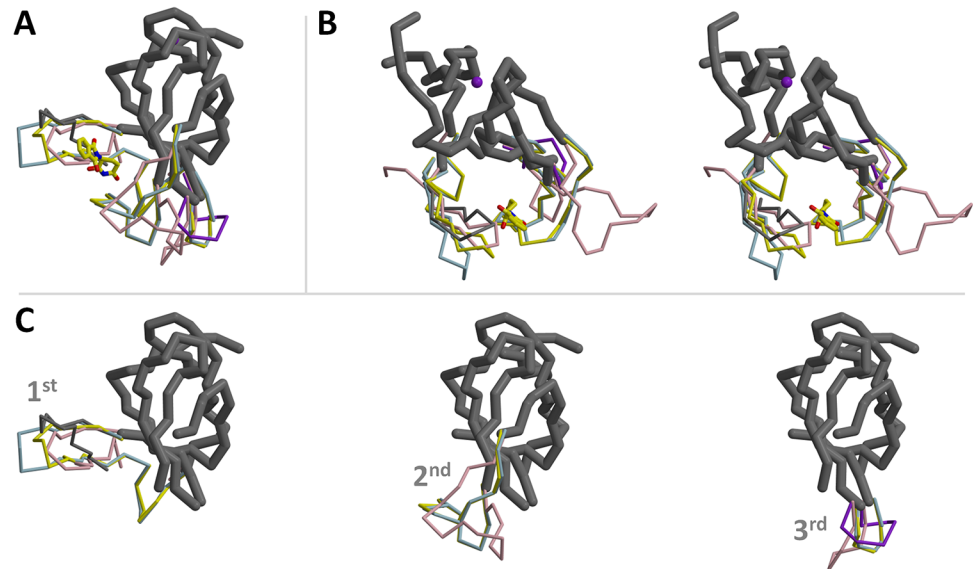


Fig 4. An ensemble of the available MsCI4 conformations. (A) Superposition of the thalidomide-bound conformation (yellow), the blue and pink monomers from Fig 3, the “Wash I” structure (thin dark grey), and the “Wash II” structure (thick dark grey). One alternate conformation of the pink monomer in the 3rd flexible region is depicted in purple. (B) same as (A), but from the other perspective and in stereo. (C) The consensus (Wash II) structure together with the individual ensembles of the 1st, 2nd and 3rd flexible regions.

doi:10.1371/journal.pone.0128342.g004

Firstly, we analyzed the thermal melting behavior of the protein in absence and presence of ligands. As judged by CD spectroscopy, apo MsCI4 is at least partially folded: the far-UV spectrum is indicative of secondary structure (Fig 5A), and a melting curve shows a thermal transition at about 70°C (Fig 5B). The effect of ligand binding was examined in a thermal shift assay with the known binders thalidomide and uridine, as well as the non-binders thymidine and cytidine. If the folding of the flexible regions was dependent on—and stabilized by—ligand binding, this effect should be manifest in increased thermal stability. Indeed, while thymidine and cytidine had no influence on MsCI4, thalidomide and uridine had a stabilizing effect, increasing the melting point by about 1°C. As a control, binding deficient MsCI4^{YW/AA} remained virtually unaffected by all ligands (Fig 5C).

Secondly, we monitored ligand binding via tryptophan fluorescence. For increased sensitivity we devised the mutant MsCI4^{WW/FF}, in which we exchanged the tryptophan residues outside the binding site, W36 and W59, to phenylalanine. The only remaining tryptophan residues are the three of the aromatic cage, of which two became disordered in the washing experiment. An inspection of MsCI4^{WW/FF} via CD spectroscopy yielded a far UV spectrum and melting behavior comparable to the wild-type protein (Fig 5A and 5B). Finally, fluorescence emission spectra were recorded in presence of uridine, thymidine and cytidine as well as in absence of ligands. The emission maximum of apo MsCI4^{WW/FF} was found at 355 nm, which is indicative for mostly solvent exposed indole side chains, and was not shifted in presence of thymidine and cytidine. In presence of uridine, the maximum was however shifted to shorter wavelength, which can be explained by a shielding from the solvent and hence folding of the binding site (Fig 5D).

The mental retardation-linked nonsense mutation does not affect thalidomide binding

As MsCI4 has proven a robust and reliable model system for which we have a set of useful spectroscopic techniques at hand, we employed it in a yet different scope: We reproduced the

