Biochemical and Structural Analysis of the Interaction between the UBA(2) Domain of the DNA Repair Protein HHR23A and HIV-1 Vpr†

Elizabeth S. Withers-Ward,‡,⊥ Thomas D. Mueller,§,⊥ Irvin S. Y. Chen,*,‡ and Juli Feigon*,§

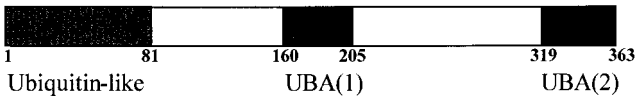
Departments of Microbiology, Immunology and Medicine, and UCLA AIDS Institute and Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095

Received July 24, 2000; Revised Manuscript Received September 27, 2000

ABSTRACT: The DNA repair protein HHR23A is a highly conserved protein that functions in nucleotide excision repair. HHR23A contains two ubiquitin associated domains (UBA) that are conserved in a number of proteins with diverse functions involved in ubiquitination, UV excision repair, and signaling pathways via protein kinases. The cellular binding partners of UBA domains remain unclear; however, we previously found that the HHR23A UBA(2) domain interacts specifically with the HIV-1 Vpr protein. Analysis of the low resolution solution structure of HHR23A UBA(2) revealed a hydrophobic loop region of the UBA(2) domain that we predicted was the interface for protein/protein interactions. Here we present results of in vitro binding studies that demonstrate the requirement of this hydrophobic loop region for interaction with human immunodeficiency virus (HIV-1) Vpr. A single point mutation of the Pro at residue 333 to a Glu totally abolishes the binding of HIV-1 Vpr to UBA(2). High resolution NMR structures of the binding deficient UBA(2) mutant P333E as well as of the wild-type UBA(2) domain were determined to compare the effect of this mutation on the structure. Small but significant differences are observed only locally at the site of the mutation. The biochemical and structural analysis confirms the function of the HHR23A UBA(2) GFP-loop as the protein/protein interacting domain.

HHR23A is one of the human homologues of the yeast Rad23 and Rhp23 proteins that functions in nucleotide excision repair (*1*−*3*). The biochemical function of HHR23A is poorly understood. In yeast, *rad23* mutants are sensitive to UV light (*4*). Rad23 promotes the assembly of a large multiprotein complex involved in nucleotide excision repair (*5*). In *Saccharomyces cerevisiae*, Rad23 is complexed with Rad4. Similarly, both of the human homologues termed HHR23A and HHR23B interact with the human ortholog, the Xeroderma pigmentosum group C (XPC) protein and promote its activity in in vitro DNA repair assays (*3*, *6*−*9*). In vitro studies have shown that the Rad23/Rad4 complex preferentially binds to damaged DNA (*10*, *11*). Studies in yeast indicate that Rad23 is required for both transcription-coupled DNA repair and efficient global genome repair (*12*).

All of the Rad23 homologues share a common domain structure, including a ubiquitin-like domain at the N-terminus and two copies of a highly conserved domain termed the ubiquitin associated domain (UBA) located in the middle [UBA(1)] and at the C-terminus [UBA(2)] of the protein (*6*) (Figure 1). The N-terminal ubiquitin-like (UBL) domain interacts with the 26S proteasome, as does that of HHR23A and B, suggesting a role for Rad23 in protein degradation via the ubiquitin−proteasome pathway (*13*−*15*). In yeast, the UBL domain is required for function (*2*). UBA domains are approximately 45 amino acids long and are found in many proteins including those involved in the ubiquitination pathway, UV excision repair, and cell signaling (*16*). The C-terminal UBA domain of HHR23A has very recently been shown to interact with the human DNA base excision repair protein 3-methyladenine DNA glycosylase (MPG) (*17*).



HHR23A

FIGURE 1: Schematic representation of the domain structure of HHR23A. The shaded boxes indicate the domains that have been identified by sequence analysis and include an N-terminal ubiquitin-like domain and two copies of highly conserved ubiquitin associated domain (UBA) which are designated UBA(1) and UBA(2).

We previously found that the human immunodeficiency virus (HIV-1) Vpr protein binds specifically to the UBA(2) domain of HHR23A (*18*). The Vpr protein is a 96-amino acid gene product of HIV-1 that is incorporated into virions (*19*). Vpr plays a role in several aspects of the viral life cycle including nuclear import of the preintegration complex, transactivation of transcription, induction of cell cycle arrest at the G2/M checkpoint, and subsequent apoptosis (*20*−*24*). These functions are mediated by distinct domains of Vpr. Residues within both the N- and C-termini are implicated in mediating cell cycle arrest (*25*, *26*). Selection for maintenance of wild-type Vpr occurs in chimpanzees and

† This work was supported by NIH Grant AI43190 to I.S.Y.C. and J.F.

* To whom correspondence should be addressed. E-mail: feigon@mbi.ucla.edu; phone: (310) 206-6922; fax: (310) 825-0982.

‡ Departments of Microbiology, Immunology and Medicine, and UCLA AIDS Institute.

§ Department of Chemistry and Biochemistry and the Molecular Biology Institute.

⊥ These authors contributed equally to this work.

10.1021/bi0017071 CCC: $19.00
Published on Web 10/27/2000

humans infected with HIV-1, supporting the idea that Vpr confers an important selective advantage to the virus (*27*).

