Received: 26 August 2021 | Revised: 2 November 2021 | Accepted: 22 November 2021

DOI: 10.1002/prot.26286

RESEARCH ARTICLE

PROTEINS WILEY

Substrate-assisted activation and selectivity of the bacterial RavD effector deubiquitinylase

Eric Schulze-Niemand[1,2] | Michael Naumann[1] | Matthias Stein[2]

[1]Institute for Experimental Internal Medicine, Medical Faculty, Otto-von-Guericke University, Magdeburg, Germany

[2]Molecular Simulations and Design Group, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Correspondence
Matthias Stein, Molecular Simulations and Design Group, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany.
Email: matthias.stein@mpi-magdeburg.mpg.de

Funding information
European Regional Development Fund of the Ministry of Economy, Science and Digitalization in Saxony-Anhalt within the Center of Dynamic Systems, Grant/Award Number: ZS/2016/04/78155; EU COST Action (European Cooperation in Science and Technology) CA20133 'ProteoCure'; Max Planck Society for the Advancement of Science

Abstract

Deubiquitinylases (DUBs) catalyze the peptide bond cleavage of specific ubiquitin linkages at distinct protein substrates. Pathogens from viruses and bacteria independently developed effector proteins with DUB activity to mimic host DUB functions and circumvent immune responses. The effector protein RavD from *Legionella pneumophila* cleaves linear ubiquitin chains with an exclusive methionine-1 selectivity. It thus performs as a functional analogue of the human DUB OTULIN, which achieves its selectivity only via a specialized proximal ubiquitin S1′ binding site as well as a substrate-assisted activation of the catalytic triad. An analysis of the crystal structures of bacterial RavD in its free and di-ubiquitin-bound forms, in order to rationalize the structural basis for its selectivity and activation mechanism, is not fully conclusive. As these ambiguities might arise from the introduced double mutation of the di-ubiquitin substrate in the RavD–di-ubiquitin complex crystal structure, biomolecular modeling, and molecular dynamics sampling (1–2 μs for each system of RavD and OTULIN) were employed to reconstitute the physiological RavD–di-ubiquitin complex. The simulations show that the distal S1 ubiquitin binding sites of RavD and OTULIN are similar in terms of interface area, composition, and ubiquitin binding affinity. The proximal S1′ site of RavD, in contrast, is significantly smaller and ubiquitin binding is weaker and more flexible than in OTULIN. Upon substrate access, the residues of the catalytic triad of RavD show a reduction of flexibility and a conformational transition toward a catalytically active state. Thus, the enzymatic activation of RavD is presumably also substrate-assisted and a clear rationale for the common M1-substrate selectivity.

KEYWORDS
substrate-assisted catalysis, deubiquitinylase, *Legionella pneumophila*, molecular dynamics, OTU, protein–protein interactions, ubiquitin recognition

1 | INTRODUCTION

Ubiquitinylation is among the most abundant prost-translational modifications (PTM) and has key regulatory functions in most cellular processes.[1] This reversible PTM is widely recognized to target proteins for proteasomal degradation or regulation[2] and has emerged as a critical signal in innate immune response,[3,4] in which it initiates inflammation, impedes pathogen growth, and triggers cell death.[5]

Ubiquitinylation is achieved through the concerted action of specific E1 activating enzymes, E2 conjugating enzymes, and E3 ligases.[6] Multiple ubiquitin (Ub) units can be conjugated via (iso)-peptide linkages of one of the exposed lysine residues (K6, K11, K27, K29, K33, K48, and K63) or the N-terminal methionine (M1) of the proximal Ub with the C-terminus of the distal Ub.[7] This chemical bonding determines the fate of the ubiquitinylated protein, for example, degradation, trafficking, or signaling.[8]

The reverse process, that is, the cleavage of ubiquitin chains, is performed by deubiquitinylase enzymes (DUBs). There are ~100 human DUBs, which are categorized into seven different families based on structural and functional characteristics.[9] While most DUBs are cysteine proteases, the JAMM subfamily employs a Zn^{2+} ion in its catalytic center.[10] DUBs possess multiple factors of selectivity control regarding protein recognition and Ub-linkage type.[9,11]

Some members of the family of OTU DUBs are known for their high selectivity toward various types of ubiquitin linkages.[12] OTUD7B is highly selective toward K11 linkage,[13] OTUB1,[14] and OTUD1[12] exclusively cleave K48-linkages and K63-linkages, respectively, and OTULIN[15,16] exhibits a unique activity against linear methionine 1 (M1)-linked polyubiquitin chains. Methionine 1-linked ubiquitin chains are critical regulators of inflammation and immunity to pathogens.[17] Such a unique M1-selectivity requires a sophisticated multifactorial recognition mechanism involving multisite recognition and substrate-assisted catalysis.[16] Such a high-level control of selectivity can only be addressed by structural investigations.[9,12]

Viral and bacterial pathogens have independently evolved numerous deubiquitinylating effector proteins to mimic host DUBs as a strategy to counteract innate immune response.[18–21] Recently, the *L. pneumophila* effector protein RavD[22] was identified to also cleave linear M1-polyubiquitin chains and thus inhibit downstream M1-ubiquitinylation-dependent NF-κB signaling.

Protein crystal structures of RavD and OTULIN in absence and when in complex with substrate M1-di-ubiquitin (DiUb) are available (see Figure 1) but not conclusive in terms of structural control of selectivity and its link to the activation mechanism.[16,22] For OTULIN, a substrate-assisted catalysis activation mechanism to bring the active site triad into a catalytically competent state was confirmed.[16] RavD, however, did not show the change in inter-residue distances as the substrate approaches the active state (see below).[22] In addition to the active site binding, the DUB interacts with the ubiquitins by forming protein–protein contact interfaces with the proximal (S1 binding site) and distal (S1′ site) ubiquitin molecules.[22]

Here, we identify critical structural features that are able to rationalize the shared exclusive M1-linkage selectivity of RavD and OTULIN OTU DUBs. Molecular dynamics (MD) simulations on the microsecond-scale for the apo- and substrate-bound forms of RavD and OTULIN are employed to probe the persistence of the assembly of the active site residues and the ubiquitin–DUB interactions at the binding sites S1 and S1′. The protein–protein interface areas and contact residues of the distal (S1) and proximal (S1′) ubiquitin binding sites of RavD are compared with OTULIN. Structural differences between the active sites residues in the free and substrate-bound states are used to shed light on the discussion of DUB activation mechanism.[23]

2 | RESULTS AND DISCUSSION

2.1 | Structural differences between free and substrate-bound RavD and OTULIN

RavD is a papain-like deubiquitinylase with a Cys–His–Ser catalytic triad, exhibits an overall structure that is dissimilar to OTULIN and may be considered as the founding member of a novel class of DUBs.[24] The binding mode of linear DiUb to RavD is, nevertheless, almost identical to the one in OTULIN (see Figure 1). Different activation mechanisms for RavD and OTULIN, however, were reported.

The prime criteria for a clear identification of a substrate-assisted enzymatic reaction mechanism is the analysis of the structural arrangement within the catalytic center of the unbound enzyme in comparison with the enzyme–substrate complex.[25] For example, human OTULIN's exclusive M1-linkage specificity originates from two Ub-recognition sites S1 and S1′ and, in addition, from substrate-assisted catalysis, in which the catalytic triad is only activated upon tight substrate binding.[16] In particular, when the substrate is in a reactive conformation, the scissile bond comes as close as 4.0 Å to Cys, and the inter-residue distances for the catalytic triad residues decrease from 6.5 to 3.4 Å for Cys–His and from 8.3 to 3.1 Å for His–Asn residues (see Scheme 1). Only this tight complex allows deprotonation of Cys by His and thereupon the activation of the catalytic triad to form the zwitterionic state. The structural differences of substrate co-crystallized RavD and OTULIN were discussed recently.[23]

However, compared with OTULIN–DiUb, in the RavD–DiUb crystal structure, catalytic inter-residue distances are as large as 8.8 Å for the scissile bond to cysteine, and 7.1 Å (for Cys–His) and 2.7 Å (His–Ser). This is beyond a reactive distance for a cysteine protease.[26] The co-crystal of RavD-DiUb was obtained when in complex with a non-hydrolysable DiUb substrate analogue, in which the two terminal glycine residues of the distal Ub are mutated to serine residues (referred to as RavD–DiUbGGSS in the following). These changes may be responsible for a hindered substrate insertion into the catalytic groove, and hence obstruct the necessary conformational change of the catalytic triad.