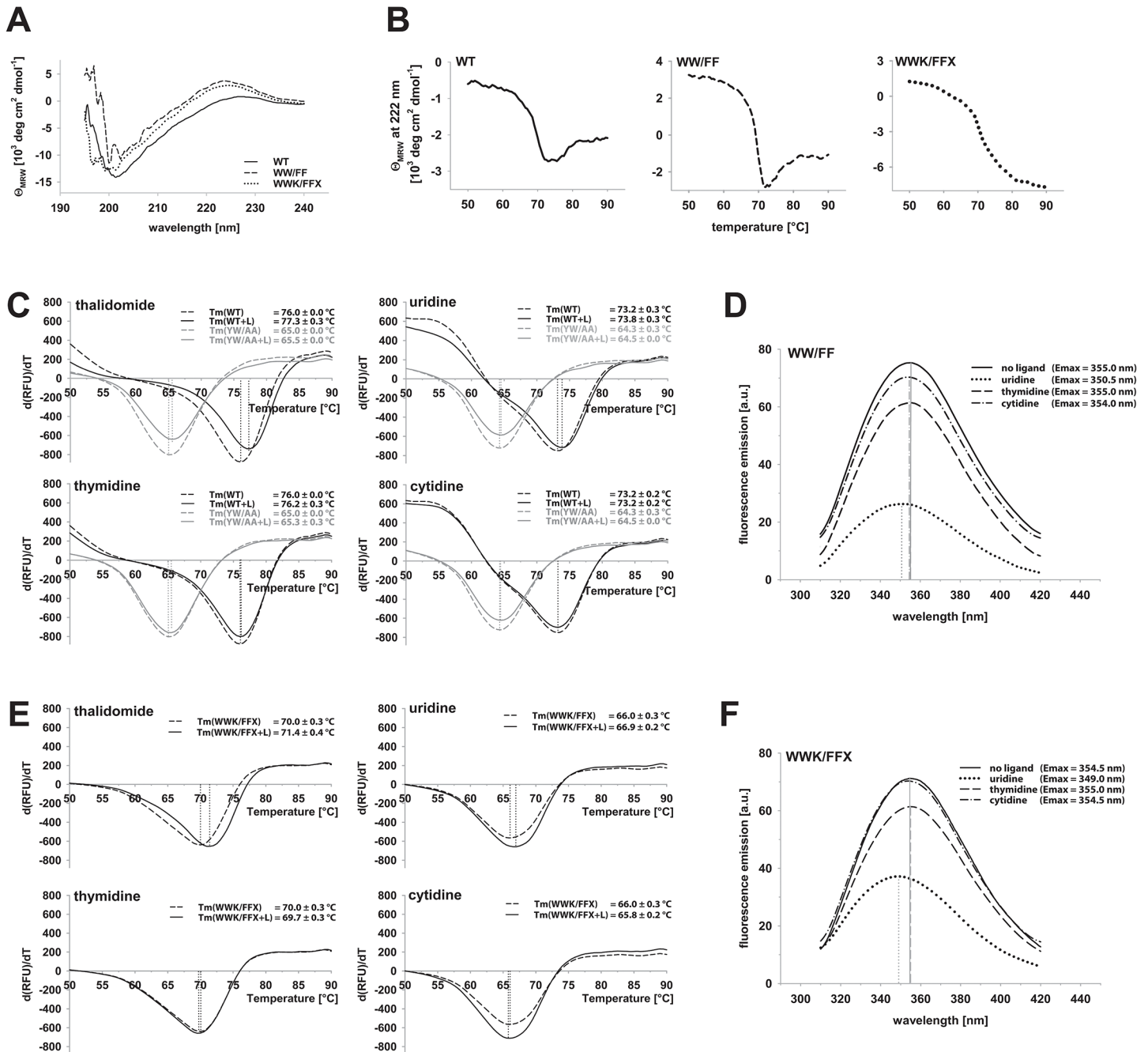


Fig 5. Biophysical characterization of MSCI4 and mutants. (A) Far-UV CD spectra of wild-type (WT) MSCI4 and mutants MSCI4^{WW/FF} and MSCI4^{WWK/FFX}, and (B) corresponding melting curves, monitored at 222 nm. (C) Thermal stability of MSCI4 and binding deficient MSCI4^{YW/AA} in presence of thalidomide, uridine, thymidine and cytidine, as determined in the thermal shift assay. Only thalidomide and uridine lead to a shift in melting temperature with (two-tail) p-values of 1.5e-3 and 4.9e-2 as determined in a two sample equal variance t-test. (D) Tryptophan fluorescence spectra of MSCI4^{WW/FF} in presence of uridine, thymidine and cytidine. The emission maximum is shifted to shorter wavelength only in presence of the binder uridine, which can be indicative for the folding of the binding site. Note also the pronounced quenching effect indicative for binding. (E) same as (C) for the MSCI4^{WWK/FFX} mutant. As for the WT, thalidomide and uridine lead to a thermal shift, with p-values of 4.0e-4 and 1.2e-3. (F) same as (D) for the MSCI4^{WWK/FFX} mutant. The presence of uridine has the same effect as on the WT.

doi:10.1371/journal.pone.0128342.g005

mental retardation-linked hCRBN^{R419X} nonsense mutation in MSCI4 with the corresponding K115X substitution for *in vitro* characterization. With tryptophan fluorescence studies in

mind, we directly introduced the substitution in the MsCI4^{WW/FF} background, yielding the mutant MsCI4^{WWK/FFX}.

The R419X mutation in hCRBN renders the protein irresponsive to thalidomide—The overall ubiquitination behavior of the E3 ligase however seems unaffected [23]. With R419 being located directly at the end of the penultimate β -strand, the premature stop effectively leads to the deletion of the last strand of the N-terminal β -sheet. It was therefore unclear whether the domain was still properly folded, which is the first question we wanted to answer with MsCI4^{WWK/FFX}. After the mutant's behavior during expression and purification was similar to the wild-type, we examined it via CD spectroscopy. The resulting far-UV spectrum was comparable to MsCI4 and MsCI4^{WW/FF} and—strikingly—the melting point was almost identical (Fig 5A and 5B). Consequently, the premature stop had seemingly no impact on the stability of the domain, so we examined its ligand binding abilities.

In analogy to MsCI4 and MsCI4^{WW/FF}, we studied MsCI4^{WWK/FFX} in a thermal shift assay and via tryptophan fluorescence spectroscopy in the presence of thalidomide, uridine, thymidine and cytosine. Essentially, all results were as for full-length MsCI4: Thalidomide and uridine yielded an increase in thermal stability of about 1°C, and uridine caused a shift of the tryptophan fluorescence emission maximum to shorter wavelength, while thymidine and cytosine had no effect (Fig 5E and 5F). Therefore, ligand binding is unaffected by the premature stop, which we ultimately verified in our NMR ligand binding assay: The affinities for both thalidomide and uridine lie in the low micromolar range as determined for the wild-type protein. Consequently, the phenotype of thalidomide irresponsive hCRBN^{R419X} is apparently not due to a defect in ligand binding.

Structural comparison to mouse cereblon

Further support for our findings on the structural dynamics is provided by the available crystal structures of the thalidomide binding domain of mouse cereblon [8]. Here, crystal structures originating from two different crystal forms show the thalidomide binding domain in several different conformations, as depicted in Fig 6.

The first crystal form contains four thalidomide-bound monomers in the ASU which form two pairs of interlaced molecules. In these pairs, the interface is formed by the thalidomide binding sites and thalidomide molecules, such that the latter form contacts with the aromatic-cages of both monomers. As a consequence, the β 3- β 4 hairpin cannot fold and remains mostly disordered. In this region, the otherwise identical four monomers of the ASU differ from each other: the region is disordered to different extents and the traceable parts are found in different conformations. The whole range of this flexible region matches the first one of MsCI4. Despite the missing stabilizing interaction between the β 3- β 4 hairpin and the first tryptophan of the cage, the binding site is in the usual thalidomide-bound conformation. This is facilitated by contacts to the thalidomide molecule of the other chain in the interlaced pairs. The neighboring thalidomide molecule forms a hydrogen bond with the tryptophan, which substitutes for the missing bond with the hairpin, and stabilizes the indole side chain in the bound conformation.

The second crystal form contains two monomers without ligands in the ASU. They adopt a conformation similar to the thalidomide-bound one. However, these two monomers also do not form the β 3- β 4 hairpin: the whole segment corresponding to the first flexible region is found in varying conformations. Via hydrophobic side chains from that region, the molecules stabilize the binding sites of neighboring monomers: While monomer A has the side chain of M349 stuck into the opening of the aromatic cage of monomer B, monomer B has the side chain of P348 stuck into the opening of the aromatic cage of another monomer A. In this fashion they form an endless array of intertwined molecules throughout the crystal. However, as

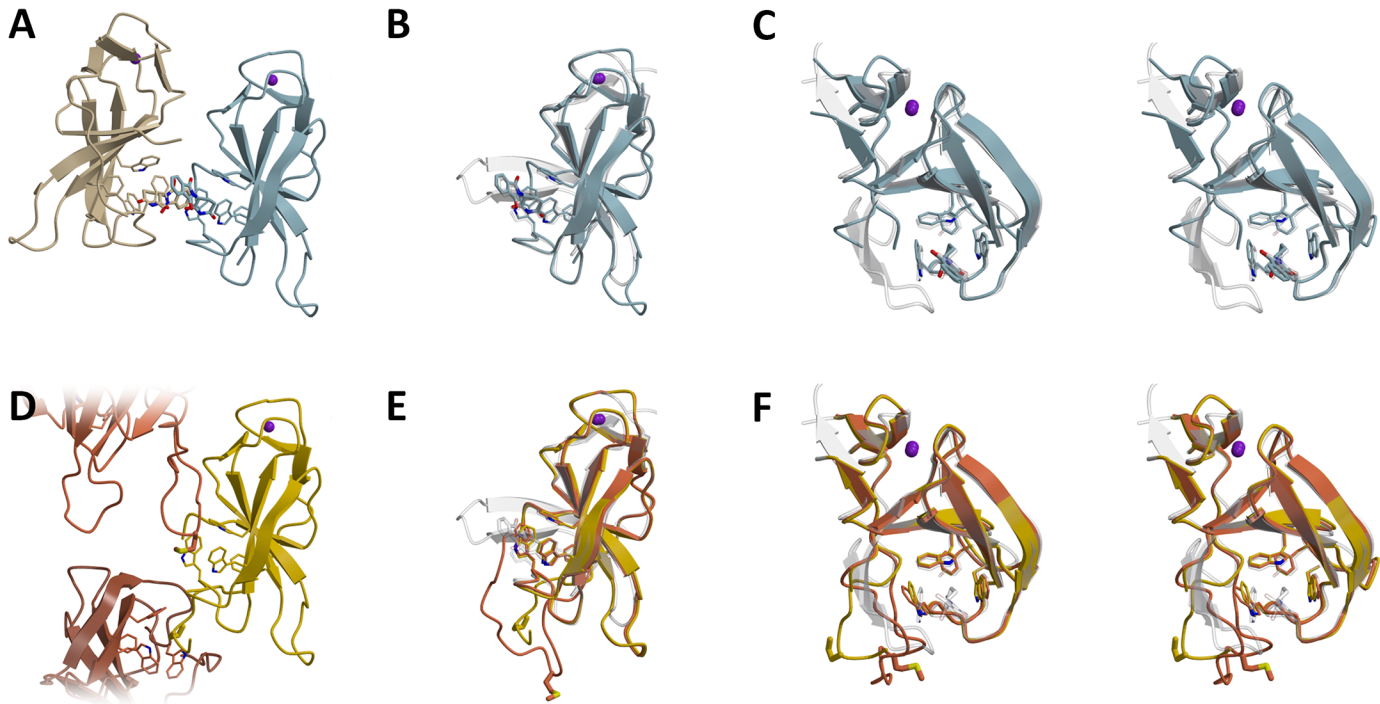


Fig 6. The thalidomide binding domain of mouse cereblon exhibits flexibility comparable to MsCI4. The mouse domain in complex with (blue, sand) and without thalidomide (yellow, shades of brown) is compared to MsCI4•thalidomide (transparent white). (A) Two thalidomide-bound domains from 4TZC arranged as an intertwined dimer. (B) Superposition of one monomer to MsCI4•thalidomide, illustrating the unfolded nature of the first flexible region. (C) Same as (B) but from another perspective, in stereo. (D) Apo mouse domains in 3WX2 are found in two conformations, “yellow” and “brown”, forming an endless array of interactions in the crystal lattice. (E) Superposition of both apo conformations onto MsCI4•thalidomide, showing the first flexible region in different conformations. (F) Same as (E) but from another perspective, in stereo. Note that the conformation of the tryptophans on the left is reminiscent of the flipped tryptophan in the intertwined MsCI4 dimer.

