Saturation scanning of ubiquitin variants reveals a common hot spot for binding to USP2 and USP21

Isabel Leung[a], Ayelet Dekel[b], Julia M. Shifman[b,1], and Sachdev S. Sidhu[a,1]

[a]Banting and Best Department of Medical Research and Department of Molecular Genetics, The Donnelly Centre, University of Toronto, Toronto, ON, Canada M5S 3E1; and [b]Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 9190401, Israel

Edited by David Baker, University of Washington, Seattle, WA, and approved June 20, 2016 (received for review December 14, 2015)

A detailed understanding of the molecular mechanisms whereby ubiquitin (Ub) recognizes enzymes in the Ub proteasome system is crucial for understanding the biological function of Ub. Many structures of Ub complexes have been solved and, in most cases, reveal a large structural epitope on a common face of the Ub molecule. However, owing to the generally weak nature of these interactions, it has been difficult to map in detail the functional contributions of individual Ub side chains to affinity and specificity. Here we took advantage of Ub variants (Ubvs) that bind tightly to particular Ub-specific proteases (USPs) and used phage display and saturation scanning mutagenesis to comprehensively map functional epitopes within the structural epitopes. We found that Ubvs that bind to USP2 or USP21 contain a remarkably similar core functional epitope, or "hot spot," consisting mainly of positions that are conserved as the wild type sequence, but also some positions that prefer mutant sequences. The Ubv core functional epitope contacts residues that are conserved in the human USP family, and thus it is likely important for the interactions of Ub across many family members.

USP | ubiquitin | phage display | saturation scanning | hot spots

Ubiquitin (Ub) is a posttranslational modifier that is covalently attached to substrate lysines, either singly or in chains of differing topologies. These modifications serve to target substrate for degradation or to alter protein–protein interactions that orchestrate a variety of cellular processes, including gene silencing, receptor internalization (1, 2), DNA repair (3), and cell cycle progression (4, 5). Ub attachment to substrates is catalyzed by the sequential action of three classes of enzymes: E1, E2, and E3 (6–8). Ubiquitination also can be reversed by deubiquitinases (DUBs), proteases that cleave Ub moieties from substrates. To date, 79 human DUBs have been identified (9). In addition to the plethora of enzymes that recognize Ub, there are at least 16 structural families of Ub-binding domains (UBDs), which recognize Ub noncovalently and serve to translate Ub modifications into signaling cascades (10–12). Thus, the biological function of Ub is the product of finely balanced interactions between Ub and diverse enzymes and binding domains.

Ub binds to thousands of proteins in the human proteome, but typically with low affinity, in the 10^{-5}–10^{-3} M range. Structural studies have shown that most interactions occur on the β-sheet face (12, 13), which is highly versatile in recognizing a wide variety of domain folds and surface topologies. Despite the low affinity, many Ub interactions bury a substantial surface area that can exceed 2,000 Å^2 (14), defying the conventional view that buried surface area positively correlates with the binding affinity of protein–protein interactions (15).

We have shown that the affinity of particular Ub–protein interactions can be enhanced by several orders of magnitude by mutating the surface to produce UB variants (Ubvs) that function as potent and specific inhibitors of the enzymes they target (14). Although that work demonstrated that Ub interactions are not optimal and can be improved by mutations, we still do not understand which residues contribute the most to binding energetics and which residues are suboptimal for binding. Many questions remain to be resolved; for example: What is the functional epitope? How are the binding determinants spatially organized within the large interface? How do binding energetics differ depending on the interacting partner?

To understand molecular recognition by Ub, we need detailed molecular binding landscapes of Ub–protein interactions. Such binding landscapes can be obtained by introducing all possible single mutations in the binding site and assessing the consequent effects on affinities (16–21). However, the low affinities of typical Ub–protein interactions (11) make it difficult to accurately measure the effects of mutations, and thus detailed empirical binding landscapes of Ub–protein interactions have not been reported.

Here we used two Ubvs, which were evolved as high-affinity inhibitors of different Ub-specific proteases (USPs), as proxies to explore the functional details of Ub–protein interactions (14). USPs constitute the largest class of DUBs, and several have been implicated as potential therapeutic targets (22, 23). USPs share a structurally similar catalytic domain, which binds Ub and positions its C-terminal tail into the conserved catalytic cleft. We analyzed Ubv.2.1 and Ubv.21.4, distinct triple mutants with high affinity for USP2 and USP21, respectively (14). Comparisons of crystal structures of each Ubv or wild type (WT) Ub (Ub.wt) in complex with USP2 or USP21 have revealed that each Ubv binds its cognate enzyme in a manner virtually identical to the binding mode of Ub.wt (14). Thus, the fact that only three mutations yield Ubvs that bind in a similar manner but with dramatically enhanced affinities relative to Ub.wt makes these Ubvs ideal candidates for mutational studies to shed light on both the

Significance

Although ubiquitin (Ub) interactions are essential to cell physiology, there is a paucity of functional data on how Ub recognizes the diverse enzymes in the Ub proteasome system. We subjected two high-affinity Ub variants (Ubvs) to saturation scanning analysis to obtain comprehensive information on how they interact with Ub-specific proteases (USPs). Our data reveal the molecular details of how individual Ubv side chains mediate weak but specific interactions with these proteases. We find that the enzymes are recognized by similar functional epitopes on the Ubvs, and that these epitopes interact with a region that is conserved across the human USP family, suggesting that Ub may interact with many USPs by using a common core functional epitope.

Author contributions: I.L., J.M.S., and S.S.S. designed research; I.L., A.D., and J.M.S. performed research; I.L., J.M.S., and S.S.S. analyzed data; and I.L., J.M.S., and S.S.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jshifman@mail.huji.ac.il or sachdev.sidhu@utoronto.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1524648113/-/DCSupplemental.