Molecular dynamics simulations are able to give insight into the dynamics and accessible conformational ensembles of proteins and protein–protein complexes in aqueous solution and at finite temperature.[27] This information is complementary to that of static protein crystal structures in the solid form. We here also recover the physiological RavD–DiUbGG complex, which we refer to as "RavD–DiUb" in the following.

2.2 | Molecular dynamics simulations of RavD–DiUb and OTULIN–DiUb protein–protein complexes

Figure 2A shows the RMSD plot[28] of the M1-linked di-ubiquitin substrate in complex with RavD and OTULIN. For RavD–DiUb, in total

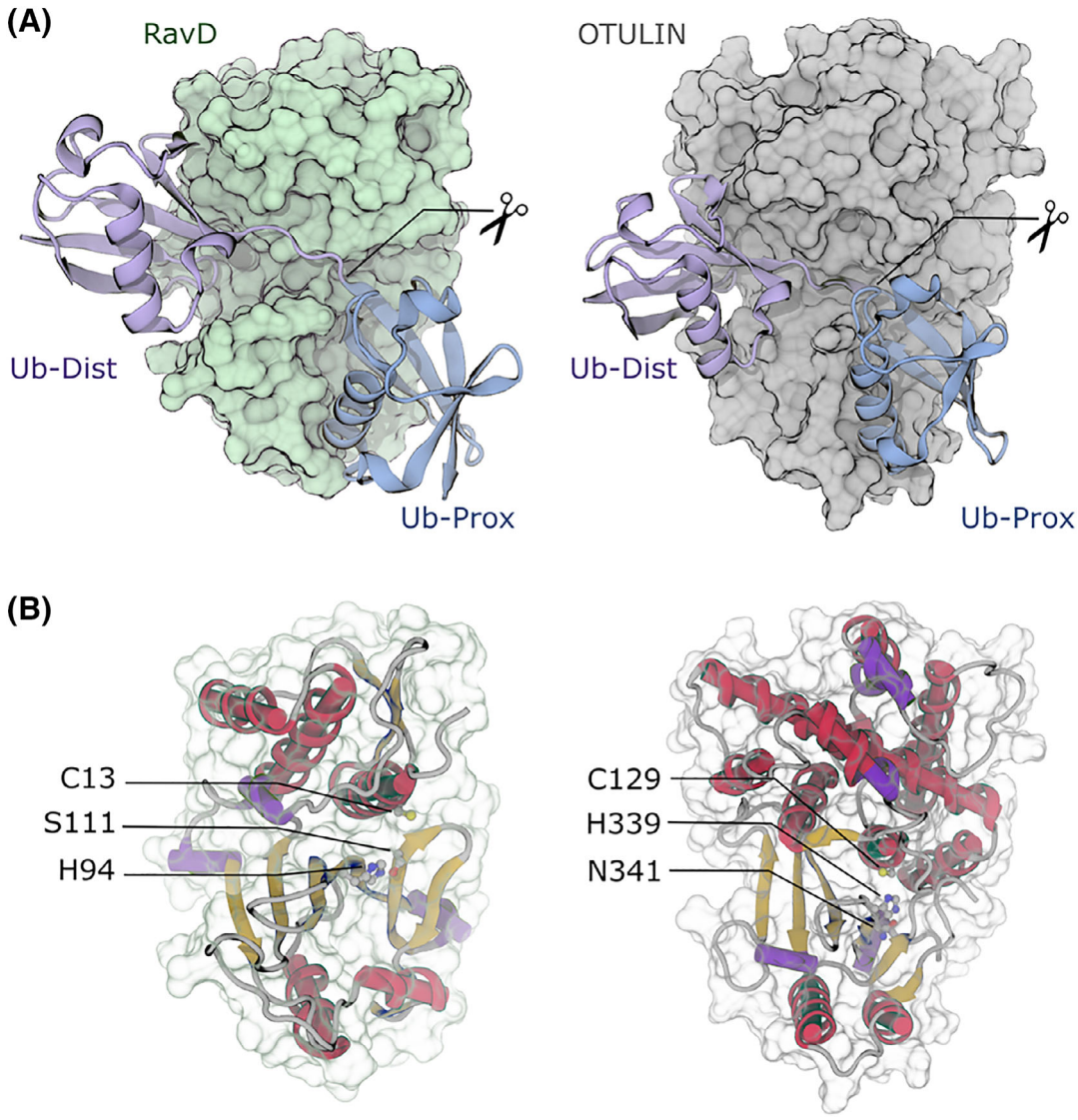


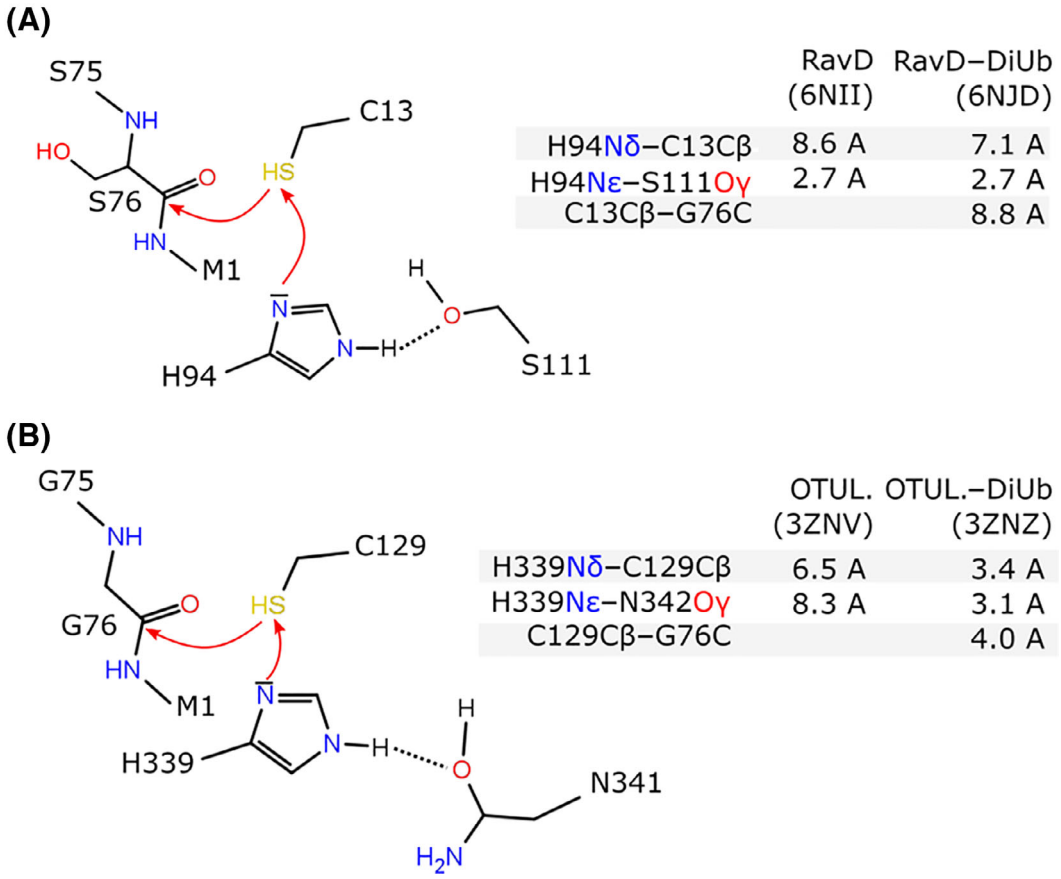
FIGURE 1 Comparison of M1-selective deubiquitinylases from *L. pneumophila* RavD and human OTULIN. (**A**) Surface representation of the DUBs and ribbon display of the di-ubiquitin. The scissile bond of di-ubiquitin is labeled. (**B**) Annotation of the conserved catalytic triad residues

2 μs (4 × 500 ns) production sampling was performed after an extended equilibration period of 500 ns since the crystallized structure was modified to re-constitute the native glycine residues at the C-terminus of the distal ubiquitin. For OTULIN–DiUb simulations, four replicates of 250 ns each with only a short equilibration period were sufficient to give reliable results.