doi:10.1371/journal.pone.0128342.g006

the stabilizing hydrogen bond between the first tryptophan of the cage and the $\beta 3$ - $\beta 4$ hairpin is lost and not substituted, the geometry of the aromatic cage is impaired. In analogy to the intertwined dimer of MsCI4 in the hexagonal crystal form, the corresponding tryptophan deviates from the thalidomide-bound conformation.

Although the conformations found in both mouse crystal forms are highly artificial, they emphasize that the flexible nature of the binding site as identified and characterized in MsCI4 is indeed not a special feature of the bacterial representative but of general significance, also for animal cereblon. As in the intertwined MsCI4 dimer, the conformation of the unliganded mouse structure is only brought about by the crystal packing.

The thalidomide binding domain embedded in the E3 ligase complex

To understand their functional relevance, we mapped the flexible regions onto the structure of human full-length cereblon (Fig 7). This structure [8] is bound to lenalidomide and is in a virtually identical conformation to the structures of chicken cereblon in complex with thalidomide, lenalidomide or pomalidomide [9]. Therein, the thalidomide binding domain is packed tightly to the N-terminal LON domain. A large portion of the interface between the domains is formed by the first flexible region. Further, the LON domain has a short N-terminal extension, which folds along the thalidomide binding site. As this extension, which was initially predicted to be unstructured, is exclusively in contact with the flexible regions, it seems plausible that its folding is concerted with the folding of the binding site; upon substrate binding, it would fold

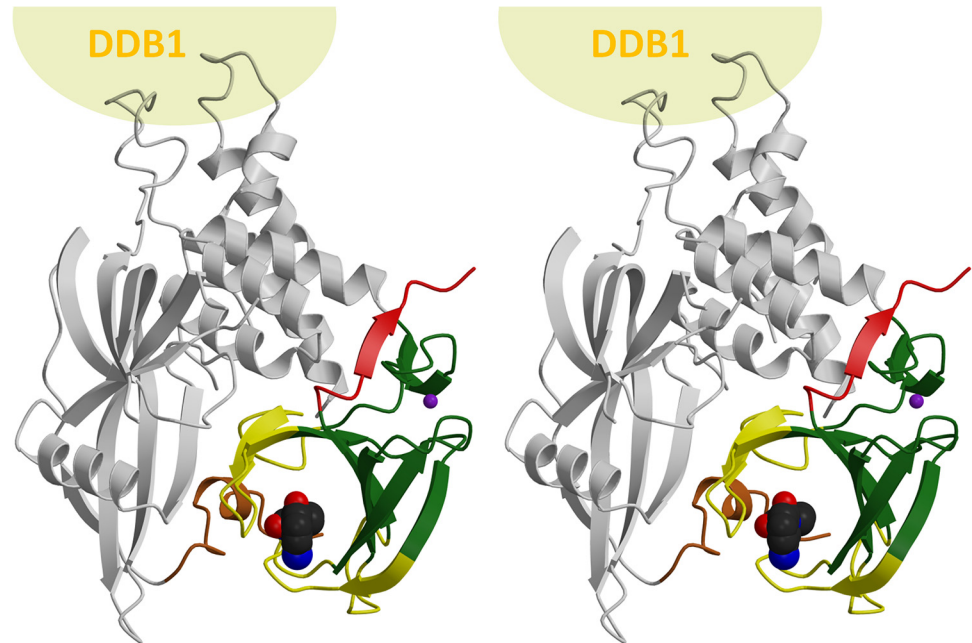


Fig 7. The flexible regions and mental retardation mutation in the context of human full length cereblon. At the top, interaction partner DDB1 is represented by a sphere. The thalidomide binding domain is in green (rigid core), yellow (flexible regions) and red (deleted C-terminal part in the MsC14^{WWK/FFX} mutant). The exact boundaries of the segments are defined in Fig 2. The N-terminal extension of cereblon is in brown, the remainder of the protein grey.

doi:10.1371/journal.pone.0128342.g007

together with the binding site to stabilize the bound conformation, and possibly anchor the domains in their relative orientations. As determined by the PISA web server (<http://www.ebi.ac.uk/pdbe/pisa/>), the interface area between the domains in the fully folded, ligand-bound crystal structure amounts to 1596 Å².

Reversely, in the apo state—in which the binding site would not be formed and the extension would be potentially unfolded—the interface between the domains would be significantly smaller, possibly allowing for conformational flexibility in the relative orientation of the domains: When the flexible regions are fully excluded from the calculation, the interface area between the domains is reduced by a factor of 2.6, to 609 Å². A central part of this reduced interface is the region that is deleted in the retardation-linked hCRBN^{R419X} nonsense mutant (Fig 7), which mediates mostly hydrophobic contacts between the domains. When this region is omitted as well, the interface area is further reduced from 609 Å² to 363 Å². Hence, it is likely that the interface, and thereby the alignment of domains, is disrupted in the mutant. As we have shown that the R419X mutation does not affect ligand binding, it seems plausible that its irresponsiveness to thalidomide is due to the impaired inter-domain interface: With the thalidomide binding domain unable to dock correctly in the bound conformation, cereblon is not a functional receptor of the E3 ligase complex.

Conclusions

This work has given insight into thalidomide binding at atomic resolution and revealed an unexpected degree of immanent structural flexibility of the thalidomide binding domain. We identify three interconnected flexible regions that fold only upon substrate binding, accounting for a third of the whole domain. A previously identified important tyrosine residue is found within the hydrophobic interface in the folded state of these regions, which explains the

binding deficiency of the hCrbn^{Y384A} mutant. As deduced from the available crystal structures, the binding site can fold into a geometry deviating from the thalidomide bound one. This could imply yet further, different classes of natural ligands, possibly tertiary ammonium groups which are geometrically reminiscent of the DMSO molecule bound in one of the structures and which could be bound via cation- π interactions within the aromatic cage. Moreover, our results imply that within the E3 ligase complex, substrate binding to the thalidomide binding domain is only properly recognized when the domain is correctly docked to the remainder of the cereblon protein—which is potentially corrupted in the mental retardation linked hCrbn^{R419X} mutant. In summary, our results advance the structural basis of substrate recognition beyond a static picture and provide a mechanistic framework for the understanding of the functional role of cereblon.

Acknowledgments

We thank Silvia Deiss and Kerstin Bär for technical assistance in sample preparation, Reinhard Albrecht for crystallographic data collection, and Stefanie Jonas for assistance with the thermal shift assay. We are continuously grateful to the staff of the PX beamlines at the Swiss Light Source for excellent technical support.

Author Contributions

Conceived and designed the experiments: MDH BHA. Performed the experiments: MDH IB MC BHA. Analyzed the data: MDH IB MC ANL BHA. Wrote the paper: MDH.