Vpr-induced cell-cycle arrest and apoptosis of HIV-1 infected cells would have a profound effect on T-cells, the primary target of HIV-1, contributing to the immune deregulation associated with AIDS (*28*). HIV-1 infected $CD4^+$ T-cells would fail to undergo activation and clonal expansion. Vpr-mediated apoptosis would contribute to depletion of the $CD4^+$ cells that is seen in HIV-infected individuals. Therefore, Vpr may be an important viral protein to target in the development of therapeutic reagents for treatment of HIV-1-infected individuals. In addition to finding that Vpr interacts specifically with the UBA(2) domain of HHR23A, we previously demonstrated that overexpression of HHR23A or a fragment encompassing the UBA(2) domain would partially alleviate cell cycle arrest mediated by Vpr (*18*). This suggests a significant functional role for the interactions between Vpr and HHR23A. We hypothesized that Vpr may mediate cell cycle arrest and/or apoptosis by binding to the UBA(2) domain of HHR23A, thereby leading to deregulation of normal cellular processes involved in the control and signaling of DNA repair. Thus, an understanding of the structural requirements for interaction between Vpr and UBA(2) may not only provide insight into the role of HHR23A in Vpr function but may also lead to strategies for the development of therapeutic reagents that would interfere with Vpr function.

We have previously reported a low resolution structure of UBA(2) of HHR23A determined by homonuclear NMR spectroscopy (*29*). The structural motif of the UBA(2) domain is composed of three α-helices folded around a hydrophobic core. Analysis of the NMR structure revealed an unusually large hydrophobic area of the protein that we predicted to be an interface for protein/protein interactions and might be the Vpr-interacting region of the HHR23A UBA(2) domain. This hydrophobic area of the UBA(2) domain is in the vicinity of the GFP-loop ($_{331}$GFP$_{333}$) which forms a sharp turn between α-helices 1 and 2. To further characterize the amino acids that are critical for interaction with HHR23A UBA(2), synthetic peptides derived from the UBA domains of *S. cerevisiae* (*2*) and an unrelated protein c-cbl (*30*) that is involved in signal transduction in hematopoietic cells were assayed for binding to Vpr in vitro. The amino acid sequence of these UBA domains differs at and around the GFP-loop. The results indicate that the Pro within the GFP-loop of the HHR23A UBA(2) domain is required for interaction with Vpr. Further studies with mutant peptides demonstrate that the Pro within the GFP-loop is necessary but not sufficient for binding to Vpr. Furthermore, the HHR23A UBA(1) domain and the UBA domain from c-cbl do not interact with Vpr, suggesting that the interaction between the HHR23A UBA(2) domain and Vpr is a highly specific one. High-resolution structures of both the wild type and the P333E mutant of UBA(2) were determined by NMR to address the question of how this mutation affects the structure and/or hydrophobic surface of the protein. Comparison of wild-type and P333E mutant structures reveals that substitution of the Pro residue 333 with a Glu results in changes in the shape and charge of the hydrophobic patch predicted to be the Vpr-interacting domain. These changes can account for the loss of the ability to bind Vpr. The structural analysis of a Vpr-interacting protein and identification of a Vpr-interacting interface provide a rationale for the future development of therapeutic reagents that interfere with Vpr function.

MATERIALS AND METHODS

Expression of GST Fusion Proteins in Bacteria. The HIV-1 Vpr GST fusion protein was expressed and purified as described with some modifications (*18*). Briefly, the HIV-1_{NL4-3} coding sequence, inserted into the pGEX-KG expression plasmid, was expressed in *Escherichia coli* BL21 (DE3). Following disruption of the cells by sonication and centrifugation, the cells were dialyzed overnight at 4 °C against PBS. The dialysate was then applied directly to a prepacked glutathione sepharose column (Pharmacia Biotech) and purified according to the manufacturer's instructions. Fractions were collected and analyzed by SDS−polyacrylamide gel electrophoresis (PAGE). Proteins were visualized by staining the polyacrylamide gels with Coomassie Blue or by transferring the proteins to nitrocellulose and staining with Ponceau S. These assays confirmed that all fractions were free of contaminating bacterial proteins. Those fractions containing the bulk of the GSTVpr fusion protein were pooled and dialyzed overnight at 4 °C against PBS. Protein concentration was determined by the Bradford method. Purified proteins were stored at −70 °C.

Synthesis and Preparation of Peptides. Peptides corresponding to UBA domains were biochemically synthesized by Fmoc (fluorenylmethoxycarbonyl)/*tert*-butyl based solid-phase peptide chemistry method on an ABI 431A (Perkin-Elmer) peptide synthesizer utilizing a single coupling program to carry out the chain assembly. The crude peptide was purified to 95% homogeneity by preparative reverse-phase HPLC on a Vydac C18 column. The composition and purity of each peptide was assessed using electro-spray ionization (ESI) mass spectrometry and amino acid analysis. For detection purposes, all peptides were synthesized with a biotin molecule at their N-termini. The sequences of the peptides used in these studies are listed below with mutations indicated in boldface type.

HHR23A UBA(2): QEKEAIERLKALGFPESLVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(2)P333E: QEKEAIERLKALGF**E**ESLVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(2)$_{333}$GFP/ERD$_{335}$: QEKEAIERLKALGF**ERD**LVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(1): SEYETMLTEIMSMGFERERVVAALRASYNNPHRAVEYLLTGIPGS; *S. cerevisiae* UBA(2): EDDQAISRLCELGFERDLVIQVYFACDKNEEAAANILFSDHAD; *S. cerevisiae* UBA(2)E370P: EDDQAISRLCELGF**P**RDLVIQVYFACDKNEEAAANILFSDHAD; *S. cerevisiae* UBA(2)$_{370}$PES$_{372}$: EDDQAISRLCELGF**PES**LVIQVYFACDKNEEAAANILFSDHAD; UBA c-cbl: SPQLSSEIENLMSQGYSYQDIQKALVIAQNNIEMAKNILRE.