BIOCHEMISTRY

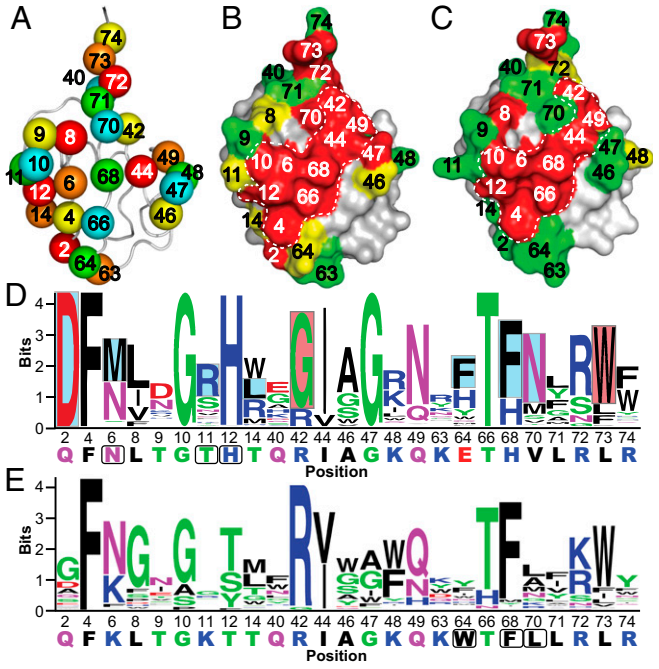

Fig. 1. Saturation scans of Ubvs. (*A*) Library design. The USP binding site contains 25 residues distributed across three regions: region 1 (residues 2–14), region 2 (residues 40–49), and region 3 (residues 63–74). Five libraries were designed to scan five residues each, in a manner that maximized the distances between residues in any one library. Scanned residues are shown as spheres and are color-coded according to the library that they share. The structure was derived from the coordinates of WT Ub (PDB ID code 1UBQ). (*B* and *C*) The results of saturation scanning mapped on the structure of Ubv.2.1 (PDB ID code 3V6C) (*B*) or Ubv.21.4 (PDB ID code 3MTN) (*C*). The Ubv is shown as a surface colored according to bit values derived from amino acid frequencies (Fig. S2), as follows: green ≤ 2; 2 < yellow <2.7; red ≥ 2.7; gray, unscanned. The white dashed lines demarcate the core functional epitope. (*D* and *E*) The results of saturation scanning presented as sequence logos for Ubv.2.1 (*D*) or Ubv.21.4 (*E*). The sequence of each Ubv is shown at the bottom, and boxes demarcate sequences that differ from Ub.wt. Blue and red boxes indicate sequence changes in Ubv.2.1 predicted to enhance inhibition of USP2 for which enzyme inhibition assays did and did not validate predictions, respectively (Table 1).

molecular bases for enhanced affinity and on the molecular details of natural Ub–USP interactions.

To obtain a comprehensive assessment of the mutational tolerance of each Ubv-binding surface, without having to make and test thousands of point mutants, we used combinatorial saturation scanning by phage display. This method uses combinatorial libraries of protein variants designed for the rapid and facile assessment of the effects of all possible point mutations across a large protein-binding surface and has been used previously to survey the binding energy landscape of human growth hormone (hGH) for its receptor (hGHR) (17). Saturation scanning analysis of the USP-binding sites of Ubv.2.1 and Ubv.21.4 reveals many common elements and some differences between the two interactions. Moreover, interpretation of the results in the context of the human USP family shows that many conserved functional interactions likely are generalizable to most family members. We have recently reported structures of additional Ubvs in complex with E3 ligases of the Homologous to E6AP C terminus (HECT), Skp1 Cul1 F-box (SCF), and Really Interesting New Gene (RING) families (24, 25), and the extension of the methods described here to these and other types of Ub–protein interactions may identify commonalities and differences in how Ub recognizes diverse structural folds to mediate biological effects.

Results

Saturation Scanning of Ubv.2.1 and Ubv.21.4. In saturation scanning, precisely designed combinatorial libraries of phage-displayed protein variants are used to assess the functional importance of individual residues for binding to a partner protein (26) The basic principle of the method is that the frequency of a particular amino acid at a particular position within a protein interaction site reflects its favorability for the function of the site. By allowing all 20 genetically encoded amino acids to compete against one another, the relative frequencies can be used to simultaneously compare the effects of all possible substitutions across the interface.

We defined the USP-binding site on Ub as 25 residues across the β-sheet face, and for each Ubv, we divided these residues into five groups of five for randomization in phage-displayed libraries (Fig. 1*A* and Fig. S1). To minimize cooperative effects between mutated positions, each library was designed to randomize positions that are not in direct contact with one another (17). Ubv.2.1 and Ubv.21.4 libraries, displayed on M13 bacteriophage as fusions to the N terminus of the gene-3 minor coat protein, were subjected to selections for binding to the catalytic domain of USP2 and USP21, respectively. Individual clones were tested for cognate USP binding by phage ELISA, and 50–100 clones with specific binding signals were sequenced from each library. Binding sequences were used to create a position weight matrix (PWM) consisting of frequencies of each amino acid type at each position (Fig. S2). Each matrix column depicts the amino acid preference at each position as a probability distribution. From this PWM, the distribution at each position was visualized as a sequence logo (Fig. 1 *D* and *E*). The importance of each Ubv position to USP binding was quantified as a bits score, plotted along the *y*-axis of the sequence logo (Fig. 1 *D* and *E*). The bits score measures the amount of information required to describe the amino acid frequency pattern at a particular position and is related to Shannon entropy, which measures the degree of randomness within a population (27) and is routinely used to identify and predict DNA- (28, 29), RNA- (30), and protein-binding motifs (31–33). For a frequency distribution of 20 amino acids, the bits score varies between 0 and 4.32 for positions that are completely random and completely conserved, respectively.