References

1. Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, et al. Identification of a primary target of thalidomide teratogenicity. *Science*. 2010; 327(5971):1345–50. Epub 2010/03/13. doi: [10.1126/science.1177319](https://doi.org/10.1126/science.1177319) PMID: [20223979](https://pubmed.ncbi.nlm.nih.gov/20223979/).
2. Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP. A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology*. 2004; 63(10):1927–31. Epub 2004/11/24. PMID: [15557513](https://pubmed.ncbi.nlm.nih.gov/15557513/); PubMed Central PMCID: PMC1201536.
3. Liu J, Ye J, Zou X, Xu Z, Feng Y, Zou X, et al. CRL4A(CRBN) E3 ubiquitin ligase restricts BK channel activity and prevents epileptogenesis. *Nature communications*. 2014; 5:3924. Epub 2014/05/23. doi: [10.1038/ncomms4924](https://doi.org/10.1038/ncomms4924) PMID: [24845235](https://pubmed.ncbi.nlm.nih.gov/24845235/).
4. Lee KM, Jo S, Kim H, Lee J, Park CS. Functional modulation of AMP-activated protein kinase by cereblon. *Biochimica et biophysica acta*. 2011; 1813(3):448–55. Epub 2011/01/15. doi: [10.1016/j.bbamcr.2011.01.005](https://doi.org/10.1016/j.bbamcr.2011.01.005) PMID: [21232561](https://pubmed.ncbi.nlm.nih.gov/21232561/).
5. Chang XB, Stewart AK. What is the functional role of the thalidomide binding protein cereblon? *International journal of biochemistry and molecular biology*. 2011; 2(3):287–94. Epub 2011/10/18. PMID: [22003441](https://pubmed.ncbi.nlm.nih.gov/22003441/); PubMed Central PMCID: PMC3193296.
6. Lopez-Girona A, Mendy D, Ito T, Miller K, Gandhi AK, Kang J, et al. Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia*. 2012; 26(11):2326–35. doi: [10.1038/Leu.2012.119](https://doi.org/10.1038/Leu.2012.119) PMID: [25569776](https://pubmed.ncbi.nlm.nih.gov/25569776/); ISI:000310791300004.
7. Lupas AN, Zhu H, Korycinski M. The thalidomide-binding domain of cereblon defines the CULT domain family and is a new member of the beta-tent fold. *PLoS computational biology*. 2015; 11(1):e1004023. Epub 2015/01/09. doi: [10.1371/journal.pcbi.1004023](https://doi.org/10.1371/journal.pcbi.1004023) PMID: [25569776](https://pubmed.ncbi.nlm.nih.gov/25569776/); PubMed Central PMCID: PMC4287342.
8. Chamberlain PP, Lopez-Girona A, Miller K, Carmel G, Pagarigan B, Chie-Leon B, et al. Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nature structural & molecular biology*. 2014; 21(9):803–9. Epub 2014/08/12. doi: [10.1038/nsmb.2874](https://doi.org/10.1038/nsmb.2874) PMID: [25108355](https://pubmed.ncbi.nlm.nih.gov/25108355/).
9. Fischer ES, Bohm K, Lydeard JR, Yang H, Stadler MB, Cavadini S, et al. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature*. 2014; 512(7512):49–53. Epub 2014/07/22. doi: [10.1038/nature13527](https://doi.org/10.1038/nature13527) PMID: [25043012](https://pubmed.ncbi.nlm.nih.gov/25043012/).

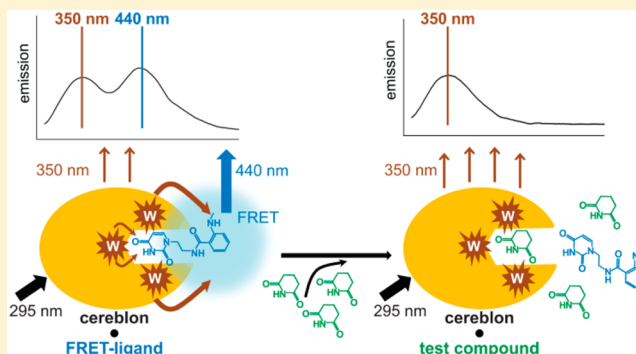
10. Hartmann MD, Boichenko I, Coles M, Zanini F, Lupas AN, Hernandez Alvarez B. Thalidomide mimics uridine binding to an aromatic cage in cereblon. *Journal of structural biology*. 2014; 188(3):225–32. Epub 2014/12/03. doi: [10.1016/j.jsb.2014.10.010](https://doi.org/10.1016/j.jsb.2014.10.010) PMID: [25448889](https://pubmed.ncbi.nlm.nih.gov/25448889/).
11. Gandhi AK, Kang J, Havens CG, Conklin T, Ning Y, Wu L, et al. 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.). *British journal of haematology*. 2014; 164(6):811–21. Epub 2013/12/18. doi: [10.1111/bjh.12708](https://doi.org/10.1111/bjh.12708) PMID: [24328678](https://pubmed.ncbi.nlm.nih.gov/24328678/); PubMed Central PMCID: [PMC4232904](https://pubmed.ncbi.nlm.nih.gov/PMC4232904/).
12. Kronke J, Udeshi ND, Narla A, Grauman P, Hurst SN, McConkey M, et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science*. 2014; 343(6168):301–5. Epub 2013/12/03. doi: [10.1126/science.1244851](https://doi.org/10.1126/science.1244851) PMID: [24292625](https://pubmed.ncbi.nlm.nih.gov/24292625/); PubMed Central PMCID: [PMC4077049](https://pubmed.ncbi.nlm.nih.gov/PMC4077049/).
13. Dougherty DA. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science*. 1996; 271(5246):163–8. Epub 1996/01/12. PMID: [8539615](https://pubmed.ncbi.nlm.nih.gov/8539615/).
14. Gayatri S, Bedford MT. Readers of histone methylarginine marks. *Biochimica et biophysica acta*. 2014; 1839(8):702–10. Epub 2014/03/04. doi: [10.1016/j.bbagr.2014.02.015](https://doi.org/10.1016/j.bbagr.2014.02.015) PMID: [24583552](https://pubmed.ncbi.nlm.nih.gov/24583552/); PubMed Central PMCID: [PMC4099268](https://pubmed.ncbi.nlm.nih.gov/PMC4099268/).
15. Yun MY, Wu J, Workman JL, Li B. Readers of histone modifications. *Cell Res*. 2011; 21(4):564–78. doi: [10.1038/Cr.2011.42](https://doi.org/10.1038/Cr.2011.42) PMID: [19000004](https://pubmed.ncbi.nlm.nih.gov/19000004/).
16. Kabsch W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr*. 1993; 26:795–800. doi: [10.1107/S0021889893005588](https://doi.org/10.1107/S0021889893005588) PMID: [19930007](https://pubmed.ncbi.nlm.nih.gov/19930007/).
17. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D*. 2000; 56:1622–4. doi: [10.1107/S0907444900013780](https://doi.org/10.1107/S0907444900013780) PMID: [165509100012](https://pubmed.ncbi.nlm.nih.gov/165509100012/).
18. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography*. 2004; 60(Pt 12 Pt 1):2126–32. Epub 2004/12/02. doi: [10.1107/S0907444904019158](https://doi.org/10.1107/S0907444904019158) PMID: [15572765](https://pubmed.ncbi.nlm.nih.gov/15572765/).
19. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D*. 1999; 55:247–55. doi: [10.1107/S090744499801405x](https://doi.org/10.1107/S090744499801405x) PMID: [000078314000031](https://pubmed.ncbi.nlm.nih.gov/000078314000031/).
20. Kraulis PJ. Molscript—a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Crystallogr*. 1991; 24:946–50. doi: [10.1107/S0021889891004399](https://doi.org/10.1107/S0021889891004399) PMID: [1991GL62900080](https://pubmed.ncbi.nlm.nih.gov/1991GL62900080/).
21. Esnouf RM. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr D*. 1999; 55:938–40. doi: [10.1107/S0907444998017363](https://doi.org/10.1107/S0907444998017363) PMID: [000079855200042](https://pubmed.ncbi.nlm.nih.gov/000079855200042/).
22. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. *Methods in enzymology*. 1997; 277:505–24. Epub 1997/01/01. PMID: [18488322](https://pubmed.ncbi.nlm.nih.gov/18488322/).
23. Xu GQ, Jiang XG, Jaffrey SR. A Mental Retardation-linked Nonsense Mutation in Cereblon Is Rescued by Proteasome Inhibition. *J Biol Chem*. 2013; 288(41):29573–85. doi: [10.1074/jbc.M113.472092](https://doi.org/10.1074/jbc.M113.472092) PMID: [000330615300033](https://pubmed.ncbi.nlm.nih.gov/000330615300033/).