Samples for binding assays were prepared by dissolving the purified lyophilized peptide in 10 mM Hepes, pH 6.8 (1 mg/mL). In some cases, the pH was adjusted to 7.0 by the addition of 1 N NaOH. Aliquots were stored at −70 °C and thawed just prior to use.

In Vitro Binding Assay. This assay is a slightly modified version of a binding assay that we have previously developed and tested with all of the relevant controls (*18*). Glutathione Sepharose 4B (Pharmacia Biotech) was prepared by washing 0.5 mL of the suspension obtained from the manufacturer three times in prebinding buffer (100 mM NaCl, 25 mM Tris-

HCl [pH 7.5], 0.1% Triton X-100) (1 mL/wash) and then resuspending it in prebinding buffer in a total volume of 2 mL. GSTVpr or GST (10 μg/reaction) was added to 100 μL of the glutathione Sepharose suspension. After samples were incubated at room temperature (30−60 min), prebound complexes were washed once in prebinding buffer (1 mL/reaction) and then twice with buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM DTT). Peptides were diluted 4-fold in buffer A, and 50 μg was added to each reaction. After a 30-min incubation at room temperature, samples were washed three times in buffer B (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 1 mM DTT, 0.5% NP40) (1 mL/wash). Samples were then incubated in blocking buffer [1× TNT (20 mM Tris, 500 mM NaCl, pH 7.5, 0.3% Tween 20 (v/v), 4.0% nonfat dry milk (w/v), 1.0% BSA (w/v)] (0.2 mL/sample) for 30 min at room temperature, and then washed three times in buffer B. After a 30-min incubation in a streptavidin horseradish peroxidase conjugate (Amersham Life Science Inc.) [diluted 1/5000 in TNT that was 0.2% in BSA (w/v)], samples were washed three times in buffer B. 1-Step Turbo TMB-ELISA solution (Pierce Chemical Co.) was then added to each sample (0.25 mL/sample). After a 10-min incubation at room temperature, 150 μL of sulfuric acid (1.2 N) was added to each sample, and the absorbance was read at 450 nm in a spectrophotometer.

Preparation of the UBA(2) Domain. A DNA fragment encoding the UBA(2) domain (residues 319−363) was amplified by PCR from full-length HHR23A gene (*18*) and inserted into the pGEX-2T vector (Pharmacia) between the *Bam*HI and *Eco*RI restriction sites. *E. coli* strain BL21(DE3)-pLysS was then transformed with the plasmid, and the transformed bacteria were cultured in either LB or M9 media. To prepare ^{15}N- or ^{15}N- and ^{13}C-labeled protein, $^{15}NH_4Cl$ or [$^{13}C_6$]glucose was used as sole nitrogen or carbon source in the growth medium. Expression of GST-UBA(2) fusion protein was induced with 1 mM isopropyl-β-D-thiogalactopyranoside when the absorbance of the culture reached approximately 0.7 optical density units at 550 nm. The temperature was lowered from 37 to 30 °C during the expression to increase the yield of soluble fusion protein. The cells were harvested 3 h (4.5 h for the preparation of labeled protein in M9 medium) after induction and resuspended in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.375 M sucrose (15 mL/L of culture). Cells were lysed by adding 1 vol of BugBuster (Novagen), 1 mg/mL lysozyme (Sigma), and 0.5 U of Benzonase (Novagen). The GST fusion protein was cleaved using biotinylated thrombin (Novagen). The protease was then removed with streptavidin agarose, and the UBA(2) protein was finally purified by an anion-exchange chromatography step (high performance Q sepharose, Pharmacia). The final purified UBA(2) protein consisted of residues 319−363 of the HHR23A protein and an N-terminal Gly-Ser peptide derived from the expression vector. Typically, between 5 (^{15}N- and ^{13}C-labeled) and 12.5 mg (^{15}N-labeled) of purified UBA(2) protein was obtained from 1 L of M9 medium. Initial NMR studies for the wild-type protein as well as the NMR analysis of the UBA(2) mutant P333E were performed using peptides derived from chemical synthesis as previously described (*29*).

NMR Analysis and Structure Calculation of the Wild-Type UBA(2) Domain and the Mutant P333E. NMR samples of wild-type UBA(2) contained between 1.5 and 3 mM protein in 50 mM sodium phosphate, pH 6.5, 100 mM sodium chloride, 2 mM dithiothreitol-d_8 (DTT), 0.02% sodium azide, and 5% ^2H_2O. All spectra were acquired at 300 K with a Bruker DRX 500 or DRX 600 spectrometer. A series of three-dimensional experiments (^{15}N-HSQC-NOE-HSQC, ^{15}N-HSQC TOCSY, CBCA(CO)NH, CBCANH, HBHA(CO)NH, HBHANH, HCCH-COSY, and HCCH-TOCSY) were acquired for the assignments of the ^{1}H, ^{15}N, and ^{13}C resonances (*31*). All NMR data was processed using XWINNMR (Bruker) and analyzed with the software XEASY (*32*). Distance information was obtained from ^{15}N- or ^{13}C-edited 3D NOESY experiments as well as a series of 2D NOESY experiments with mixing times ranging from 80 to 300 ms. NOE connectivities obtained from 3D NOESY experiments were categorized into strong, medium, weak, and very weak, corresponding to upper limits for proton−proton distances of 3.0, 4.0, 4.5, and 5.0 Å, respectively. 2D NOESY spectra were analyzed and integrated using the software packages XEASY (*32*) and SPSCAN (*33*). No stereospecific assignments of methylene protons or methyl groups were used. Upper limit distance restraints were then generated from the NOE volumes using DYANA (*34*). In total, 779 NOE-derived distance restraints (173 intraresidue, 196 sequential, 219 medium range, and 191 long range) were used for the wild-type UBA(2). In the final stages of refinement, 17 hydrogen bonds were introduced based on a regular pattern of NOEs of the type $H_{\alpha i}H_{\beta i+3}$ and $H_{\alpha i}NH_{i+3}$, as well as carbon and proton chemical shift information suggesting a regular α-helix.