Usually, the root mean square deviation (RMSD) is computed after least-square fitting of the trajectory to a reference structure so that rotational and translational motions are removed and only atomic fluctuations around the reference state are measured. However, in an enzyme–substrate complex, explicit inclusion of rotational and translational motions of the substrate molecule relative to the host molecule is more instructive.

In the case of the RavD–DiUb complex, we find two distinct states with converged RMSDs of the distal Ub of 0.5 nm (in two replicates) and 0.2 nm (in two replicates) with fluctuations of 0.1 nm each (Figure 2A). For a small globular protein such as ubiquitin, a value of 0.4 nm can only occur through a systematic movement from the initial, crystallized binding mode. The fluctuations, on the other hand, rather belong to minor conformational dynamics and orientation tumbling. For comparison, we note that the fluctuations in the RMSD of the distal Ub bound to OTULIN are slightly smaller. The RMSD reaches 0.3 nm, which shows that there are no large structural rearrangements in OTULIN-DiUb. In both cases, a stable binding of the distal Ub to the S1 binding sites can be observed in absence of large structural fluctuations.

Binding of the proximal Ub to the S1′ binding site of RavD, however, is much more flexible and less structurally defined. In two of the four replicates, the RMSD has a base value of 0.3 nm but occasionally reaches 0.6 nm, indicating transitions to short-lived binding modes. In the two other replica, the RMSD shows an average of ~0.5 nm with large fluctuations of 0.4 nm. Thus, the proximal Ub not only diverges



SCHEME 1 Comparison of structural parameters of catalytic triad residues in M1-specific DUBs (**A**) RavD (top) and (**B**) OTULIN (bottom). All distances to cysteine were measured from the Cβ atom for reasons of comparability with the OTULIN–DiUb crystal structure in which cysteine was mutated to alanine

from its initial, as crystallized binding mode, but also occupies a number of transient binding modes. We note that even if the RMSD appears large, these binding modes may be explained by small rotations or translations of the ubiquitin molecules (Figure S1). In OTULIN, the proximal ubiquitin binds in a significantly more stable mode. The RMSD is 0.2 nm and only exhibits marginal fluctuations. In the DUB enzymes RavD and OTULIN (as opposed to the above discussed ubiquitin units), the fluctuations are only minor and the RMSD converges to an average of 0.2 nm. This is indicative of the absence of large conformational changes of RavD and OTULIN during the simulation of the protein–protein complexes (Figure S2).

As a means to resolve the large RMSD of the proximal Ub in the RavD–ubiquitin complex, the pairwise RMSD matrix was calculated as a basis for further conformational clustering. A two-dimensional (2D) embedding of the RMSD distance matrix can, for example, be obtained via the popular UMAP algorithm (Figure 2C).[29] Interestingly, the UMAP embedding puts conformations with a large RMSD from the crystal structure to the left, which allows us to use the u1 coordinate as an approximate binding-unbinding coordinate. We note, however, that such an observation may be coincidal and UMAP embedding axis generally do not correspond to physical coordinates. It has also to be mentioned, that the distance in the UMAP space is not identical to the distance in the RMSD matrix. In order to take further advantage of the RMSD matrix and the UMAP embedding, a *k*-medoids clustering[30] was applied to the RMSD matrix. The advantage of *k*-medoids clustering is the use of these medoids as representative cluster conformations. A small number of clusters ($k = 16$) was sufficient to cover a large fraction of the sampled conformational space (precisely the orientational space) and to identify a small number of representatives along a transition path. Here, two paths were chosen and were further examined in more detail. The first one corresponds to clusters labeled **1**, **10**, **8**, and **4** with increasing RMSD relative to the crystal structure. Visual inspection of the conformations reveals that this transition path corresponds to a tilting motion of the proximal ubiquitin from the RavD binding site (Figure 2D). The origin of this tilting motion appears to be the interaction of the ubiquitin main helix with a hydrophobic patch close to RavD residue Ala138. It results in a reduction of protein–protein contact area and a partial dislocation of the proximal ubiquitin. The second path can be followed along clusters **4**, **0**, **14**, and **3** and appears to correspond to a rotation of ubiquitin around the main helix (Figure 2E). This rotation does not lead to an additional increase of the RMSD relative to the crystal structure. However, it repositions the charged ubiquitin residues in a more solvent accessible orientation, which might eventually lead to