To assess the accuracy of the saturation scanning data, we focused on inhibitors of USP2 and measured the effects of Ub.wt and a panel of Ubvs on the hydrolysis of the model substrate Ub-AMC (Table 1 and Fig. S3). Ub.wt exhibited an IC_{50} value of ~60 μM, and Ubv.2.1 (K6N/K11T/T12H) was ~1,000-fold more potent (IC_{50} = 56 nM). Among the three mutated positions in Ubv.2.1, position 11 exhibited a low bits score and Thr was relatively uncommon (Fig. 1*D*). Thus, we predicted that Thr^{11} does not contribute significantly to the enhanced affinity of Ubv.2.1

Table 1. IC_{50} values for inhibition of USP2 by Ub.wt and Ubvs

Ub substitution	IC_{50}, nM	$IC_{50,K6N/T12H}/IC_{50}$
WT	60,000 ± 20,000	0.001
K6N	200 ± 10	0.3
T12H	24,000 ± 1,000	0.002
K6N/T12H	58 ± 7	1.0
K6N/T12H/K11T	56 ± 2	1.0
K6N/T12H/Q2D	22 ± 4	2.5
K6N/T12H/N6M	70 ± 20	0.8
K6N/T12H/KIIR	12 ± 5	4.7
K6N/T12H/T14L	21 ± 5	2.7
K6N/T12H/E64F	32 ± 3	1.8
K6N/T12H/H68F	44 ± 7	1.3
K6N/T12H/V70N	52 ± 6	1.1
K6N/T12H/R42G	1,400 ± 200	0.04
K6N/T12H/L73W	2,300 ± 800	0.02

for USP2, and confirmed this by showing that a variant containing two substitutions (K6N/T12H) was virtually equipotent (IC_{50} = 58 nM) to Ubv.2.1. We assessed the effects of individual substitutions at positions 6 and 12 and found that, relative to Ub.wt, the substitution Asn^6 or His^{12} enhanced potency by ∼300-fold (IC_{50} = 200 nM) or ∼2-fold (IC_{50} = 24 μM), respectively. Taken together, these results show that the two mutations K6N/T12H are necessary and sufficient to confer high-potency inhibition of USP2.

Next, in the background of the double-mutant K6N/T12H, we tested the effects of substitutions that were abundant in the selected sequences. Substitutions at seven positions (Q2D, N6M, KllR, T14L, E64F, H68F, and V70N) either enhanced or did not affect inhibitor potency, confirming the predictions of the saturation scan analysis. However, substitutions at two positions (R42G and L73W) reduced inhibitor potency, and we speculate that these substitutions might have been selected because of favorable effects on protein stability or levels of protein display on phage, factors that were not corrected for in the analysis. Taken together, these results generally confirm the accuracy of predictions derived from the saturation scan data, but the method was inaccurate for two of nine positions at which substitutions to the Ubv.2.1 sequence were predicted to enhance binding to USP2.

To visualize the saturation scanning data in a structural context, we assigned the positions on the basis of bits scores into high-, moderate-, or low-conservation groups and mapped the data onto the structure of Ubv.2.1 (Fig. 1*B*). The high conservation positions form a large, contiguous functional epitope, including the C-terminal tail region and a band of residues across the center of the interface. Strikingly, the functional epitope of Ubv.21.4 (Fig. 1*C*) is very similar to that of Ubv.2.1, both exhibiting high sequence conservation at nine positions across the central band (4, 6, 10, 12, 42, 44, 49, 66, and 68), which we henceforth refer to as the core functional epitope.

To further validate and understand the saturation scanning data, we used an in silico saturation mutagenesis protocol (20, 21) to predict the tolerance of each Ub position to mutation. In this protocol, we substituted all Ub residues randomized in this study with the 19 other amino acids and calculated changes in free energy of Ub binding ($\Delta\Delta G_{bind}$) to USP2 or USP21 (Fig. S4*B*). The positions were then classified according to the number of substitutions that produced negative $\Delta\Delta G_{bind}$ values. Our computational results largely confirm the conservation patterns observed by saturation scanning, producing a similar Ub functional epitope for binding to USP2 or to USP21 (Fig. S4*A*). A few Ub positions exhibited slightly higher variability in the computational analysis relative to the empirical analysis, which may be due to the inability to model certain experimental phage selection pressures, such as higher stability and expression.

The Ub-Binding Site of USPs. We next examined the interactions of Ubv.2.1 and Ubv.21.4 with their cognate USPs. We examined the structure of the USP2:Ubv.2.1 complex and defined the Ub-binding site on USP2 as 52 residues that are within 4.5 Å of Ubv.2.1 (Fig. 2). In a manner analogous to the analysis of the saturation scan data for the Ubvs, an alignment of these 52 positions across 48 human USPs (Fig. S5) was used to create a PWM that was visualized as a sequence logo (Fig. 2*C* and Fig. S6), and the data were used to calculate bits scores to quantify sequence conservation within the human USP family. Based on the same cutoffs used for the saturation scan analysis of the Ubvs, the Ub-binding site contains 18 residues exhibiting high conservation and another 12 residues exhibiting moderate conservation. The sequence conservation data were visualized by mapping the bits scores onto the USP2 structure (Fig. 2*A*).

Viewed in a structural context, the conserved residues form two sites that interact with Ubv.2.1. Eight highly conserved residues and one moderately conserved residue form the active site cleft (site 1) that interacts with the C-terminal tail of Ubv.2.1