The proton resonances of the UBA(2)P333E were assigned by analysis of a series of 2D ^{1}H NMR experiments composed of TOCSY, DQF-COSY, and NOESY. NOE connectivities were derived from a series of 2D NOESY experiments acquired with mixing times ranging from 100 to 350 ms. Distance restraints were generated as described above. A total of 826 NOE-derived distance restraints (204 intraresidue, 182 sequential, 226 medium range, and 214 long range) were used for the structure calculation of UBA(2)P333E. Although almost all NOE-derived distance restraints were derived from the 2D NOESY data acquired on the mutant sample, an additional 148 distance restraints were taken from the wild-type UBA(2) 3D data set due to spectral overlap in the 2D NOESY spectra of the mutant. All of these were located within the hydrophobic core and were far from the site of mutation. Identical proton chemical shifts between the wild-type and the mutant protein indicate that the structural environment for those protons is the same, and use of those restraints does not bias the structure of the mutant protein.

The structures of wild-type UBA(2) and its mutant P333E were calculated from an extended peptide conformation with a simulated annealing protocol using XPLOR (*35*). The simulated annealing protocol consisted of 35 ps dynamics at 2000 K (integration time steps of 5 fs) and 35 ps slow cooling (10 K/cycle). Final minimization was performed using 50 steps of Adopted Raphson Newton algorithm (Quanta98, MSI). For each protein, 100 structures were calculated, of which 18 (wild-type) and 21 (mutant P333E) were selected based on lowest total and NOE energy.

Coordinates for the 18 lowest energy structures of UBA(2) (accession code 1DV0) and UBA(2)P333E (accession code 1F4I) have been deposited in the RCSB protein data bank.

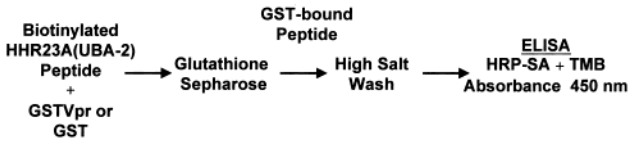


FIGURE 2: Schematic of the in vitro binding assay (*18*). Binding reactions were performed in the following manner. Ten micrograms of GSTVpr or GST bound to glutathione sepharose was mixed with 50 μg of biotinylated UBA(2) peptide and incubated at room temperature for 1 h. GST-containing complexes were selectively recovered and incubated with streptavidin conjugated to horseradish peroxidase. The presence of bound biotinylated peptide was determined by reading the absorbance at 450 nm after the addition of tetramethyl benzidine substrate solution (1-Step Turbo TMB-ELISA, Pierce Chemical Co.) and sulfuric acid.

RESULTS AND DISCUSSION

Binding of Vpr to UBA Domains Is Specific to the UBA-(2) of HHR23A. Our previous studies showed that Vpr binds specifically to UBA(2) of HHR23A and HHR23B. Since the UBA domain is a common structural motif, the binding to Vpr of a variety of other UBA domains was tested using a binding assay based on one that we have previously described (*18*) (Figure 2). Binding of GSTVpr-associated proteins was detected by ELISA. The amino acid sequences and alignment of the peptides that were tested are shown in Figure 3, panel A. The results of the in vitro binding assays using peptides derived from the internal UBA domain of HHR23A [UBA-(1)], *S. cerevisiae* Rad23A UBA(2), and the UBA domain of an unrelated protein, c-cbl (*30*), which is involved in signal transduction in hematopoeitic cells, are shown in Figure 3, panel B. None of these peptides show binding to GSTVpr in this assay. The *S. cerevisiae* Rad23 UBA(2) domain contains several hydrophilic residues in the vicinity of the GFP-loop. Western blot analysis (*29*) also did not show any detectable binding of *S. cerevisiae* Rad23 UBA(2) to GSTVpr (unpublished results). The absence of binding to the UBA domains of c-cbl, *S. cerevisiae* Rad23, as well as the UBA(1) of HHR23A indicates that the interaction between the HHR23A UBA(2) domain and Vpr is a highly specific one and not simply the result of the promiscuous interaction of Vpr with a domain found within many proteins with diverse cellular functions. Furthermore, these results support the prediction that the presence of hydrophilic residues at and around the region of the GFP-loop might preclude Vpr binding and lend support to the proposal that the GFP-loop is the interface for Vpr interaction.