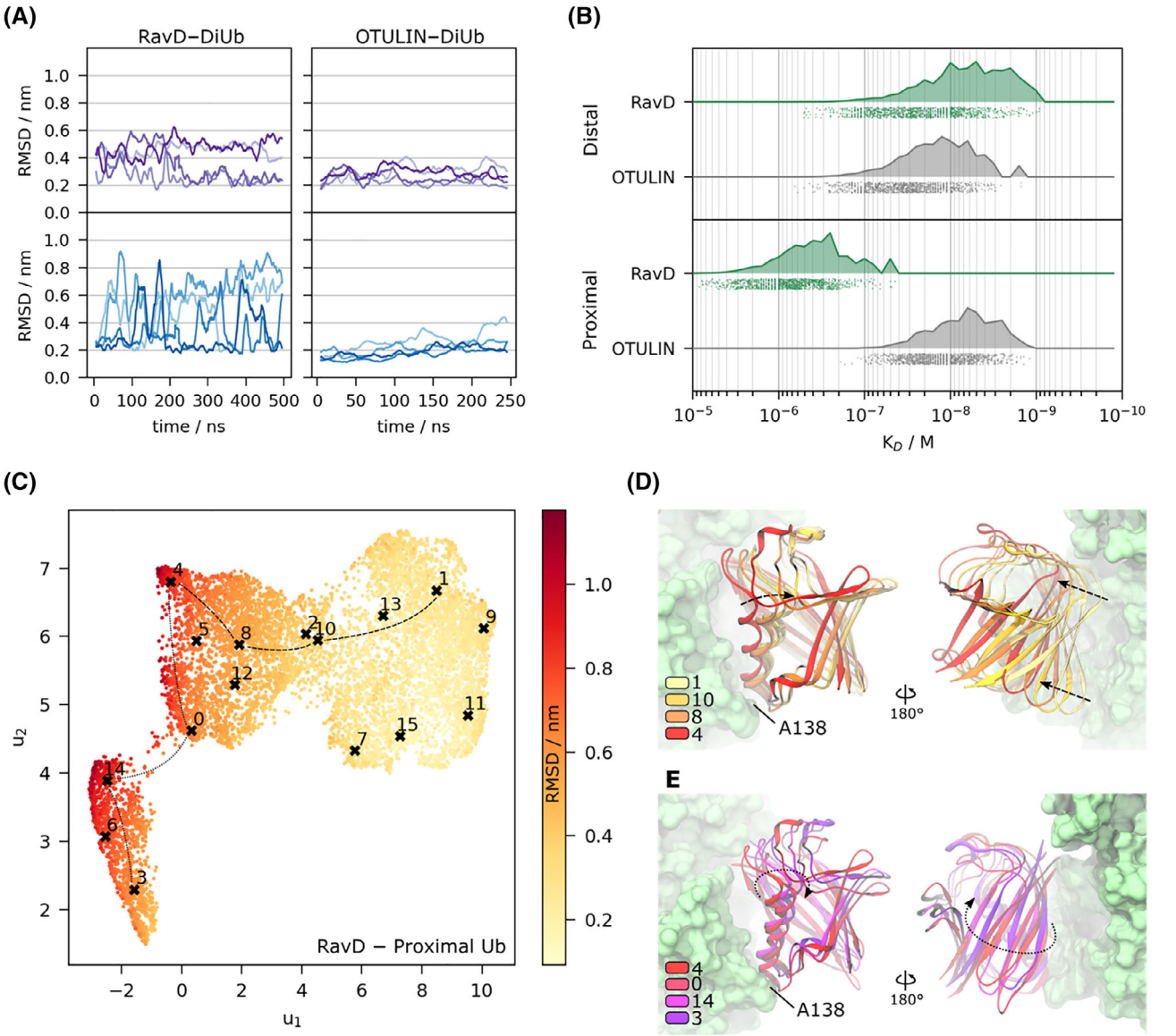


FIGURE 2 (**A**) Cα root mean squared deviation (RMSD) of the individual ubiquitin moieties relative to RavD and OTULIN (see text for details), respectively. The reference structures are the crystal structures of the RavD–DiUb and OTULIN–DiUb complexes. Only moving averages are shown for clarity. Ubiquitin binding to the distal S1 site is given in purple (top), proximal ubiquitin binding in blue (bottom panels). Different shades represent different replicas of simulation. (**B**) Calculated binding free energies (PRODIGY software) for distal and proximal ubiquitin to RavD and OTULIN, respectively. The points are the results for 1000 individual snapshots for RavD and OTULIN. The filled curves represent the probability density functions estimated from 50 equidistant bins (in log space). (**C**) UMAP embedding of the pairwise Cα RMSD distances between the sampled orientations of the proximal ubiquitin of RavD. The points are colored according to their Cα RMSD relative to the crystal structure. The crosses mark the cluster medoids of a *k*-medoids clustering ($k = 16$) on the RMSD matrix. The dashed (in panel **E**) and dotted paths (panel **D**) indicate exemplary transitions along the two embedding vectors. (**D**) and (**E**) snapshots of the transition states along the dashed and dotted transition paths of panel (**C**). The green surfaces represent the distal Ub binding site of RavD. The colored ribbon models represent the distal ubiquitin. The arrows indicate the movements of ubiquitin along the transition path. All motions evolve around RavD residue Ala138 (annotated) as an anchoring point

total ubiquitin dissociation when the isopeptide bond is catalytically cleaved. The combined embedding/clustering analysis gives a strong indication of a dissociation pathway of the proximal ubiquitin that is characterized by an initial tilting away from the RavD protein and a subsequent re-orientation of ubiquitin to expose Ub charged residues to the solvent.

2.3 | Ubiquitin binding affinities to distal and proximal sites

As quantitative differences in Ub binding modes, we estimate the relative binding affinities in terms of K_D values (dissociation constant) to the S1 and S1′ binding sites of RavD and OTULIN. For this, the

interaction-based approach PRODIGY[31,32] was used. From the MD trajectories, unbiased and equidistant snapshots at every 1 ns for OTULIN and every 2 ns for RavD were chosen (Figure 2B).

For the distal Ub binding to the S1 binding site, estimated K_D values between 2 and 12 nM for RavD and between 10 and 20 nM for OTULIN were obtained. This shows that Ub binding to the S1 sites of RavD and OTULIN is energetically comparable. Ubiquitin binding to the S1′ binding sites of RavD and OTULIN, in contrast, differs. The proximal site of OTULIN exhibits a K_D between 4 and 8 nM which is even stronger than for the distal Ub. The affinity of Ub to the S1′ RavD proximal site is substantially weaker with a K_D in the range between 200 and 1000 nM.

For the prediction of binding affinities, PRODIGY is among the most accurate methods to date and superior to the commonly used MM/GBSA approach. In the PRODIGY benchmark set, 81 protein complexes with affinities between −6 and −16 kcal/mol were considered. This positions the calculated binding energies of ubiquitin to the S1 and S1′ sites of RavD and OTULIN well into the predictive range. Experimental data on ubiquitin binding free energies to DUBs are sparse. Dissociation constants (K_D) of ubiquitin binding domains are mostly in the micromolar range.[33,34]

OTULIN shows a high binding affinity of Ub to the S1 and S1′ sites. Ub binding to the distal S1 site on RavD is similar but it is significantly weaker for the proximal S1′ site, which corresponds to the large structural fluctuations seen in the MD trajectories. Apparently, the distal site S1 contributes most to the total binding affinity of DiUb to RavD. A binding affinity in the high nanomolar range can be considered an intermediate between transient and permanent binding.[35] In free solution, the proximal Ub might shuttle between bound and unbound states. Such a motion is, under these particular circumstances, locally restrained by conjugation with the strongly binding distal ubiquitin, which then retains it close to the RavD surface. It can be suggested that first Ub recognition by the distal site of RavD occurs, which then allows proximal Ub binding close to the catalytic triad. The low affinity binding to the S1′ proximal site then tentatively leads to a swift dissociation of mono-Ub from RavD after iso-peptide bond cleavage.

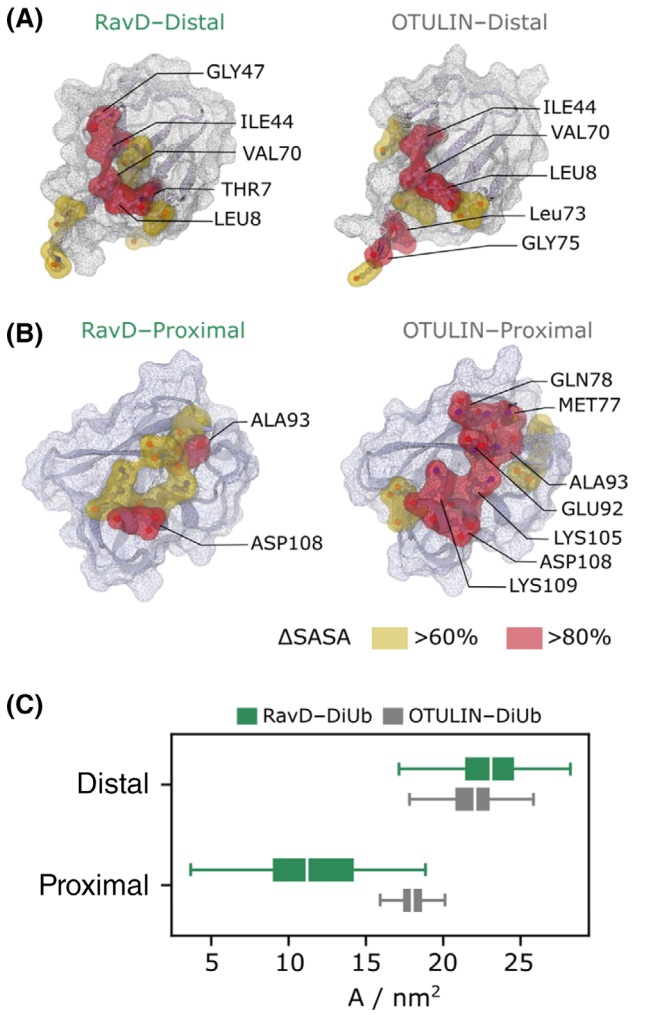


FIGURE 3 Changes in protein solvent-accessible surface area (ΔSASA) upon DiUb binding to RavD and OTULIN. DUB binding to (**A**) distal and (**B**) proximal ubiquitin. The boxplot (**C**) shows the total protein–protein interface areas. Whiskers show smallest and largest areas of the entire ensemble. The box length represents the interquartile range (50% of data). The central line is the median

2.4 | Structural details of RavD and OTULIN distal ubiquitin recognition sites S1

The ubiquitin–DUB protein–protein interaction interfaces and critical hotspots of interactions are determined from pairwise residue interactions and buried surface area calculations (see Section 3). This analysis of long-living protein–protein contacts helps to rationalize the strong binding of Ub to the RavD S1 site. Upon Ub binding, the S1–Ub interface area is 23 nm^2 for RavD and 22 nm^2 for OTULIN (Figure 3C). In RavD, residues which undergo largest changes in solvent accessible surface area (>80%) form a single patch whereas in OTULIN they are distributed (Figure 3A). The calculated distal DUB–ubiquitin interface area is relatively large as compared with 46 single patch protein–protein complexes in the literature[36] and an average interface area of 15 nm^2. Three out of five distal ubiquitin interface residues are conserved among RavD and OTULIN. Upon Ub binding to RavD, residues Gly47, Ile44, Val70, Leu8, and Thr7 become 80% less solvent accessible. In OTULIN, the most buried residues are Ile44, Val70 and Leu8 plus Leu73, and Gly75 which is close to the catalytic center.