A FRET-Based Assay for the Identification and Characterization of Cereblon Ligands

Iuliia Boichenko, Silvia Deiss, Kerstin Bär, Marcus D. Hartmann,* and Birte Hernandez Alvarez*

Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

ABSTRACT: Cereblon serves as an ubiquitin ligase substrate receptor that can be tuned toward different target proteins by various cereblon-binding agents. This offers one of the most promising avenues for targeted protein degradation in cancer therapy, but cereblon binding can also mediate teratogenic effects. We present an effective assay that is suited for high-throughput screening of compound libraries for off-target cereblon interactions but also can guide lead optimization and rational design of novel cereblon effector molecules.



INTRODUCTION

The protein cereblon is a substrate receptor of the CRL4A E3 ubiquitin ligase complex and the primary target of thalidomide,¹ a drug that gained notoriety as a teratogen, inducing severe malformations in newborns and thus causing the greatest pharmaceutical catastrophe of the last century. Nevertheless, thalidomide and its derivatives pomalidomide and lenalidomide belong to a group of therapeutically important immunomodulatory drugs (IMiDs), used in the treatment of multiple myeloma and other B cell malignancies because of their antiangiogenic and antineoplastic properties. As established recently, they mediate both their therapeutic and teratogenic effects via their interaction with cereblon. The first identified substrates of the cereblon–CRL4A E3 ligase complex were the B cell specific transcription factors IKZF1 and IKZF3. Both proteins are ubiquitinated in the presence of IMiDs, thus affecting downstream targets like interleukin-2 (IL-2) or interferon regulatory factor 4 (IRF4).^{2–4} The homeobox transcription factor MEIS2 was later identified as a first endogenous target for which IMiD binding had the opposite effect and prevented its degradation.⁵ More recently, casein kinase 1A1 (CK1 α) has been identified as another target of cereblon,^{6,7} which was surprisingly only recognized and ubiquitinated in the presence of lenalidomide, while thalidomide and pomalidomide had no effect.⁶ This implies that the affinity of cereblon can be modulated and directed toward different protein substrates by the binding of different effector molecules.^{8,9} Consequently, pharmaceutical interest in cereblon is dipartite. For the majority of pharmaceuticals, to avoid teratogenic side effects, cereblon poses an off-target to be circumvented in early drug development. Conversely, the specific development of effector molecules mediating targeted ubiquitination of disease related proteins via cereblon is an important emerging field in pharmacology.

Cereblon is composed of an N-terminal LON-like domain, which attaches to the E3 ubiquitin ligase complex, and a C-terminal domain, which binds thalidomide. Homologues of the thalidomide-binding domain, also referred to as CULT, are found as single-domain proteins in bacteria and, preceded by an N-terminal secretion signal, in lower animals up to fish.¹⁰ Recent studies of the domain from human, chicken, mouse, and *Magnetospirillum* revealed an unusually high degree of conservation between the species.^{5,11,12} The glutarimide ring of the IMiDs slots into an aromatic pocket of this domain, built from three invariant tryptophan residues, and forms hydrogen bonds with the protein backbone of the pocket. Uracil, the nucleobase moiety of uridine and the only natural ligand identified to date, employs the same binding mode. It however remains unclear, what cereblon recognizes in the cell in absence of IMiDs. With the exception of MEIS2, the known substrates IKZF1, IKZF3, and CK1 α are ubiquitinated in an IMiD-dependent manner, suggesting that the IMiDs might emulate a posttranslational modification targeting these proteins for degradation. Because of the aromatic nature of the thalidomide-binding pocket, a modified cationic protein residue seems a possible candidate.¹² However, no such ligand or protein modification has been identified yet.

In this work, we present a robust assay for the identification and characterization of cereblon effector molecules, natural or pharmaceutical. Taking advantage of the invariant tryptophan residues of the binding pocket, we designed a custom reporter ligand with a fluorophore that has an overlapping absorption band with the emission band of tryptophan. Monitoring FRET signaling, we follow the replacement of the reporter ligand by potential test compounds and thereby determine IC₅₀ values and specific affinity constants. We demonstrate the potential of

Received: November 6, 2015

Published: January 5, 2016

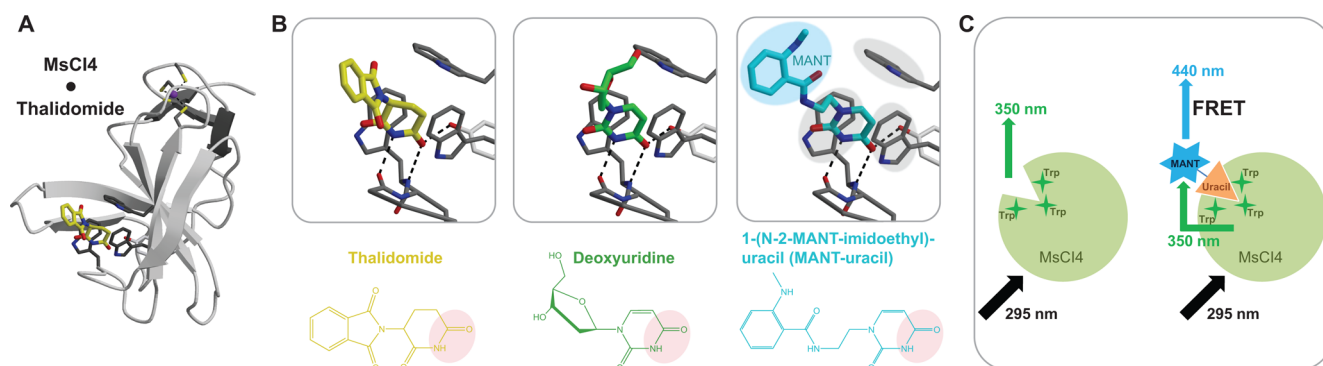


Figure 1. The ligand-binding mode of cereblon and its exploitation for FRET. (A) Structure of MsCI4 with thalidomide bound to the pocket formed by three invariant tryptophan residues. (B) The first two panels show details of thalidomide and deoxyuridine bound to MsCI4. In the right panel, the fluorogenic compound 1-(N-2-MANT-imidoethyl)-uracil (MANT-uracil) is modeled into the pocket. The average distance between the tryptophan residues (gray background) and the MANT-fluorophore (blue background) is approximately 0.7 nm in this model. (C) Principle of the FRET assay. Upon excitation at 295 nm, the tryptophan residues emit at 350 nm. When MANT-uracil is bound, MANT is excited by this fluorescence due to its close proximity, resulting in a FRET signal at 440 nm.

the assay in a comparative characterization of thalidomide-binding domains from human, *Magnetospirillum gryphiswaldense*, and *Caenorhabditis elegans*.