Identification of Critical Residues on HHR23A for Binding to Vpr. As a next step in characterizing the amino acids that are critical for interaction with HHR23A UBA(2), we performed in vitro binding studies with synthetic mutant peptides derived from the UBA(2) domains of *S. cerevisiae* and HHR23A. A comparison of the amino acid sequence of the UBA domains in the vicinity of the GFP-loop indicated the presence of several charged residues in the *S. cerevisiae* UBA(2) domain at and around the GFP-loop that might contribute to the loss of binding to GSTVpr. In particular,

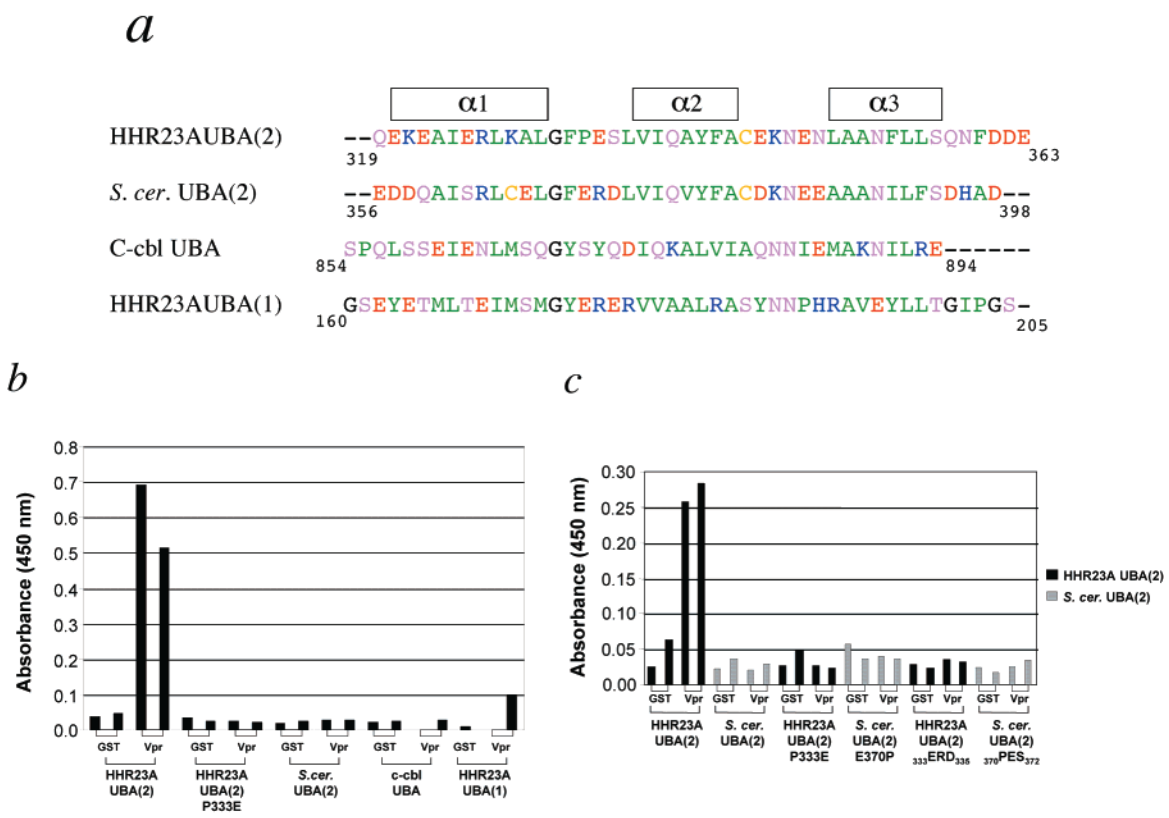


FIGURE 3: (A) Amino acid sequences and alignment of UBA domains: HHR23A UBA(2), *S. cerevisiae* Rad23A UBA(2), c-cbl UBA, and HHR23A UBA(1). The amino acids that comprise α-helices 1, 2, and 3 are indicated above the sequence. The sequence of amino acids are colored green (hydrophobic), red (negatively charged), orange (polar uncharged), and blue (positively charged). (B) Results of binding assays with HHR23A UBA(2), HHR23A UBA(2) P333E, *S. cerevisiae* Rad23A UBA(2), c-cbl UBA, and HHR23A UBA(1). (C) Results of binding assays with HHR23A and *S. cerevisiae* Rad23 UBA(2) mutant peptides. Binding reactions were performed as described in the Materials and Methods section. Data shown are values obtained from duplicate samples from a typical experiment.