Upon formation of the DUB–DiUb complex, not every inter-residue interaction is fully present and not all residues become fully solvent-inaccessible. Some residues form persistent interactions but yet remain partially solvent-accessible at the edge of the protein–protein interface. In the case of RavD, those residues are Lys6, Thr9, Gly10, Lys11, Thr11, Glu34, Gly35, Ile36, Ala46, Gly47, His68, Leu71, Arg74, Gly75, and Gly76 (Figure 4A). OTULIN interacts with similar residues. However, the interactions are made up by fewer residues in number and a larger part of interactions is established by C-terminal residues. Those identified protein–protein interaction patches are characteristic Ub interaction motifs as frequently seen in crystal

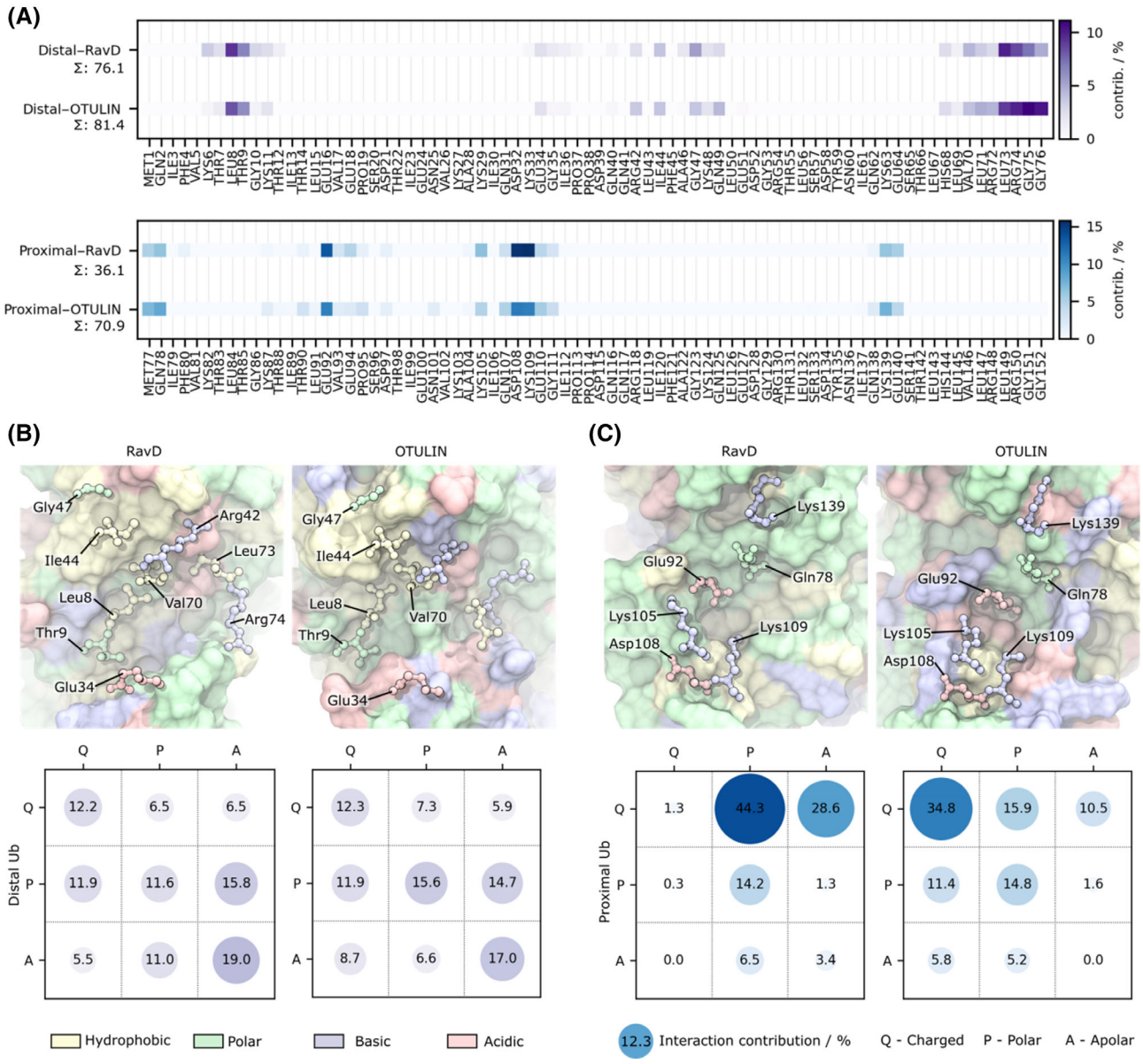


FIGURE 4 Analysis of persistent DUB–DiUb interactions. (**A**) RavD and OUTLIN residue interactions with the distal and proximal ubiquitin molecules. Key protein–protein contact residues of DiUb are given as their residual contribution to the total number of inter-residue interactions in percent (see Section 3 for details). (**B**) Snapshots of distal Ub binding to the S1 sites of RavD and OTULIN and inter-residue interactions between charged (Q), polar (P), and apolar (A) DUB S1 residues with their distal Ub counterparts. (**C**) Snapshots of proximal Ub binding to the S1′ sites of RavD and OTULIN and inter-residue interactions between charged (Q), polar (P), and apolar (A) S1′ DUB residues with their proximal Ub counterparts

structures.[37] The C-terminal patch contributes more in OTULIN, whereas the canonical hydrophobic patch is more pronounced in RavD. Here, for both enzymes, Leu8 has a noticeably higher contribution as compared with the eponymic Ile44. RavD, additionally, shows weak interactions with a rarely reported patch in vicinity of Ile36, which has been discussed as an alternative ubiquitin binding region to maintain the Ile44 patch accessible to other effectors, or to enhance binding affinity (as seen in RavD).[37]

Recognition of the conserved distal ubiquitin interaction patches is achieved by RavD and OTULIN upon employing overall similar exposed residue types (Figure 4B). Of all monitored pairwise interactions, RavD exhibits an interaction residue composition of 30% charged (Q), 30% polar (P), and 40% apolar (A) amino acids. Likewise, OTULIN features 33% Q, 29% P, and 38% A residues. The differences are only marginal and agree with the previously described trend, that RavD has evolutionary developed a high-affinity S1 ubiquitin binding site comparable to that of OTULIN.

2.5 | The proximal S1′ ubiquitin binding site of RavD and OTULIN

RavD binding of Ub via the proximal S1′ site is more flexible and weaker than via the distal S1 site. During the simulations, the most

persistent contacts are mediated by residue Asp108, which becomes deeply buried upon formation of the protein complex (Figure 3B), and neighboring residues 105–111. This patch of residues is also found in OTULIN, but there it is significantly larger and more extended. A second patch of interacting residues is made up by ubiquitin residues Glu92, Val93, Glu94 and residues Pro95, and Asp97 in close vicinity to the Asp108 patch.

In OTULIN, these interactions are responsible for the substrate-assisted activation mechanism: Substrate DiUb residue Glu92 inserts into the catalytic core and stabilizes the catalytically competent conformational state of the triad.

RavD binds to the same proximal Ub residues as OTULIN (Figure 4A), but its interactions are mainly mediated by different residue types (Figure 4C). In particular, 56% of OTULIN's proximal–Ub interactions are made up by electrostatic interactions (charged OTULIN-charged Ub residues 35%; charged OTULIN with polar proximal Ub residues 11%) and constitute highly favorable, solvent-buried electrostatic interactions.[38] RavD exposes mostly polar (44%) and hydrophobic residues (29%) to interact with charged residues of the proximal Ub. This apparently weaker binding of key ubiquitin residues leads to a reduced buried surface area of only 12 nm^2 compared to 18 nm^2 for OTULIN (Figure 3B).