RESULTS AND DISCUSSION

We have previously established *Magnetospirillum gryphiswaldense* cereblon isoform 4 (MsCI4) as a robust and tractable model protein for the thalidomide-binding domain.¹² For tryptophan fluorescence experiments, we devised the construct MsCI4^{WW/FF}, which has two surface tryptophan residues outside the binding pocket replaced by phenylalanine to minimize intrinsic fluorescence. In initial measurements, exciting at 295 nm and monitoring at 350 nm, we observed a decrease in tryptophan fluorescence with increasing ligand concentrations,¹³ which could be used as a trivial binding assay. However, these measurements were hampered by differences in the quenching properties of the ligands, unspecific binding, and unspecific quenching. To minimize such effects and to increase overall sensitivity, we designed the reporter ligand 1-(N-2-MANT-imidoethyl)-uracil (MANT-uracil) that binds specifically to the thalidomide-binding pocket (Figure 1). The N-methylanthraniloyl (MANT) group constitutes an optimal acceptor fluorophore for a FRET-based competition assay: its excitation maximum at 330–350 nm overlaps with the emission maximum of tryptophan fluorescence and has its own emission maximum at 430–450 nm. Uracil serves as the binding moiety; a model in Figure 1 shows how binding of MANT-uracil positions the MANT fluorophore in close proximity to the tryptophan residues for efficient FRET. The expected FRET was verified measuring single fluorescence emission spectra of MsCI4^{WW/FF} alone and in the presence of MANT-uracil (Figure 2A). Whereas MANT-uracil alone is only slightly excited at 295 nm, a strong emission signal at 440 nm arises in the presence of equal amounts of protein; this signal increases upon titration of the ligand to the protein in a concentration-dependent manner (Figure 2B). On the basis of this observation, we determined the affinity of MsCI4^{WW/FF} to MANT-uracil, monitoring the signal at 440 nm, and scaled the experimental setup to a 96-well format. To ensure a fixed concentration of the cosolvent, a master dilution series of the MANT-uracil stock was prepared in DMSO.

Titration of MANT-uracil to the protein yielded a curve indicative of binding. MANT-uracil as a control shows a linear

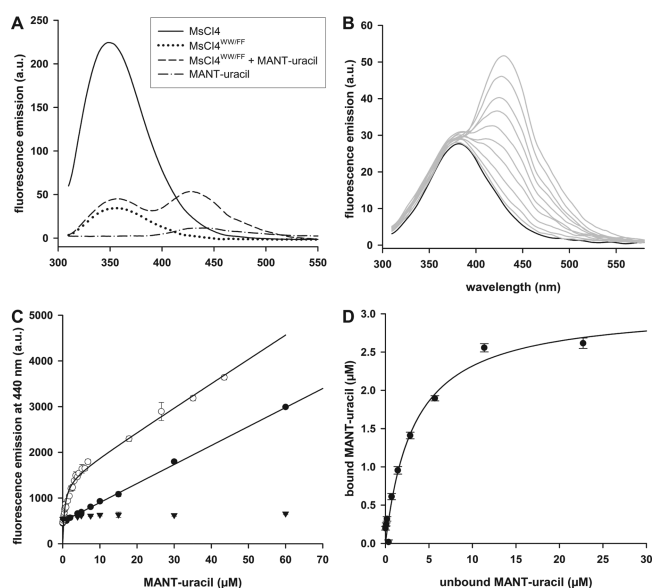


Figure 2. Specific binding of MANT-uracil to MsCI4. (A) Fluorescence spectra of 10 μM MsCI4, MsCI4^{WW/FF}, and MsCI4^{WW/FF} in the presence of 10 μM MANT-uracil and MANT-uracil diluted in buffer as a control. The emission maxima of MsCI4^{WW/FF} and wild-type MsCI4 differ in intensity due to the lacking surface tryptophan residues. (B) Fluorescence spectra of MsCI4^{WW/FF} (10 μM) in absence (black) and upon titration of 0.2–44 μM MANT-uracil. (C) FRET signal for MANT-uracil titrated to 3 μM MsCI4^{WW/FF} (○), MANT-uracil in assay buffer (●), and 3 μM MsCI4^{WW/FF} in assay buffer plus DMSO (corresponding to MANT-uracil diluted from DMSO stock) (▼). Data points represent the mean of three measurements including standard deviation. (D) Data analysis of steady-state FRET binding assay. Fractions of MsCI4^{WW/FF}::MANT-uracil and of free ligand were calculated from FRET data and fitted according to the one-binding site model, yielding $K_D = 3.3 \pm 0.3 \mu\text{M}$. All experiments were excited at 295 nm.

increase of the emission signal resulting from the autofluorescence of the fluorophore (Figure 2C). A blank-corrected titration series resulted in the binding curve in Figure 2D. Given the known stoichiometry of 1:1 and applying a hyperbolic fit according to the one binding site model, we obtained an affinity constant of $K_D = 3.3 \pm 0.3 \mu\text{M}$ for MANT-uracil.

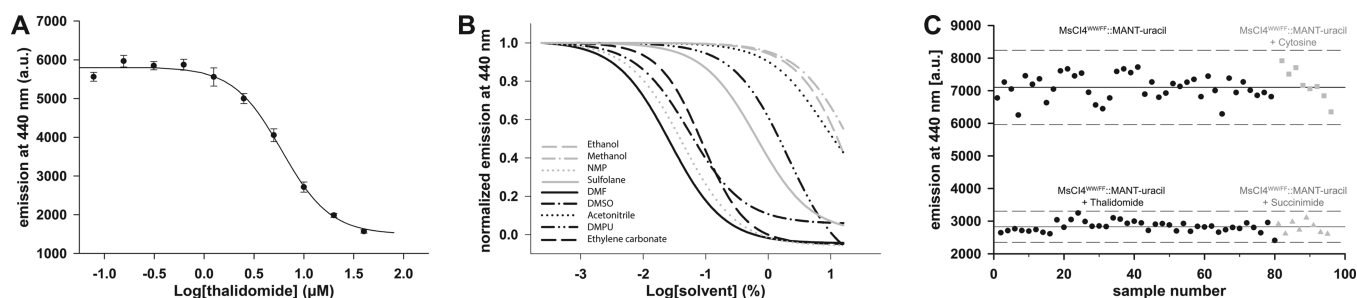


Figure 3. Competitive binding assay and solvent effects. (A) Titration of thalidomide (0–80 μM) yields an IC_{50} of $7.8 \pm 0.3 \mu\text{M}$, corresponding to $K_{i(\text{Thal})} = 4.4 \mu\text{M}$. (B) Influence of different polar aprotic solvents (0–20% (v/v)) on the assay. (C) Evaluation of assay quality for calculation of Z' -factor. Scatter plot of FRET values obtained from a 96-well plate containing 40 samples of each, positive control ($\text{MsCI4}^{\text{WW/FF}}::\text{MANT-uracil}$) and negative control ($\text{MsCI4}^{\text{WW/FF}}::\text{MANT-uracil}$, 30 μM thalidomide). Succinimide (30 μM) competing for MANT-uracil binding reduces FRET, whereas cytosine (30 μM) does not. Means of positive and negative controls are shown by solid lines; dashed lines indicate 3-fold standard deviations. All experiments contained 2.5 μM $\text{MsCI4}^{\text{WW/FF}}::\text{MANT-uracil}$ and were excited at 295 nm. A and B show means of three measurements. DMF, *N,N*-dimethylformamide; DMSO, dimethyl sulfoxide; DMPU, *N,N'*-dimethylpropylene urea; NMP, 1-methyl-2-pyrrolidinone.

Table 1. Binding Constants of Cereblon Ligand-Binding Domains from Human, *C. elegans*, and *Magnetospirillum*^a

protein	MANT-uracil		\pm -thalidomide		succinimide		glutarimide		DMSO	
	K_D [μM]	IC_{50} [μM]	K_i [μM]	IC_{50} [μM]	K_i [μM]	IC_{50} [μM]	K_i [μM]	IC_{50} [μM]	K_i [μM]	
$\text{MsCI4}^{\text{WW/FF}}$	3.3 ± 0.3	7.8 ± 0.3	4.4	7.6 ± 1.0	4.3	48.9 ± 8.9	27.8	10000 (0.1%)	5690 (0.06%)	
<i>hCRBN</i> Δ 1-315	7.9 ± 0.7	29.9 ± 7.8	22.7	34.5 ± 4.8	26.2	135.8 ± 19	103.1	181000 (1.3%)	137490 (1%)	
<i>CeCRBNsec</i> Δ 1-15	16.0 ± 0.4	18.6 ± 1.9	16.1	6.7 ± 1.0	5.8	50.6 ± 6.2	43.8	361000 (2.6%)	312200 (2.2%)	
$\text{MsCI4}^{\text{WW/FF, Y101F}}$	22.9 ± 1.1	10.7 ± 0.8	9.6	53.0 ± 6.6	47.8	275.3 ± 24	248.2	737000 (5.3%)	664460 (4.8%)	

^aData are means of three replicates with standard errors. K_D values for MANT-uracil were calculated from FRET data according to the one binding site model. IC_{50} curves were obtained from titrations of 0–80 μM thalidomide, 0–5 mM succinimide, 0–20 mM glutarimide, and 0–20% DMSO. K_i values were calculated from IC_{50} values using Cheng–Prusoff equation.