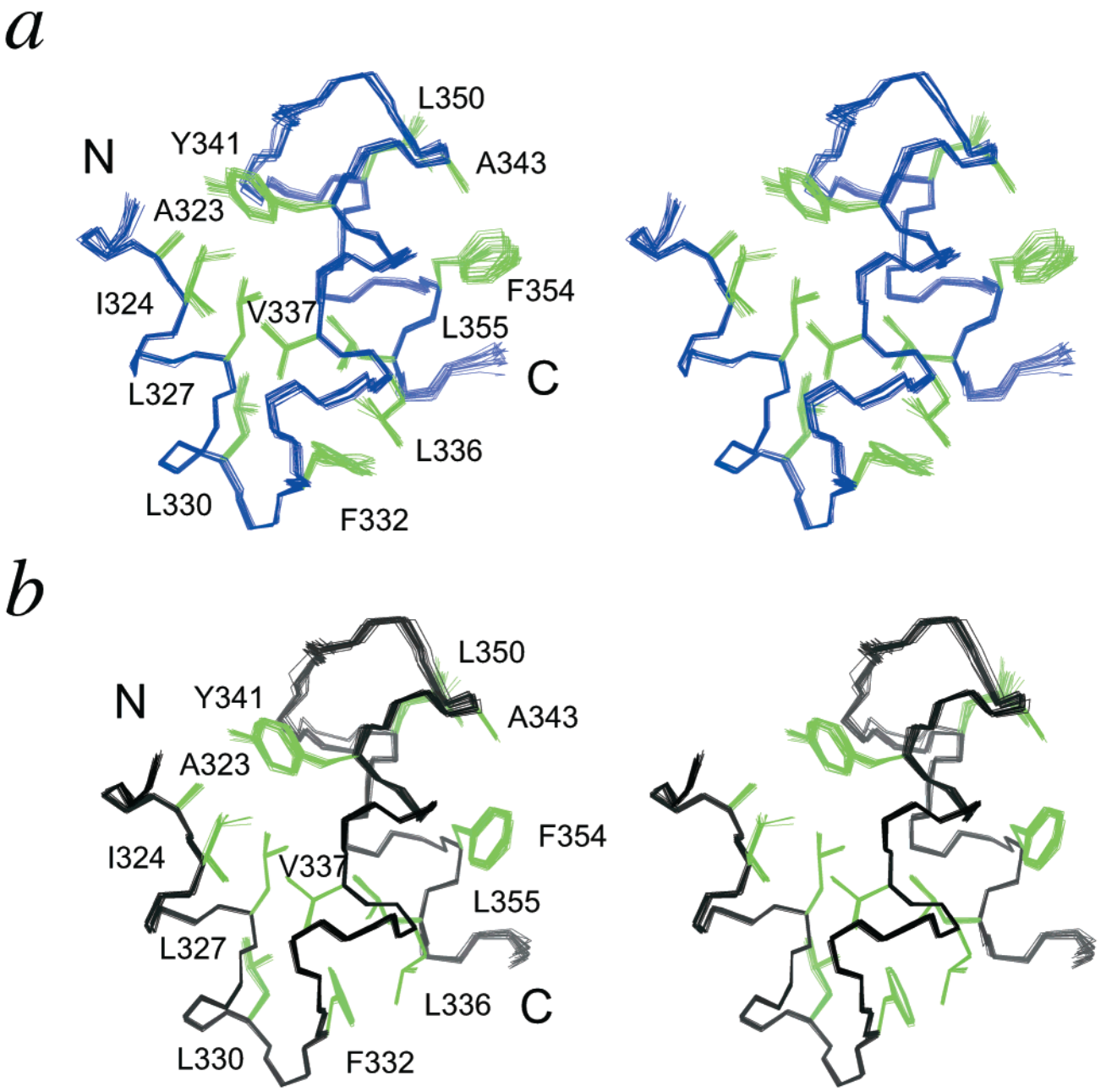


FIGURE 4: Stereoviews of superpositions of the lowest energy structures of (A) wild-type UBA(2) (18 structures) and (B) UBA(2)P333E (21 structures). Backbone atoms for residues 320−357 are shown. The C-terminal residues 358−363 are disordered and are not shown here. Side chain atoms for the residues contributing to the hydrophobic core are displayed in green.

we predicted that the Pro residue at position 333 of HHR23A UBA(2) might be critical for interaction with Vpr. To assess the contribution of this residue to the HHR23A UBA(2)/Vpr interaction, a mutant HHR23A UBA(2) peptide containing a glutamic acid residue at position 333 [HHR23A UBA(2)P333E)] was chemically synthesized and assayed for binding. This P333E substitution completely abolished binding of the peptide to GSTVpr (Figure 3, panel C). The HHR23A UBA(1) domain contains a glutamic acid residue at the position corresponding to position 333 in the HHR23A UBA(2) domain, and the c-cbl UBA domain contains a serine at that position. These results indicate that the proline at position 333 of HHR23A UBA(2) is necessary for interaction with GSTVpr.

Further studies with mutant synthetic UBA(2) peptides demonstrate that the Pro at position 333 is critical for binding but not sufficient on its own for the interaction of UBA(2) with GSTVpr. Neither an E370P substitution nor a $_{370}$PES$_{372}$ substitution made within the *S. cerevisiae* UBA(2) confers upon those peptides the ability to bind GSTVpr (Figure 3, panel C). These in vitro binding results demonstrate the importance of the hydrophobic GFP-loop region for interaction of UBA(2) with Vpr.

Refined High-Resolution Structures of Wild-Type UBA-(2) of HHR23A. To further explore the role of the critical P333 residue on binding of UBA(2) to Vpr, we undertook to determine the structure of UBA(2) P333E. During the assignment of this mutant protein using 2D proton NMR data recorded from a synthetic peptide, we discovered a mismatch between the assignments for the wild-type UBA(2) (*29*) and the mutant. To resolve the conflict, the structure of HHR23A UBA(2) was redetermined using heteronuclear NMR spec-

troscopy on a ^{13}C-,^{15}N-labeled protein. The high resolution structure was necessary for the structure comparison, to analyze the small structural changes that might be caused by the point mutation of Pro333 to Glu. For the preparation of labeled protein, the cDNA of UBA(2) encoding the residues 319−363 was cloned into the expression vector pGEX-2T. Analysis of the isotope-edited 3D data revealed a misassignment in the original data due to spectral overlap. Specifically, NOEs were incorrectly assigned between the side chain of Leu330 and the aromatic ring protons of Tyr341. The correct NOE connectivities are found to be between the side chains of Ile324 and Tyr341 as well as Glu348 and Tyr341. For the highly refined structure of the wild-type UBA(2) domain 779 NOE-derived distance restraints (vs 380 in the previously reported structure) were obtained. Simulated annealing calculations were employed to generate an ensemble of 18 structures (Figure 4, panel A) consistent with the NMR data. The structures exhibit good covalent geometry and have no NOE violations greater than 0.5 Å. Figure 4, panel A, shows the quality of the calculated structures with a root-mean-square (rms) deviation between the atomic coordinates of the backbone and all heavy atoms with respect to the average coordinates for the residues within the three α-helices (residues 320−330, 337−344 and 350−357) of only 0.178 ± 0.037 Å and 0.686 ± 0.064 Å, respectively. Complete restraint and structural statistics are presented in Table 1.