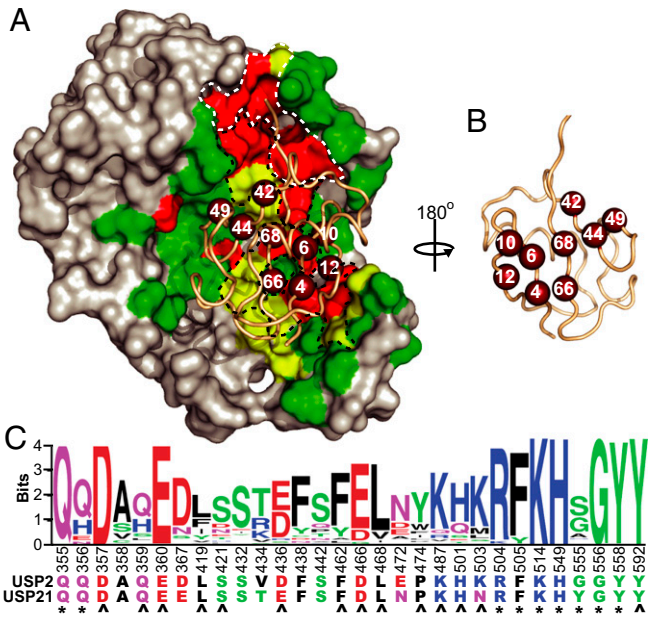

Fig. 2. Sequence conservation across the Ub-binding sites of human USPs. (*A*) The structure of USP2 in complex with Ubv.2.1 (PDB ID code 3V6C). USP2 is shown as a gray surface with residues in the Ub-binding site (residues within 4.5 Å of any residue on Ubv.2.1) colored according to bits scores derived from the alignment of 48 human USPs, as follows: green ≤ 2, 2 < yellow < 2.7, red ≥ 2.7. The main chain of Ubv.2.1 is shown as a wheat tube, and nine residues that show high conservation in the saturation scan of both Ubv.2.1 and Ubv.21.4 (core functional epitope) are shown as ruby spheres. White and black dashed lines demarcate site 1 and site 2, respectively. (*B*) Ubv.2.1 rotated 180° relative to *A*. (*C*) The sequence logo for highly conserved (bits ≥ 2.7) and moderately conserved positions (2 < bits < 2.7) within the Ub-binding sites of human USPs. Positions are numbered according to the sequence of USP2, and sequences of USP2 and USP21 are shown. Asterisks and arrowheads indicate residues that interact with the Ub C-terminal tail and the core functional epitope, respectively.

(residues 70–74). Ten of the remaining highly conserved residues and the remaining 11 moderately conserved residues form a second site (site 2) adjacent to site 1 (Fig. 2*B*). Site 2 interacts with the flat β-sheet face of Ubv.2.1, and notably, 14 of the residues in site 2 interact with the nine residues that form the core functional epitope shared by Ubv.2.1 and Ubv.21.4.

We speculated that the USP2 site 2 revealed to be important for binding to Ubv.2.1 also may be involved in the recognition of native Ub during catalysis. We explored this possibility using alanine-scanning mutagenesis (34) by constructing a panel of USP2 variants containing single-site alanine substitutions at each of the 14 conserved site 2 residues that make contact with the core functional epitope of Ubv.2.1 and then measuring their effects on the catalytic efficiency for the hydrolysis of Ub-AMC substrate. These assays showed that 10 of the 14 alanine substitutions reduced catalytic efficiency by at least twofold relative to WT USP2 (Fig. 3 and Table S1), thus verifying that USP2 uses site 2 for the recognition of Ub substrates.

Interactions at the USP:Ubv Interface. Saturation scanning enabled elucidation of the core functional epitope of the two Ubvs for binding to their cognate USPs, and examination of the USP:Ubv complex structures identified the USP residues interacting with this epitope. Notably, the core functional epitopes of Ubv.2.1 and Ubv.21.4 differ from the sequence of Ub.wt at only two positions and one position, respectively, suggesting that each Ubv likely interacts with its cognate USP in a manner very similar to that of Ub.wt. Indeed, comparisons of the USP2:Ubv.2.1 and USP2:Ub.wt

BIOCHEMISTRY

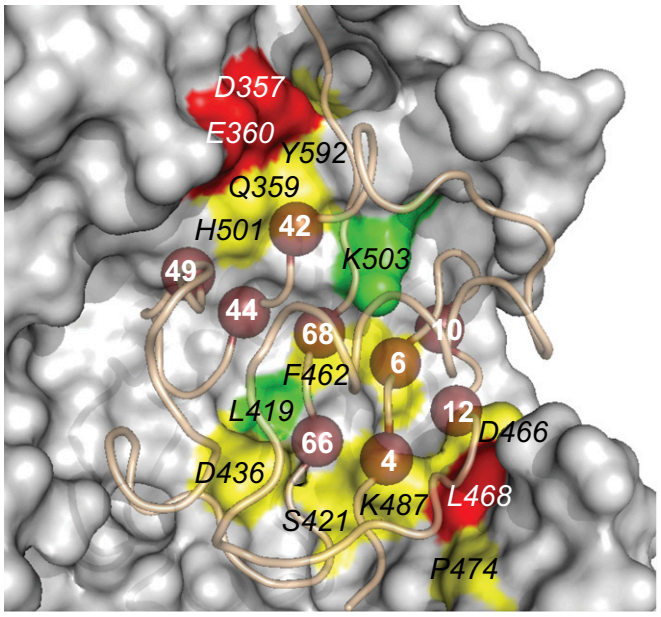

Fig. 3. Alanine-scanning of USP2 site 2. The structure of USP2 bound to Ubv.2.1 (PDB ID code 3V6C) is shown. USP2 is shown as a surface colored according to the fold reduction in catalytic efficiency relative to WT caused by alanine substitutions at the labeled positions, as follows: green < 2; 2 ≤ yellow < 10; red > 10; gray, unscanned. The Ubv02.1 main chain is shown as a tube, and residues in the core functional epitope are shown as numbered spheres.

complex structures (Fig. 4*A*) and of the USP21:Ubv.21.4 and USP21:Ub.wt complex structures (Fig. 4*B*) revealed excellent superposition of the Cα atoms, with an rmsd <0.28 Å and <0.32 Å, respectively. Moreover, there was excellent superposition of the six and seven side chains of the core functional epitope that are identical between Ub.wt and Ubv.2.1 and between Ub.wt and Ubv.21.4, respectively (rmsd <0.13 Å and <0.20 Å, respectively) and for the 19 USP2 side chains and the 18 USP21 side chains that make contact with any of the eight side chains of the Ubv.2.1 and Ubv.21.4 core functional epitope (rmsd <0.18 Å and <0.36 Å, respectively).