In a next step, we performed a first competitive binding assay with thalidomide as a test compound. Again, a master dilution series of the ligand was prepared in DMSO. As anticipated, titration of thalidomide to the $\text{MsCI4}^{\text{WW/FF}}::\text{MANT-uracil}$ complex in assay buffer decreased the FRET signal in a concentration-dependent manner. Plotting of the data using a log (dose) response curve yielded an IC_{50} value of $7.8 \pm 0.3 \mu\text{M}$ for thalidomide (Figure 3A). Applying the Cheng–Prusoff equation¹⁴ yielded an absolute inhibition constant of $K_{i(\text{Thal})} = 4.4 \mu\text{M}$.

As the stock solution and master dilution series of MANT-uracil and thalidomide are prepared in 100% DMSO, we examined for a possible inner filter effect.¹⁵ We found that DMSO concentrations of up to 5% do not have any effect on the fluorescence measurements. Further testing for deleterious effects of the solvent on the measurements, we performed a competition assay using merely DMSO as the test compound, within a range of 0–20%. The according log (dose) response curve implies that DMSO competes with MANT-uracil in the low-affinity range, yielding an $\text{IC}_{50(\text{DMSO})} = 0.1\%$ and an according inhibition constant of $K_{i(\text{DMSO})} = 0.06\%$, which corresponds to 10 and 5.7 mM DMSO, respectively. The binding curve in Figure 3B shows that DMSO concentrations higher than about 0.03% start to affect the measurement significantly. This result is not surprising, as it supports observations in a crystal structure of MsCI4 containing a DMSO molecule bound in the pocket.¹³ In the experiments measuring binding of MANT-uracil and thalidomide, the final DMSO concentrations were below the threshold of 0.03% (0.01% and 0.02%, respectively). However, the low-affinity binding of DMSO generally limits its use as a solvent for water-insoluble compounds if the experimental setup requires concentrations higher than 0.03%. We therefore tested several

alternate aprotic polar solvents for their suitability in the assay (Figure 3B). Many of them show low-affinity binding or otherwise impair the protein in a range similar to DMSO. However, acetonitrile, methanol, and ethanol represent reasonable alternatives to DMSO; they appear applicable up to a concentration of about 1%.

The stability of $\text{MsCI4}^{\text{WW/FF}}$ and MANT-uracil in the assay was evaluated by monitoring fluorescence at 440 nm over a period of 48 h, which did not show any significant loss in signal intensity. As a statistical indicator of assay quality, we further estimated the Z' -factor and obtained value of $Z' = 0.6$, indicating very good quality (Figure 3C).¹⁶ Finally, we performed a comparative analysis of thalidomide-binding domains. We chose representatives from all three subgroups of cereblon proteins:¹⁰ human *hCRBN* Δ 1-315 from an E3 ligase substrate receptor, the secreted *C. elegans* variant *CeCRBNsec* Δ 1-15, and the bacterial $\text{MsCI4}^{\text{WW/FF}}$. As MsCI4 differs from the two eukaryotic proteins by an F \rightarrow Y substitution in the binding pocket, we also assayed the canonical variant $\text{MsCI4}^{\text{WW/FF, Y101F}}$.¹² All proteins were assayed with the same set of ligands: MANT-uracil itself as a uracil derivative, thalidomide and its binding moiety glutarimide, and additionally succinimide as a potential new ligand. We suspected succinimide to bind because of its structural similarity to glutarimide as the two compounds merely differ in ring size, 5 for succinimide and 6 for glutarimide. All compounds were bound by all proteins, with glutarimide generally showing the lowest affinity (Table 1). $\text{MsCI4}^{\text{WW/FF}}$ has the highest binding affinities for the different ligands, with K_i values in the very low micromolar range. The affinities of *CeCRBNsec* Δ 1-15 and *hCRBN* Δ 1-315 to the different compounds were significantly lower, which we accredited to the F \rightarrow Y substitution in MsCI4 : in comparison to *hCRBN* and

CeCRBN, MsCI4 forms an additional hydrogen bond with the ligands via the Y101 hydroxy group in the binding pocket. Structurally, substitution of this tyrosine to phenylalanine as in the eukaryotic proteins does not affect the geometry of ligand binding but it decreases affinity, as we now verified in the comparison of MsCI4^{WW/FF, Y101F} and MsCI4^{WW/FF}. Consistently, the binding affinities for MsCI4^{WW/FF, Y101F} are lower, in a range comparable to hCRBN Δ 1-315 and CeCRBNsec Δ 1-15.

In two cases, affinity constants of hCRBN Δ 1-315 for IMiDs have been reported. For thalidomide, an IC₅₀ value of \approx 30 μ M was determined in dose–response pulldown experiments, while fluorescence polarization binding assays yielded a K_D of \approx 250 nM.^{5,17} Although these experiments are methodologically hard to compare, the IC₅₀ values determined in the pulldown are roughly comparable with our results. The K_D values determined by fluorescence polarization are noticeably low but also not directly comparable as the experiments were performed with the full hCRBN protein in complex with DDB1, which might further stabilize the ligand-bound conformation.

CONCLUSION

Here, we demonstrated the high potential and general applicability of our newly developed FRET-based assay. In a comparative substrate binding study, representatives of all cereblon proteins showed the same differential affinities to the same substrates on a relative scale. Therefore, affinities determined for one cereblon homologue can be translated to expectation values for other species. As the bacterial representative MsCI4 was the most stable, accessible, and easiest-to-handle protein in our study, it is so far the most advisable candidate for the general characterization of cereblon ligand binding. With its amenability for high-throughput screening, the assay may serve an important role in screening libraries of pharmaceuticals for potential cereblon-mediated teratogenic effects. The identification of succinimide as a new cereblon-binding moiety represents a first success of this application. Furthermore, with its high accuracy, the assay is suited for assisting the rational design of cereblon effectors for the targeted degradation of disease-related proteins as in cancer therapy.

EXPERIMENTAL SECTION

Molecular Biology. A DNA fragments encoding the *C. elegans* cereblon homologue (gil17569329) lacking the N-terminal secretion signal and the thalidomide-binding domain of hCRBN (hCRBN Δ 1-315) were codon optimized for expression in insect cells and synthesized (Genscript). CeCRBNsec Δ 1-15 was cloned with a C-terminal histidine tag and hCRBN Δ 1-315 with an N-terminal His-ZZ-tag into pFastBac/CT-TOPO vector (Life Technology). Recombinant baculoviruses were generated and amplified using the Bac-to-Bac Baculovirus expression system (Life Technology).

Protein Purification. MsCI4^{WW/FF} and MsCI4^{WW/FF, Y101F} were expressed and purified as described previously.¹² CeCRBNsec Δ 1-15 and hCRBN Δ 1-315 were expressed in Sf9 insect cells. Then 72 h post infection, cells were harvested by centrifugation. The cell pellet was resuspended in buffer A (50 mM NaH₂PO₄, pH 8.0, 500 mM NaCl, 5 mM β -mercaptoethanol, 4 mM MgCl₂) containing 1 mM PMSF, protease inhibitor mix, and DNase I. After sonication, cell debris was pelleted by centrifugation and the supernatant applied onto a NiNTA column equilibrated with buffer A. Bound proteins were eluted with a linear gradient of 0–0.5 M imidazole in buffer A. Protein-containing fractions were pooled and dialyzed against 20 mM Tris, pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol containing additionally 0.1 M L-arginine for CeCRBNsec Δ 1-15. Tags were cleaved using TEV protease. Applying a second NiNTA column, pure proteins were

finally separated from TEV protease and cleavage products. Protein containing fractions were pooled and samples were concentrated in dialysis buffer.