The specific, electrostatic interactions in OTULIN can exceptionally well be seen at the proximal ubiquitin recognition patch at the end of the helix residues Asp108, Lys109, and Glu110 (Figure S4). This motif interacts with RavD residues Ala138, Try147, Val135, Gly32, Leu134, Pro136, and His137—all of which are polar or hydrophobic. OTULIN, on the other hand, employs mostly charged residues Lys102, Glu95, and Arg97 to form a complementarily charged surface patch.

2.6 | Substrate-binding to the catalytic triad in RavD and OTULIN

A substrate-assisted activation mechanism of RavD was disregarded by the low global RMSD between the free RavD and when in complex with its substrate. However, a definite statement can only be made when analyzing the arrangement of the catalytic triad residues in the absence and presence of M1-linked DiUb.

Inter-residue distances of the catalytic triad residues as well as the distance between the nucleophile cysteine and the substrate carbonyl carbon of the scissile bond are clear indicators for the state of activation. We stress once again that the RavD–DiUb complex was obtained as a double mutant of the substrate in which the two C-terminal glycine residues of the distal Ub were substituted by serine residues (DiUb–GGSS). Those mutations were necessary to avoid substrate cleavage and product dissociation. The OTULIN–DiUb complex, on the other hand, was crystallized with a mutation within the catalytic triad of OTULIN (C129A). Thus, we re-constituted the physiological RavD–DiUb and compare with re-constituted WT OTULIN–DiUb.

Molecular dynamics simulations of the RavD and OTULIN DUBs in absence and presence of physiological substrates plus the RavD–DiUbGGSS substrate analogue were performed. Critical inter-residue distances between Cys–His and His–Ser (His–Asn for OTULIN, respectively) were monitored. For unbound RavD, we mostly found conformations in which the Cys–His distance is large and not in a catalytically competent configuration (Figure 5A,B). Histidine and serine distances fluctuate between 0.3 and 0.9 nm.

For RavD in complex with DiUbGGSS, the populations shift toward shorter His(H)–Ser(S) distances, and short Cys(C)–His(H) distances occur more frequently. However, there is still a distinct local energy minimum at a His–Ser distance of 0.8 nm and a Cys–His distance of 0.6 nm. This shows that RavD–DiUbGGSS rarely adopts a catalytically competent conformation since the inter-residue distances remain large. Upon re-constitution of RavD–DiUb, the energy minimum at His–Ser of 0.8 nm vanishes. Instead, mostly conformations with short His–Ser distances are occurring and the Cys–His distances are between 0.4 and 0.8 nm. For OTULIN–DiUb, the inter-residue distances between Cys–His and His–Ser are below 0.4 nm throughout most of the trajectories, whereas they are significantly larger (0.7 and 0.8 nm) in the crystal structure of substrate-free OTULIN.

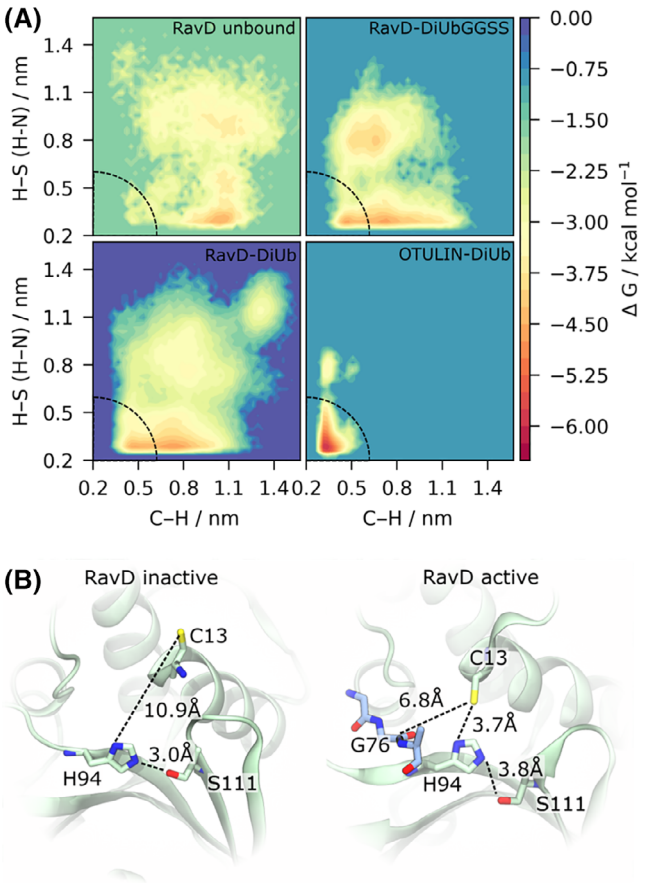


FIGURE 5 Free energy contour maps (**A**) of relevant inter-residue distances of catalytic residues for unbound RavD, and substrate-bound RavD and OTULIN (see text for details). The area in which the catalytic triad is in a catalytically competent state is marked. The free energy landscapes are truncated at the highest sampled energy. (**B**) Representative MD snapshots of catalytically active and inactive states based on the energy wells of free RavD and RavD–DiUb

When applying a contact cutoff criterion of 0.6 nm, the probability of occurrence of a catalytically competent arrangement of the catalytic triad is calculated. The probalility for an activated catalytic triad conformation is 0.02 for free RavD, 0.16 for RavD–DiUbGGSS, 0.20 for RavD–DiUb, and 0.98 for OTULIN–DiUb (Figure S2). Compared with apo-RavD, the binding of DiUbGGSS, and even more so the binding of WT DiUb increases the number of active conformations. However, the ratio of competent conformations is still smaller than in OTULIN. Apparently, substrate binding induces a structural approach of the catalytic cysteine towards histidine, which then leads to the assembly and activation of the catalytic triad in RavD. The introduction of two polar serine residues in the double mutant DiUbGGSS apparently obstructs the complete assembly of the catalytic triad. Also the Cys-substrate nucleophilic attack distance is too large for the reaction to occur. The probability for the catalytic cysteine to approach the carbonyl carbon of Gly(G)76 or Ser(S)76, respectively, closer than 0.6 nm is negligibly small in both the complexes RavD–DiUbGGSS and RavD–DiUb. In the OTULIN–DiUb complex it is significant with a value of 0.7 (Figure S2).

Based on a critical analysis of the crystal structures and from the analysis of MD simulations, we observe a strong indication that the activation of RavD is in fact substrate-assisted: The binding of DiUb clearly leads to stabilization of the catalytic triad and a structural rearrangement with shorter inter-residues distances which are closer to a catalytically competent state. The simulations, however, do not sample the full transition of the catalytic triad and the complete accommodation of the terminal Gly75–Gly76–Met1 ubiquitin motif into the catalytic cleft. The experimentally reported exclusive selectivity of RavD towards M1-linked di-ubiquitin cannot be explained by the high affinity distal ubiquitin binding site alone and substrate-induced activation appears to be essential for this selective recognition.

The use of the inactive co-crystal structure of RavD–DiUbGGSS as a starting point for molecular dynamics investigations of the reconstituted substrate makes the sampling very cumbersome and the sampling of the transition slow. It represents a deep local energy minimum from which a transition to the fully competent catalytic arrangement is only achievable on a timescale of hundreds of microseconds. Enhanced sampling methods might be able to accelerate such a transition[39] and will be part of further studies.

2.7 | Structural and functional ambiguities in substrate-modified DUB structures

OTULIN seems to be perfectly adapted to specifically recognize and cleave M1-linear polyubiquitin chains. In OTULIN, both the distal and proximal ubiquitin binding sites are well defined and form stable interactions with DiUb throughout the MD trajectories. The catalytic site shows almost no dynamic changes and the nucleophilic attack distance to the substrate remains short and persistent throughout the simulations. For RavD, our simulations reveal a well-defined distal ubiquitin binding site to which Ub binds with high affinity. The proximal site binding is more dynamic and has a low affinity towards Ub. The catalytic inter-residue distances remain large and are highly fluctuating, even in presence of a substrate analogue. In such a scenario, the activity of RavD toward linear M1-linked Ub chains cannot be rationalized due to the significantly reduced lower affinity of the proximal site and the small number of catalytically competent conformations.