We next compared the USP2:Ubv.2.1 and USP21:Ubv.21.4 complex structures to understand how the core functional epitopes on the Ubvs interact with USPs (Fig. 5*A*). Superposition of the two complexes showed the overall structures to be highly similar (rmsd <0.55 Å for Cα atoms). Notably, six of nine residues in the core functional epitopes of the Ubvs are identical, and 12 of 19 USP residues that contact this epitope are identical. Thus, in terms of both the overall fold and side chain interactions, the two Ubvs bind to their cognate USPs in a very similar manner.

The USP fold comprises three subdomains, likened to a palm, thumb, and fingers, with the catalytic site located between the palm and the thumb (35). The eight side chains in the Ubv core functional epitope form two patches, one (patch A) that docks onto the palm and one (patch B) that docks onto the fingers (Fig. 5*A*). Patch A contains three residues (Arg42, Ile44, and Gln49), all of which are conserved as the Ub.wt sequence in both Ubvs (Fig. 5*B*). Patch B contains five residues, which can in turn be grouped into two clusters, with one cluster containing two residues (Phe4 and Thr66) that are conserved as the Ub.wt sequence in both Ubvs (Fig. 5*D*) and the other cluster containing three positions (6, 12, and 68) that differ between the Ubvs and show a preference for particular non-WT sequences in the saturation scans (Fig. 5*C*).

Each patch is anchored by a highly conserved hydrophobic residue: Ile44 in patch A and Phe4 in patch B. In patch A of Ubv.2.1, the hydrophobic side chain of $Ile44^{2.1}$ resides in a predominantly hydrophilic environment formed by the side chains of $Gln359^{USP2}$ and $Arg363^{USP2}$, which are either identical (position 359) or conservatively substituted (position 363) in USP21 (Fig. 5*B*). The other two residues in patch A that are conserved as the Ub.wt sequence (Arg42 and Gln49) also reside in a hydrophilic environment that is conserved in the two USPs. In particular, the $Arg42^{2.1}$ side chain forms hydrogen bonds with the side chains of $Gln359^{USP2}$ and $Gln49^{2.1}$ and makes favorable electrostatic interactions with the side chain of $Asp357^{USP2}$, and the $Gln49^{2.1}$ side chain also forms a hydrogen bond with the $Glu360^{USP2}$ side chain. Notably, all of these side chains are identical in USP2 and USP21 and are highly conserved within the USP family (Fig. 2*A* and Fig. S5), and, consequently, the interactions between Ubv.21.4 and USP21 are virtually identical in this region of the interface. In patch B of Ubv.2.1, the $Phe4^{2.1}$ side chain interacts with $Pro474^{USP2}$, the side chains of $Leu423^{USP2}$ and $Leu468^{USP2}$, and the aliphatic portion of the $Lys487^{USP2}$ side chain. The amino group of the $Lys487^{USP2}$ side chain makes a hydrogen bond with the side chain of $Thr66^{2.1}$, the second residue in patch B that is conserved as the Ub.wt sequence (Fig. 5*D*). In addition, the aliphatic portion of the $Thr66^{2.1}$ side chain packs against the side chains of $Asp436^{USP2}$ and $Phe489^{USP2}$. Notably, aside from positions 436 and 489, which have conservative substitutions, these residues are identical and have very similar interactions in the USP21:Ubv.21.4 complex structure.

Between the conserved hydrophobic anchor residues (Ile44 and Phe4) and the associated conserved hydrophilic residues (Arg42, Gln49, and Thr66) resides a cluster of three residues (6, 12, and 68) that differ between Ubv.2.1 and Ubv.21.4 and tend to vary from the sequence of Ub.wt (Fig. 5*C*). At position 68, Ubv.2.1 contains a WT His residue, whereas Ubv.21.4 contains a

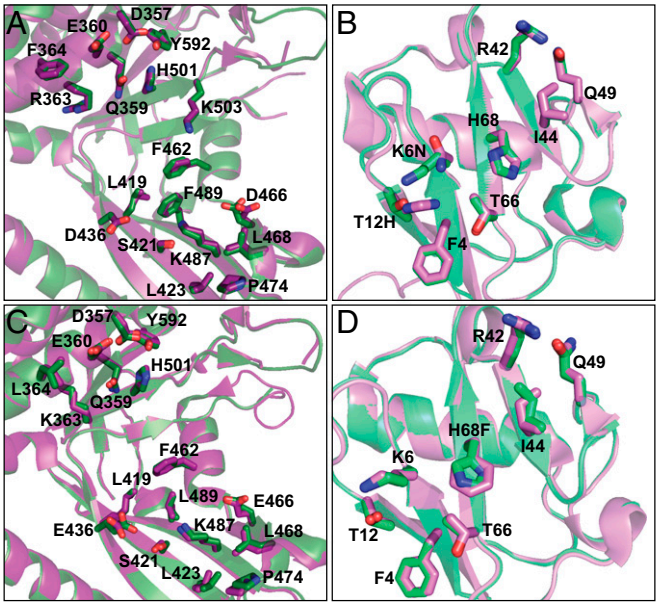

Fig. 4. Superposition of USPs in complex with Ub.wt or Ubvs. Shown are ribbon representations of superposed USP2 from USP2:Ub.wt (PDB ID code 2HD5) and USP2 from USP2:Ubv2.1 (PDB ID code 3V6C) (*A*), Ub.wt from USP2:Ub.wt (PDB ID code 2HD5) and Ubv2.1 from USP2:Ubv2.1 (PDB ID code 3V6C) (*B*), USP21 from USP21:Ub.wt (PDB ID code 3I3T) and USP21 from USP21:Ubv21.4 (PDB ID code 3MTN) (*C*), and Ub.wt from USP21:Ub.wt (PDB ID code 3I3T) and Ubv21.4 from USP21:Ubv21.4 (PDB ID code 3MTN) (*D*). USP and Ub.wt in complex are colored green and light green, respectively. USP and Ubv in complex are colored purple and pink, respectively. Side chains are shown as sticks for the Ubv core functional epitope and for USP side chains that contact the core functional epitope (USP side chains that are within 4.5 Å of side chains in the Ubv core functional epitope). For residues that differ between Ub.wt and Ubv, the Ubv sequence is shown after the residue number. Both USP2 and USP21 are numbered according to the sequence of USP2, based on a sequence alignment (Fig. S5).