Inner Filter Effect. Assay buffer (20 mM Tris, pH 7.5, 150 mM NaCl, and 0.5 mM β -mercaptoethanol) containing varying DMSO concentrations (0–20%) and 2.5 μ M MsCI4^{WW/FF}::MANT-uracil was recorded on a Synergy microplate reader in black 96-well plates. Parameter settings were as described for the multiwell plate assay, excepting excitation: (350/9 nm).

Fluorescence Spectroscopy. Fluorescence spectra were measured using a 2 mL cuvette in a JASCO FP-6500 spectrofluorometer at a protein concentration of 20 μ M in assay buffer. Samples were excited at 295 nm, and emission spectra were recorded at a data pitch of 0.5, excitation bandwidth of 1 nm, emission bandwidth of 3 nm, response of 0.5 s, and scan speed of 200 nm/min. JASCO software was used to analyze data.

Multiwell Plate Assay. Fluorescence end point measurements were performed in black 96-well plates on Synergy microplate readers (BioTek) with the following settings: excitation, 295/9 nm; emission collection, 440/9 nm; optics, top; read speed, normal; delay, 100 ms, measurements/data point, 10; read height, 8 mm. All measurements were done in triplicate; SigmaPlot 12.3 (Systat Software) was used for data evaluation, curve fitting, plotting, and determination of specific constants.

MANT-uracil was obtained by custom synthesis (Jena Bioscience). All measurements were performed in assay buffer. Stock solutions for MANT-uracil (1 M) and for thalidomide (400 mM) were prepared in DMSO, all other ligands in assay buffer. Highly concentrated master dilution series of MANT-uracil and thalidomide stocks were prepared in DMSO and subsequently diluted in assay buffer to keep the final DMSO concentration \leq 0.02%. Binding of MANT-uracil was measured by incubating protein (2.5 μ M) with increasing concentrations of this ligand in assay buffer. Exciting at 295 nm, fluorescence was monitored at 440 nm. The autofluorescence of MANT-uracil at the according concentrations served as a blank. The obtained FRET data resembling the pool of MANT-uracil bound to protein were used to calculate the B_{max} value, which corresponds to the saturation of the binding pocket by bound ligand. Given the known binding stoichiometry, for each data point the relative amount of bound and free MANT-uracil was calculated. The plots were rescaled and fitted according to the one binding site model to calculate the specific affinity constant K_p.

For determination of IC₅₀ values, ligand was titrated to a final volume of 120 μ L assay buffer containing 2.5 μ M each, MANT-uracil and protein. Binding competition was monitored by collecting emission data sets at 440 nm. Samples containing compound in buffer served as references. IC₅₀ values were determined by applying a log (dose) response curve on averaged data obtained from three to five independent measurements. K_i values were calculated according to Cheng–Prusoff equation.¹⁴

Z'-Factor. The Z'-factor was estimated according to Zhang et al.¹⁶ In a black 96-well plate, 40 samples of each, the positive control (2.5 μ M MsCI4^{WW/FF}::MANT-uracil, 0.02% DMSO in assay buffer) and the negative control (2.5 μ M MsCI4^{WW/FF}, 2.5 μ M MANT-uracil, and 30 μ M thalidomide in assay buffer) were analyzed performing fluorescence end point measurements according to the assay protocol.

AUTHOR INFORMATION

Corresponding Authors

*For M.D.H.: phone, -49 7071 601323; E-mail, marcus.hartmann@tuebingen.mpg.de.

*For B.H.A.: phone, -49 7071 601356; E-mail, birte.hernandez@tuebingen.mpg.de.

Author Contributions

I.B. and S.D. conducted the cloning and protein purification experiments; K.B. performed assay measurements; B.H.A. established the FRET assay and supervised all biochemical work; M.D.H. and B.H.A. designed the project and wrote the

manuscript; all authors discussed the results and critically read the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Andrei Lupas for continuing support.

ABBREVIATIONS USED

IMiD, immunomodulatory drug; MANT, *N*-methyl-anthraniloyl; MANT-uracil, 1-(*N*-2-MANT-imidoethyl)-uracil; NiNTA, nickel-nitrilotriacetic acid

REFERENCES

- (1) Ito, T.; Ando, H.; Suzuki, T.; Ogura, T.; Hotta, K.; Imamura, Y.; Yamaguchi, Y.; Handa, H. Identification of a primary target of thalidomide teratogenicity. *Science* **2010**, *327*, 1345–50.
- (2) 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. 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.* **2014**, *164*, 811–21.
- (3) 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, S. L.; Carr, S. A.; Ebert, B. L. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **2014**, *343*, 301–5.
- (4) Lu, G.; Middleton, R. E.; Sun, H.; Naniong, M.; Ott, C. J.; Mitsiades, C. S.; Wong, K. K.; Bradner, J. E.; Kaelin, W. G., Jr. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **2014**, *343*, 305–9.
- (5) Fischer, E. S.; Bohm, 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.; Thoma, N. H. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **2014**, *512*, 49–53.
- (6) Hagner, P. R.; Man, H. W.; Fontanillo, C.; Wang, M.; Couto, S.; Breider, M.; Bjorklund, C.; Havens, C. G.; Lu, G.; Rychak, E.; Raymon, H.; Narla, R. K.; Barnes, L.; Khambatta, G.; Chiu, H.; Kosek, J.; Kang, J.; Amantangelo, M. D.; Waldman, M.; Lopez-Girona, A.; Cai, T.; Pourdehnad, M.; Trotter, M.; Daniel, T. O.; Schafer, P. H.; Klippel, A.; Thakurta, A.; Chopra, R.; Gandhi, A. K. CC-122, a pleiotropic pathway modifier, mimics an interferon response and has antitumor activity in DLBCL. *Blood* **2015**, *126*, 779–789.
- (7) Kronke, J.; Fink, E. C.; Hollenbach, P. W.; MacBeth, K. J.; Hurst, S. N.; Udeshi, N. D.; Chamberlain, P. P.; Mani, D. R.; Man, H. W.; Gandhi, A. K.; Svinkina, T.; Schneider, R. K.; McConkey, M.; Jaras, M.; Griffiths, E.; Wetzler, M.; Bullinger, L.; Cathers, B. E.; Carr, S. A.; Chopra, R.; Ebert, B. L. Lenalidomide induces ubiquitination and degradation of CK1 α in del(5q) MDS. *Nature* **2015**, *523*, 183–8.
- (8) Lu, J.; Qian, Y.; Altieri, M.; Dong, H.; Wang, J.; Raina, K.; Hines, J.; Winkler, J. D.; Crew, A. P.; Coleman, K.; Crews, C. M. Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem. Biol.* **2015**, *22*, 755–63.
- (9) Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E. DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **2015**, *348*, 1376–81.
- (10) Lupas, A. N.; Zhu, H.; Korycinski, M. The thalidomide-binding domain of cereblon defines the CULT domain family and is a new member of the beta-tent fold. *PLoS Comput. Biol.* **2015**, *11*, e1004023.
- (11) 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. Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat. Struct. Mol. Biol.* **2014**, *21*, 803–9.
- (12) Hartmann, M. D.; Boichenko, I.; Coles, M.; Zanini, F.; Lupas, A. N.; Hernandez Alvarez, B. Thalidomide mimics uridine binding to an aromatic cage in cereblon. *J. Struct. Biol.* **2014**, *188*, 225–32.
- (13) Hartmann, M. D.; Boichenko, I.; Coles, M.; Lupas, A. N.; Hernandez Alvarez, B. Structural dynamics of the cereblon ligand binding domain. *PLoS One* **2015**, *10*, e0128342.
- (14) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (K₁) and the concentration of inhibitor which causes 50% inhibition (I₅₀) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- (15) Liu, Y.; Kati, W.; Chen, C. M.; Tripathi, R.; Molla, A.; Kohlbrenner, W. Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **1999**, *267*, 331–5.
- (16) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screening* **1999**, *4*, 67–73.
- (17) 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. Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia* **2012**, *26*, 2326–35.