Nevertheless, the bacterial effector protein RavD from *L. pneumophila* is a deubiquitinylase with a high M1-linkage selectivity. Experimental assays of ubiquitin activity and selectivity were performed on the *L. pneumophila* strain Philadelphia I, which did not crystallize in the substrate-protein co-complex. *L. pneumophila* Corby, however, when in complex with DiUbGGSS, yielded a stable protein–protein complex with sufficient X-ray diffractions and yielded a protein-protein structure at 2.05 Å resolution. The RavD Corby strain shows two mutations in the hydrophobic patch of the distal ubiquitin S1 binding site that lead to a stronger binding of the distal ubiquitin binding relative to Philadelphia I.

The conservation of the proximal Ub binding site S1′, however, is less pronounced and Ub binding more motile in RavD. A few mutations in close vicinity of RavD residue Asp108 from polar or hydrophobic to charged amino acid residues might enhance its binding affinity. The large differences of the S1′ binding sites, on the other hand, might be exploited to develop RavD-selective therapeutics by inhibiting the proximal Ub binding and suppress its activity toward selective M1-linked ubiquitin chains.

Upon co-crystallization of RavD with the non-hydrolysable DiUbGGSS substrate analogue, two new polar sidechains with hydrogen-donating and accepting properties were introduced. We argue that the introduction of these larger and polar serine residues obstructs full substrate access to the active site. The serine residues form strong interactions with RavD surface residues Gln42 and Ala92, which hinders substrate insertion into the catalytic groove, and hence a conformational change and activation of the catalytic triad. Such mechanistic insight is only possible from a detailed structural analysis of protein crystal structures augmented by extensive molecular dynamics simulations. The absence of large global conformational changes upon complexation is not a sufficient criteria to disregard substrate-assisted activation and catalysis. We suggest that, instead of mutating residues of the DiUb substrate, a modification of the proteolytic cysteine residue into a non-reactive alanine could give a more conclusive structure close to the physiological state.

3 | METHODS

3.1 | System setup

Protein structures of OTULIN (3ZNV) OTULIN–DiUb (3ZNZ), RavD (6NII), and RavD–DiUbGGSS (6NJD) were retrieved from the PDB. In the protein–protein complexes the co-crystallized di-ubiquitin was Met-1 linked. OTULIN–DiUb has four additional N-terminal residues and there are four missing C-terminal residues in RavD–DiUb. However, as the termini of DiUb do not interact with the enzyme, this is

not relevant for di-ubiquitin recognition. VMD 1.9.3[40] was used for parameter assignment and system setup including solvation with TIP3P water, neutralized and ionization to 0.15 M NaCl. All histidine residues were treated as neutral and mono-protonated at the delta nitrogen. The protein termini were charged. The periodic box was set to $10 \times 10 \times 10$ nm^3. CHARMM36m[41] was utilized as the force field.

3.2 | Molecular dynamics simulation

Molecular mechanics and molecular dynamics simulations were carried out using openMM 7.4.1.[42] We used PME[43] for long range electrostatic interactions, a non-bonded cutoff of 1.2 nm and a switch distance of 1.0 nm. Covalent bonds to hydrogen atoms were constrained. The molecular system was initially minimized for 5000 steps. Equilibration was carried using Langevin integration[44] with a 300 K thermostat, a 1.0 ps^{-1} friction coefficient and 1 fs time step. During equilibration, all the CA atoms were restrained with a force constant of 500 kcal/mol nm^2. Initial velocities were set according to a Maxwell distribution corresponding to 300 K. Production data were generated using the same integrator but with a time step of 2 fs. Additionally, a Monte–Carlo barostat[45] was engaged to maintain a constant pressure of 1 bar. The barostat was coupled every 25 integration steps. Trajectory snapshots were saved every 0.2 ns. Four replicate simulations were performed with individual initial velocity distributions. Every replicate covered 250 ns (500 ns for RavD–DiUb), accumulating to 1 μs (2 μs for RavD–DiUb) of total sampling time. In the case of RavD–DiUb, the protocol was adapted to consider large initial fluctuations of the proximal Ub. Here, we performed an initial 500 ns equilibration run from which the four replicates of 500 ns each were initiated (with new and different velocity distributions). The integration was performed on Nvidia GTX1080 GPUs in single precision mode. Topology, coordinates and MD run input files are stored and available upon request.

3.3 | Trajectory analysis

For trajectory analysis, MDTraj 1.9.5,[46] MDAnalysis 0.20.1,[47,48] NumPy 0.51.2,[49] and SciPy 1.2.1[50] were used. Visualization was achieved with VMD 1.9.3[40] and MatplotLib 3.0.2.[51]

Ubiquitin RMSDs were calculated using the CA coordinates. As reference coordinates we used the initial model from the crystal structure. Before the RMSD calculation, the whole protein complex was aligned by the CA atoms of the DUB (RavD, or OTULIN respectively). For the computation of the RavD proximal Ub pairwise RMSD matrix, 10 000 frames where chosen (every 200 ps). The two-dimensional embedding was achieved using the UMAP algorithm.[29] Clustering was performed with k-medoids[52] using *k*-means++ initialization[30] and $k = 16$ clusters.

The binding free energies were estimated using PRODIGY[31] for equidistant multiple conformations extracted from the MD simulations. For RavD, 1.000 conformations and for OTULIN 1.000 conformations (every 1 ns) were chosen.

The interface areas were calculated from the differences in the solvent accessible surface areas (SASA), that is, the sum of the SASAs of two separate proteins minus the SASA of the protein–protein complexes. For the SASAs of the substrate-free proteins also the complex trajectory was used. The per-residue SASAs were computed using the algorithm of Shrake and Rupley[53] with a probe radius of 0.14 nm and 512 sphere points. The residue-wise buried areas were computed as the relative differences between SASA in the Ub-free and the substrate-bound forms.

For intermolecular interactions, heavy-atom contacts were calculated with MDTraj, employing a distance cutoff criterion of 0.55 nm. The residue type contributions were calculated as the ratio of contacts of the single residues to the total contacts. The classification followed the standard convention. Histidine, cysteine, glycine, and proline were always considered as polar. The free energy maps were constructed from the 2D joint probability distribution functions (PDF) of the interatomic distances of Cys-S:His-Nδ and HIS-Nε:Ser-Oγ (HIS-Nε:ASN-Oγ for OTULIN). The PDFs were estimated using binning to 50×50 bins. The free energy difference was calculated as the negative natural logarithm of the PDF.

ACKNOWLEDGMENTS

Financial support by the Max Planck Society for the Advancement of Science to ESN and MS is gratefully acknowledged. This work was supported in part by grant ZS/2016/04/78155 from the European Union Program European Regional Development Fund of the Ministry of Economy, Science and Digitalization in Saxony-Anhalt within the Center of Dynamic Systems to MN. This work is part of the COST Action CA20133 "ProteoCure" (European Cooperation in Science and Technology). Open Access funding enabled and organized by Projekt DEAL.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/prot.26286.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Matthias Stein https://orcid.org/0000-0001-7793-0052

REFERENCES

1. Komander D, Rape M. The ubiquitin code. *Annu Rev Biochem*. 2012;81:203-229. doi:10.1146/annurev-biochem-060310-170328
2. Varshavsky A. The ubiquitin system, autophagy, and regulated protein degradation. *Annu Rev Biochem*. 2017;86:123-128. doi:10.1146/annurev-biochem-061516-044859
3. Zinngrebe J, Montinaro A, Peltzer N, Walczak H. Ubiquitin in the immune system. *EMBO Rep*. 2014;15:322. doi:10.1002/embr.201470030
4. Ben-Neriah Y. Regulatory functions of ubiquitination in the immune system. *Nat Immunol*. 2002;3:20-26. doi:10.1038/ni0102-20
5. Fiil BK, Gyrd-Hansen M. The Met1-linked ubiquitin machinery in inflammation and infection. *Cell Death Differ*. 2021;28:557-569. doi:10.1038/s41418-020-00702-x