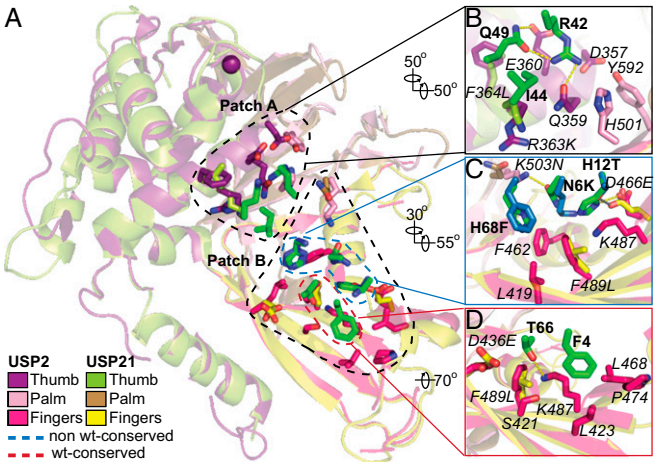

Fig. 5. Interactions between the Ubv core functional epitopes and USPs. (*A*) Superposition of USP2 (purple, pink, and magenta) and USP21 (lime, sand, and yellow) in complex with Ubv.2.1 (green) and Ubv.21.4 (blue), respectively. The USP main chains are shown as ribbons, the catalytic Cys is shown as a purple sphere, and side chains are shown as sticks for residues that contact the Ubv core functional epitope (within 4.5 Å). For clarity, the Ubv main chains are not shown, and only the side chains of the core functional epitopes are shown. (*B–D*) Interactions between USP side chains and Ubv side chains of patch A residues 42, 44, and 49 (*B*); non-WT patch B residues 6, 12, and 68 (*C*); and WT patch B residues 4 and 66 (*D*). USP residues are labeled in italics and numbered according to the sequence of USP2. For residues that differ between USP2 and USP21, the USP21 sequence is shown after the residue number. Ubv residues are labeled in bold, and for residues that differ between Ubv.2.1 and Ubv.21.4, the Ubv.21.4 sequence is shown after the residue number.

Phe substitution, but the saturation scan shows that Ubv.2.1 also prefers a Phe at this position (Fig. 1*D*). Consistent with a preference for a hydrophobic Phe, the side chains at position 68 of both Ubvs interact with three hydrophobic residues in their cognate USPs that are either identical (positions 419 and 462) or conserved (position 489). At position 6 of Ubv.2.1, the WT Lys residue is replaced by an Asn, and Met is also highly prevalent in the saturation scan dataset (Fig. 1*D*). The $Asn6^{2.1}$ side chain forms a hydrogen bond with the side chain of $Lys503^{USP2}$. The side chain at position 6 is close to the hydrophobic side chains of $Phe462^{USP2}$ and $Phe489^{USP2}$, which likely explains why a hydrophobic Met can effectively substitute for Asn at this position (Table 1).

Notably, although Ubv.21.4 contains a WT Lys at position 6, the saturation scan also shows a preference for Asn, which could be explained by a putative hydrogen bond with $Gln503^{USP21}$. At position 12 of Ubv.2.1, His dominates over the WT Thr and all other substitutions in the saturation scan (Fig. 1*D*), suggesting that this side chain contributes to recognition of USP2. This preference can be explained by a hydrogen bond between the $His12^{2.1}$ side chain and the $Asp466^{USP2}$ side chain. In contrast, in the saturation scan data for Ubv.21.4, His is completely absent at position 12, which instead prefers the WT Thr and also tolerates Ser and Tyr. In this case, the structure shows that a longer $Glu466^{USP21}$ side chain makes a hydrogen bond with $Thr12^{21.4}$, and it is likely that a larger His residue at position 12 cannot make the same favorable interactions.

Discussion

By applying phage display and saturation scanning analyses, we have obtained detailed views of how Ubv.2.1 and Ubv.21.4 bind with high affinity to their respective targets USP2 and USP21. Moreover, the availability of structures of each USP in complex with either its cognate Ubv or Ub.wt enabled us to extend our insights to the interactions between USPs and native Ub. Our structural comparison shows remarkable similarity at both the main chain and side chain levels between key residues on both sides of the interface (Fig. 3), suggesting that our conclusions regarding the Ubvs are likely applicable to native Ub as well.

Overall, the saturation scans reveal that a large number of residues in each Ubv exhibit significant sequence conservation, and thus a large surface area at the binding interface is recruited for productive binding contacts (Fig. 1). Notably, nine positions exhibit high conservation in both Ubvs, and this core functional epitope is involved in interactions with many other Ub-binding proteins (12, 36). Moreover, six of the nine positions (4, 6, 10, 42, 44, and 68) have been shown to be highly intolerant to substitutions in a study that probed the effects of point mutations on the biological function of Ub in yeast (36).

Interestingly, the core functional epitope docks on a region of the USP fold that is conserved within the family and is distinct from the active site (site 2; Fig. 2). Six of the nine residues within the core functional epitope are conserved as the WT sequence and interact with USP residues that are also highly conserved within the USP family, suggesting that these interactions may make similar energetic contributions to the recognition of Ub substrates by many USPs (Fig. 4). The importance of site 2 for the recognition of Ub substrates was confirmed by alanine-scanning mutagenesis of USP2, which showed that alanine substitutions at many positions in site 2 reduced the catalytic efficiency for hydrolysis of Ub-AMC (Fig. 3). The remaining three residues are also highly conserved, but favor non-WT sequences. Notably, these three residues cluster together (Fig. 5*C*) and include residues that were mutated in the Ubvs to gain high affinity and specificity for their cognate USPs, and thus they represent a critical cluster that can be exploited for specific inhibitor design.