In the refined structure, the three helix bundle is more compact and the orientation of the interhelical angles changed slightly (Figure 4, panel A). Interhelical angles were determined using the program RIBBONS2.6 (*36*). Helices 1 and 2 are arranged at an angle of 120° (equivalent to −60°) and helices 2 and 3 exhibit an angle of 126° (or −54°). These values are close to the optimum of 52° reported for a "knob in hole" packing of the side chains. In contrast, the interhelical angle between helices 1 and 3 is only about 30°. A tight "knob in hole" packing for the side chains of these two helices is not required due to the larger distance between them. The resulting change in the structure of the UBA(2) domain is mainly a shift of the first α-helix of the three-helical bundle along its axis of about 3 Å. The orientation of the first helix with respect to the other two helices is only slightly affected (about 15°).

The structure of the UBA(2) domain is well-ordered from residues Glu320 to Ser357. In contrast to the low resolution structure of UBA(2) (*29*), a full set of NOEs of the type $H_{\alpha i}H_{\beta i+3}$ and $H_{\alpha i}NH_{i+3}$, were found for residues Glu320 to Leu330, showing that the first helix is extended to the N-terminal end of the UBA(2) domain. The C-terminal residues Asn359 to Asp363 show no medium or long range NOE connectivities and are disordered. Carbon and proton chemical shifts for these residues are similar to those found for random coil conformations. The first loop connecting helices 1 and 2 is well-ordered, and the residues Gly331 to Leu336 show many NOE connectivities to residues located in the hydrophobic core. NOE connectivities for P333 clearly indicate that the Pro residue adopts a trans conformation. The second loop, consisting of the residues Glu345 to Asn349, is far more exposed on the surface. Line broadening for several residues in this loop, namely Lys346 and Glu348, indicates that this loop is dynamic, likely due to slow chemical exchange processes.

Table 1: Input Data for Structure Calculation and Structural Statistics UBA(2) [18 Structures] and UBA(2)P333E [21 Structures]

Distance Restraints	UBA(2)	UBA(2)P333E
total	796	843
NOE-derived		
intraresidue	173	204
sequential ($\|i-j\|=1$)	196	182
medium-range ($\|i-j\|\leq 4$)	219	226
long-range ($\|i-j\|>4$)	191	214
hydrogen bonds		
medium-range (within-helices)	17	17
Structure Statistics		
RMSD from experimental distance restraints (Å)	0.047 ± 0.001	0.052 ± 0.0003
RMSD from idealized geometry		
bonds (Å)	0.005 ± 0.00004	0.005 ± 0.00005
angles (deg)	0.616 ± 0.007	0.617 ± 0.004
energies (kcal mol^{-1})		
E_{NOE}	11.7 ± 0.4	9.8 ± 0.5
E_{bond}	17.1 ± 0.3	18.3 ± 0.4
E_{angles}	74.9 ± 1.6	74.8 ± 1.0
$E_{L.-J.}$[a]	−262± 9	−270 ± 6
RMSD from avg structure (Å)		
backbone atoms in secondary structure[b]	0.18 ± 0.04	0.12 ± 0.03
backbone atoms for residues 3−9	0.21 ± 0.06	0.16 ± 0.07
all heavy atoms in secondary structure[b]	0.69 ± 0.06	0.66 ± 0.08
all heavy atoms for residues 3−39	0.72 ± 0.07	0.71 ± 0.07
Procheck analysis		
residues in most favored region (%)	79.7	76.1
residues in additional allowed region (%)	15.9	20.9
residues in generously allowed regions (%)	3.7	2.6
residues in disallowed regions (%)	0.8	0.4

[a] Lennart−Jones energy was calculated after the accepted structures were subjected to a final minimization (50 steps Adopted Raphson Newton algorithm, QUANTA98) using the CHARMM force field and the NOE-derived distance restraints. [b] Secondary structure is composed of residues 3−12, 19−25, and 32−39. These residues are numbered 321−330, 337−343, and 350−357 in HHR23A.

One interesting feature of the UBA(2) domain is the presence of large hydrophobic patches at the protein surface (Figure 5). Several hydrophobic residues that show contacts with the hydrophobic core are partially exposed on the surface. The largest area is made up of residues Leu327, Ala329, and Leu330 in helix 1, residues Gly331, Phe332, Pro333, and Leu336 in the loop 1, and residues Ala351 and Leu355 in helix 3, covering a total of about 596 Å^2 (epitope 1). This corresponds to approximately 18% of the total solvent accessible surface area of 3288 Å^2. A second hydrophobic region (epitope 2) located on the opposite side of the UBA(2) molecule consists of the residues Ala323, Ile324 (helix 1), Ile338, Phe341, Ala342 (helix2), Leu349 (helix3), and Phe359 (C-terminus). The total surface area is 555 Å^2, but in contrast to the hydrophobic epitope 1, this area is not as compact and the residues do not form a consecutive epitope. Such large hydrophobic surface patches are rather unusual and as previously noted are likely binding sites for other proteins. The site of interaction of HIV-1 Vpr with the UBA(2) domain was therefore suggested to occur