6. Pickart CM. Mechanisms underlying Ubiquitination. *Annu Rev Biochem*. 2001;70:503-533. doi:10.1146/annurev.biochem.70.1.503
7. Swatek KN, Komander D. Ubiquitin modifications. *Cell Res*. 2016;26:399-422. doi:10.1038/cr.2016.39
8. Komander D. The emerging complexity of protein ubiquitination. *Biochem Soc Trans*. 2009;37:937-953. doi:10.1042/BST0370937
9. Mevissen TET, Komander D. Mechanisms of deubiquitinase specificity and regulation. *Annu Rev Biochem*. 2017;86:159-192. doi:10.1146/annurev-biochem-061516-044916
10. Komander D, Clague MJ, Urbé S. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol*. 2009;10:550-563. doi:10.1038/nrm2731
11. Clague MJ, Urbé S, Komander D. Breaking the chains: deubiquitylating enzyme specificity begets function. *Nat Rev Mol Cell Biol*. 2019;20:338-352. doi:10.1038/s41580-019-0099-1
12. Mevissen TET, Hospenthal MK, Geurink PP, et al. OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell*. 2013;154:169-184. doi:10.1016/j.cell.2013.05.046
13. Mevissen TET, Kulathu Y, Mulder MPC, et al. Molecular basis of Lys11-polyubiquitin specificity in the deubiquitinase Cezanne. *Nature*. 2016;538:402-405. doi:10.1038/nature19836
14. Edelmann MJ, Iphöfer A, Akutsu M, et al. Structural basis and specificity of human otubain 1-mediated deubiquitination. *Biochem J*. 2009;418:379-390. doi:10.1042/BJ20081318
15. Fiil BK, Damgaard RB, Wagner SA, et al. OTULIN restricts Met1-linked ubiquitination to control innate immune signaling. *Mol Cell*. 2013;50:818-830. doi:10.1016/j.molcel.2013.06.004
16. Keusekotten K, Elliott PR, Glockner L, et al. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell*. 2013;153:1312-1326. doi:10.1016/j.cell.2013.05.014
17. Jahan AS, Elbæk CR, Damgaard RB. Met1-linked ubiquitin signalling in health and disease: inflammation, immunity, cancer, and beyond. *Cell Death Differ*. 2021;28:473-492. doi:10.1038/s41418-020-00676-w
18. Schlüter D, Schulze-Niemand E, Stein M, Naumann M. Ovarian tumor domain proteases in pathogen infection. *Trends Microbiol*. 2021. doi:10.1016/j.tim.2021.04.002
19. Schubert AF, Nguyen JV, Franklin TG, et al. Identification and characterization of diverse OTU deubiquitinases in bacteria. *EMBO J*. 2020;39:e105127. doi:10.15252/embj.2020105127
20. Bailey-Elkin BA, van Kasteren PB, Snijder EJ, Kikkert M, Mark BL. Viral OTU deubiquitinases: a structural and functional comparison. *PLoS Pathog*. 2014;10:e1003894. doi:10.1371/journal.ppat.1003894
21. Franklin TG, Pruneda JN. Bacteria make surgical strikes on host ubiquitin signaling. *PLoS Pathog*. 2021;17:e1009341. doi:10.1371/journal.ppat.1009341
22. Wan M, Wang X, Huang C, et al. A bacterial effector deubiquitinase specifically hydrolyses linear ubiquitin chains to inhibit host inflammatory signalling. *Nat Microbiol*. 2019;4:1282-1293. doi:10.1038/s41564-019-0454-1
23. Schulze-Niemand E, Naumann M, Stein M. The activation and selectivity of the Legionella RavD deubiquitinase. *Front Mol Biosci*. 2021;8:770320. https://doi.org/10.3389/fmolb.2021.770320
24. Hermanns T, Hofmann K. Bacterial DUBs: deubiquitination beyond the seven classes. *Biochem Soc Trans*. 2019;47:1857-1866. doi:10.1042/BST20190526
25. Dall'Acqua W, Carter P. Substrate-assisted catalysis: molecular basis and biological significance. *Protein Sci*. 2000;9:1-9. doi:10.1110/ps.9.1.1
26. Hosfield CM, Elce JS, Davies PL, Jia Z. Crystal structure of calpain reveals the structural basis for Ca(2+)-dependent protease activity and a novel mode of enzyme activation. *EMBO J*. 1999;18:6880-6889. doi:10.1093/emboj/18.24.6880
27. Rakers C, Bermudez M, Keller BG, Mortier J, Wolber G. Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations? *WIREs Comput Mol Sci*. 2015;5:345-359. doi:10.1002/wcms.1222
28. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol*. 1994;235:625-634. doi:10.1006/jmbi.1994.1017
29. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*. 2018;1802.03426.
30. Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, New Orleans, Lousiana, 2007;1027–1035.
31. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*. 2016;32:3676–3678. https://doi.org/10.1093/bioinformatics/btw514
32. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein–protein complexes. *Elife*. 2015;4:e07454.
33. Hurley JH, Lee S, Prag G. Ubiquitin-binding domains. *Biochem J*. 2006;399:361-372. doi:10.1042/BJ20061138
34. Sokratous K, Roach LV, Channing D, et al. Probing affinity and ubiquitin linkage selectivity of ubiquitin-binding domains using mass spectrometry. *J Am Chem Soc*. 2012;134:6416-6424. doi:10.1021/ja300749d
35. Acuner Ozbabacan SE, Engin HB, Gursoy A, Keskin O. Transient protein-protein interactions. *Protein Eng Des Sel*. 2011;24:635-648. doi:10.1093/protein/gzr025
36. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins*. 2002;47:334-343. doi:10.1002/prot.10085
37. Winget JM, Mayor T. The diversity of ubiquitin recognition: hot spots and varied specificity. *Mol Cell*. 2010;38:627-635. doi:10.1016/j.molcel.2010.05.003
38. Zhou H-X, Pang X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem Rev*. 2018;118:1691-1741. doi:10.1021/acs.chemrev.7b00305
39. Miao Y, Huang Y-MM, Walker RC, McCammon JA, Chang C-EA. Ligand binding pathways and conformational transitions of the HIV protease. *Biochemistry*. 2018;57:1533-1541. doi:10.1021/acs.biochem.7b01248
40. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14:33-38. doi:10.1016/0263-7855(96)00018-5
41. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14:71-73. doi:10.1038/nmeth.4067
42. Eastman P, Swails J, Chodera JD, et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. 2017;13:e1005659. doi:10.1371/journal.pcbi.1005659
43. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103:8577-8593. doi:10.1063/1.470117
44. Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD. Langevin stabilization of molecular dynamics. *J Chem Phys*. 2001;114:2090-2098. doi:10.1063/1.1332996
45. Åqvist J, Wennerström P, Nervall M, Bjelic S, Brandsdal BO. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem Phys Lett*. 2004;384:288-294. doi:10.1016/j.cplett.2003.12.039
46. McGibbon RT, Beauchamp KA, Harrigan MP, et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys J*. 2015;109:1528-1532. doi:10.1016/j.bpj.2015.08.015
47. Gowers R, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, Domanski J, Dotson DL, Buchouxk S, Kenney IM, Beckstein O.

MDAnalysis: a python package for the rapid analysis of molecular dynamics simulations. In Proceedings of the 15th Python in Science Conference (SciPy2016), pp. 98–105.
48. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32:2319-2327. doi:10.1002/jcc.21787
49. Harris CR, Millman KJ, Van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585:357-362. doi:10.1038/s41586-020-2649-2
50. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17:261-272. doi:10.1038/s41592-019-0686-2
51. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9:90-95. doi:10.1109/MCSE.2007.55
52. Park H-S, Jun C-H. A simple and fast algorithm for K-medoids clustering. *Expert Syst Appl*. 2009;36:3336-3341. doi:10.1016/j.eswa.2008.01.039
53. Shrake A, Rupley JA. Environment and exposure to solvent of protein atoms, lysozyme and insulin. *J Mol Biol*. 1973;79:351-371. doi:10.1016/0022-2836(73)90011-9

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Schulze-Niemand E, Naumann M, Stein M. Substrate-assisted activation and selectivity of the bacterial RavD effector deubiquitinylase. *Proteins*. 2022;90(4):947-958. doi:10.1002/prot.26286