Taken together, our results show that site 2 is important for the recognition of Ubvs and Ub substrates, and thus this region may be an attractive alternative target for the design of small molecules that inhibit USP function with higher specificity than has been possible with inhibitors targeting the highly conserved active site (37–39). Indeed, site 2 on the USPs and the core functional epitope on the Ubvs conform to the hot spot model of protein–protein interactions, in which a subset of side chains on each side of an interface form contiguous patches that interact with one another and contribute a large fraction of the binding energy (16, 19, 26). Although inhibition of protein–protein interactions with small molecules remains challenging, significant progress has been made in targeting sites other than classical active site clefts (40–42).

Our phage-displayed technology for developing high-affinity Ubvs has proven successful not only against USPs, but also against other DUB families, E3 ligases, E2- conjugating enzymes, and Ub-binding domains (14). Indeed, we have recently reported structures of Ubvs that target E3 ligases of the HECT (24), SCF (25), and RING families (43). Thus, the methods described here can be applied to dissect the functional epitopes for Ubvs targeting many other components of the Ub proteasome system. These studies should prove useful for guiding new inhibitor design strategies, and also for understanding the molecular basis for the myriad protein interactions that mediate Ub biology.

Materials and Methods

Library Construction and Analysis. Each Ubv was displayed on M13 bacteriophage, as described previously (14). Phage-displayed Ubv libraries were constructed, sorted, and analyzed as described previously (17). In brief, a Ubv-displaying phagemid vector containing TAA stop codons in the regions to be mutagenized was used as a template for oligonucleotide-directed mutagenesis (44). Mutagenic oligonucleotides were designed to replace positions to be scanned with degenerate NNK codons (n = A/G/C/T; K = G/C) that collectively code for all 20 natural amino acids. Five libraries were designed for each Ubv, with each combining three mutagenic oligonucleotides to introduce the mutations into the Ubv sequence (Table S2). Each mutagenesis reaction was electroporated separately into *Escherichia coli* SS320, and each yielded a library of $>10^9$ unique members.

The libraries were handled separately, and each was sorted for four rounds to select for binding to immobilized USP. Individual clones from the fourth round of selection were grown in a 96-well format, and the culture supernatants were used directly in phage ELISAs (45) to detect clonal Ubv phage bound to USP. Between 50 and 100 positive binding clones from each library were subjected to DNA sequencing and statistical analysis.

Statistical Analysis. The binding sequences from each library were aligned. The occurrence of each amino acid at each scanned position was corrected for bias by dividing the counts by the number of codons for that amino acid contained within the NNK degenerate codon (46). Normalized sequences were used to produce an alignment in LOGOS consensus format (weblogo.berkeley.edu/logo.cgi) (27). Bits scores at position *i* were calculated according to the formula $R_i = \log_2 20 - (H_i - e_n)$, where H_i is Shannon entropy at position *i*, $= -\Sigma f_{a,i} \times \log_2 f_{a,i}$, and $f_{a,i}$ is the relative frequency of amino acid *a* at position *i*. Small sample correction, e_n, was assigned as 0, so that bits scores ranged from 0 to 4.32 ($\log_2 20$). The statistical analysis of the Ub-binding site of USPs was performed similarly, using an alignment of the catalytic domains of 48 human USPs (Fig. S5).

Enzyme activity assays were performed in assay buffer (50 mM Hepes pH 7.5, 0.01% Tween 20, and 10 mM DTT). Serial dilutions were performed in 96-well plates and then transferred to 384-well plates for measurements. USP2 inhibition assays were performed by measuring the release of fluorogenic amido-4-methylcoumarin (AMC) from the substrate, Ub-AMC (Boston Biochem), as described previously (14), with 1 μM Ub-AMC substrate, 7.5 nM USP2, and serial dilutions of Ubvs. USP2 and Ubv were mixed in assay buffer and incubated at room temperature for 2 min before the addition of Ub-AMC. Proteolytic activity was measured by monitoring the increase of fluorescence emission at 460 nm (excitation at 360 nm) for 30 min using a Synergy2 plate reader (BioTek Instruments). IC_{50}, defined as the concentration of Ubv that inhibits 50% of USP2 activity, was fitted using the sigmoidal 4PL equation in GraphPad Prism software.

For determination of kinetic parameters, a fixed concentration of USP2 was incubated with serial dilutions of Ub-AMC. Duplicates of initial reaction velocity (nM/s) were determined at each substrate concentration by determining the linear slope from plotting fluorescence signal vs. time and converting to molarity by interpolating from a standard curve of known AMC concentrations. Velocity vs. substrate concentration was plotted to determine the K_M and V_{max} values using GraphPad Prism with the Michaelis–Menten equation, $V_0 = V_{max}*[S]/(K_M + [S])$. k_{cat} was obtained from the equation $k_{cat} = V_{max}/[E]_o$, where $[E]_o$ is the total enzyme concentration. For velocity vs. substrate plots that could not be fitted using the Michaelis–Menten equation but could be fitted using a linear line, the k_{cat}/K_M value was estimated by dividing the linear slope ($nMs^{-1}M^{-1}$) by $[E]_o$.

Computational saturation mutagenesis of Ubv.2.1 or Ubv.21.4 in complex with USP2 or USP21 (PDB ID codes 3V6E and 3MTN) was performed as described previously (21). Each substitution was introduced into the Ubv, and the intermolecular energy, ΔE_{inter}, was calculated by subtracting the energies of the unbound chains from that of the complex. $\Delta\Delta G_{bind}$ was calculated as the difference between the ΔE_{inter} values of the Ubv and Ub.wt complexes. The backrub algorithm was used to model backbone flexibility, including movements of −15, −10, −5, +5, +10, and +15 degrees relative to the original conformation. The lowest $\Delta\Delta G_{bind}$ value from all backrub moves was assigned to each substitution, and a substitution was considered allowed if it exhibited a negative $\Delta\Delta G_{bind}$ value. The position was categorized as high, medium, or low conservation if <4, 4–7, or >7 substitutions, respectively, were allowed.

ACKNOWLEDGMENTS. We thank J. Teyra, M. Ben-David, M. Gorelik, R. Hanna, and J. Tykvart for helpful discussions. This work was supported by a Canadian Institutes of Health Research operating grant (MOP-136956, to S.S.S.). I.L. was supported by the Canadian Cancer Society (Grant 702861). J.M.S. was supported by the Israel Science Foundation (Grant 1873/15).

1. Haglund K, et al. (2003) Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* 5(5):461–466.
2. Terrell J, Shih S, Dunn R, Hicke L (1998) A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol Cell* 1(2):193–202.
3. Ulrich HD, Walden H (2010) Ubiquitin signalling in DNA replication and repair. *Nat Rev Mol Cell Biol* 11(7):479–489.
4. Jin L, Williamson A, Banerjee S, Philipp I, Rape M (2008) Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* 133(4):653–665.
5. Spence J, et al. (2000) Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* 102(1):67–76.
6. Schulman BA, Harper JW (2009) Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* 10(5):319–331.
7. Ye Y, Rape M (2009) Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol* 10(11):755–764.
8. Berndsen CE, Wolberger C (2014) New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* 21(4):301–307.
9. Komander D, Clague MJ, Urbé S (2009) Breaking the chains: Structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* 10(8):550–563.
10. Husnjak K, Dikic I (2012) Ubiquitin-binding proteins: Decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* 81:291–322.
11. Hurley JH, Lee S, Prag G (2006) Ubiquitin-binding domains. *Biochem J* 399(3):361–372.
12. Harrison JS, Jacobs TM, Houlihan K, Van Doorslaer K, Kuhlman B (2016) UbSRD: The Ubiquitin Structural Relational Database. *J Mol Biol* 428(4):679–687.
13. Lange OF, et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471–1475.
14. Ernst A, et al. (2013) A strategy for modulation of enzymes in the ubiquitin system. *Science* 339(6119):590–595.
15. Chen J, Sawyer N, Regan L (2013) Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Prot Sci* 22(4):510–515.
16. Wells JA (1996) Binding in the growth hormone receptor complex. *Proc Natl Acad Sci USA* 93(1):1–6.
17. Pál G, Kouadio JL, Artis DR, Kossiakoff AA, Sidhu SS (2006) Comprehensive and quantitative mapping of energy landscapes for protein–protein interactions by rapid combinatorial scanning. *J Biol Chem* 281(31):22378–22385.
18. Wells JA, de Vos AM (1996) Hematopoietic receptor complexes. *Annu Rev Biochem* 65:609–634.
19. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots: A review of the protein–protein interface determinant amino-acid residues. *Proteins* 68(4):803–812.
20. Aizner Y, et al. (2014) Mapping of the binding landscape for a picomolar protein–protein complex through computation and experiment. *Structure* 22(4):636–645.
21. Sharabi O, Erijman A, Shifman JM (2013) Computational methods for controlling binding specificity. *Methods Enzymol* 523:41–59.
22. Yang Y, Kitagaki J, Wang H, Hou DX, Perantoni AO (2009) Targeting the ubiquitin-proteasome system for cancer therapy. *Cancer Sci* 100(1):24–28.
23. Sippl W, Collura V, Colland F (2011) Ubiquitin-specific proteases as cancer drug targets. *Future Oncol* 7(5):619–632.
24. Zhang W, et al. (2016) System-wide modulation of HECT E3 ligases with selective ubiquitin variant probes. *Mol Cell* 62(1):121–136.
25. Gorelik M, et al. (2016) Inhibition of SCF ubiquitin ligases by engineered ubiquitin variants that target the Cul1 binding site on the Skp1-F-box interface. *Proc Natl Acad Sci USA* 113(13):3527–3532.
26. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267(5196):383–386.
27. Schneider TD, Stephens RM (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* 18(20):6097–6100.
28. D'haeseleer P (2006) What are DNA sequence motifs? *Nat Biotechnol* 24(4):423–425.
29. Badis G, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324(5935):1720–1723.
30. Ray D, et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499(7457):172–177.
31. Tonikian R, et al. (2009) Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. *PLoS Biol* 7(10):e1000218.
32. Gfeller D, et al. (2011) The multiple-specificity landscape of modular peptide recognition domains. *Mol Syst Biol* 7:484.
33. Andreatta M, Nielsen M (2012) Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using NNAlign. *Immunology* 136(3):306–311.
34. Morrison KL, Weiss GA (2001) Combinatorial alanine scanning. *Curr Opin Chem Biol* 5(3):302–307.
35. Hu M, et al. (2002) Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* 111(7):1041–1054.
36. Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN (2013) Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol* 425(8):1363–1377.
37. Pal A, Young MA, Donato NJ (2014) Emerging potential of therapeutic targeting of ubiquitin-specific proteases in the treatment of cancer. *Cancer Res* 74(18):4955–4966.
38. Colland F, et al. (2009) Small-molecule inhibitor of USP7/HAUSP ubiquitin protease stabilizes and activates p53 in cells. *Mol Cancer Ther* 8(8):2286–2295.
39. Colland F (2010) The therapeutic potential of deubiquitinating enzyme inhibitors. *Biochem Soc Trans* 38(Pt 1):137–143.
40. Arkin MR, Tang Y, Wells JA (2014) Small-molecule inhibitors of protein–protein interactions: Progressing toward the reality. *Chem Biol* 21(9):1102–1114.
41. Guo W, Wisniewski JA, Ji H (2014) Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett* 24(11):2546–2554.
42. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat Rev Drug Discov* 3(4):301–317.
43. Brown NG, et al. (2016) Dual architectures regulate multiubiquitination and ubiquitin chain elongation for a RING E3. *Cell* 165(6):1440–1453.
44. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA* 82(2):488–492.
45. Sidhu SS, Lowman HB, Cunningham BC, Wells JA (2000) Phage display for selection of novel binding peptides. *Methods Enzymol* 328:333–363.
46. Bond CJ, Wiesmann C, Marsters JC, Jr, Sidhu SS (2005) A structure-based database of antibody variable domain diversity. *J Mol Biol* 348(3):699–709.