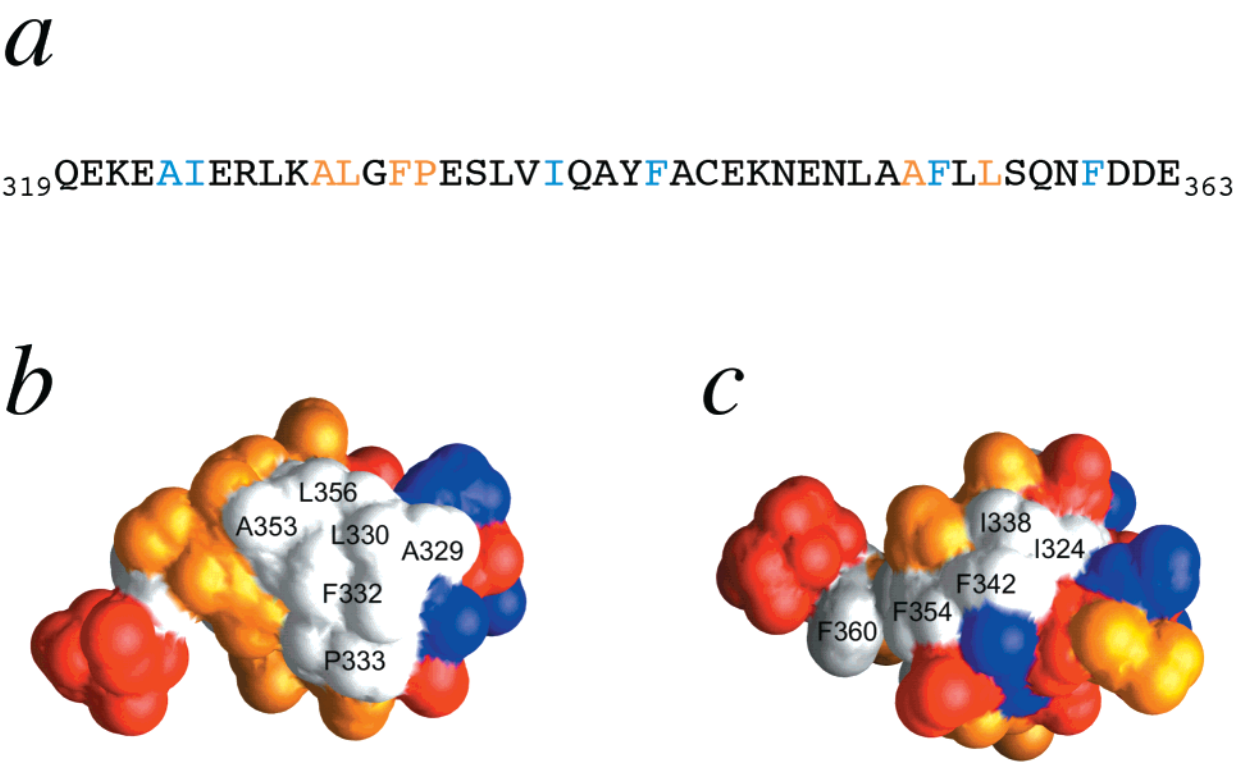


FIGURE 5: (A) Amino acid sequence of HHR23A UBA(2) with the amino acids that comprise the hydrophobic epitope 1 and 2 highlighted in yellow and blue, respectively. (B) Hydrophobic surface view of UBA(2) showing the hydrophobic epitope 1, which is the proposed binding site for HIV-1 Vpr. (C) View of the surface representation as in panel B but rotated upon the *x*-axis about 180°, showing the hydrophobic epitope 2. Hydrophobic amino acids (A, G, F, I, L, M, P, V) are colored in white, negatively charged amino acids (D, E) in red, positively charged in blue (K, R), and polar residues (N,Q, S, T) are in orange.

between the hydrophobic region of Vpr and epitope 1. Epitope 2 is a potential site for other protein−protein interactions in the XPC complex.

Comparison of the Structures of Wild-Type UBA(2) and Mutant P333E. As discussed, a single point mutation of the residue Pro333 to a Glu has been shown to completely abolish the binding of HIV-1 Vpr to UBA(2). To investigate whether the impaired binding is due to a change in structure or the change in a critical interaction with the amino acid side chain, we determined the structure of the UBA(2) mutant P333E (Figure 4, panel B). A synthetic peptide was used for the NMR analysis. For the final structure calculations, 826 NOE-derived distance restraints were used, of which 204 are intraresidual, 182 are sequential, 226 are medium range, and 214 are long-range NOE connectivities (Table 1). Twenty-one of 100 calculated structures were selected based on lowest total and NOE energy. Figure 4, panel B, shows the high structural precision of 0.121 ± 0.031 Å for the backbone atoms and 0.661 ± 0.080 Å for the heavy atoms for the three helices. The high-resolution structures for the wild-type and mutant UBA(2) domain were of similar resolution enabling us to do a detailed structure/function analysis.

The coordinates for the backbone atoms within the three helices of the wild-type UBA(2) and mutant P333E are nearly identical, showing that the only structural changes between the wild-type and mutant UBA(2) occur locally around the mutated residue 333 (Figure 4). Comparison of the proton chemical shifts of the wild-type and the mutant resonances also indicates that there is no change in the overall structure (data not shown). Differences in the proton chemical shift larger than 0.1 ppm are observed for the residues close to the site of mutation, i.e., for the residues Gly331, Phe332, Glu334, and Ser335.

Figure 6 shows the superposition of the structural ensemble of wild-type UBA(2) and the mutant P333E for the residues Leu327 to Val337 around the GFP-loop. The structures were superimposed using the backbone atoms within the three helices. Structural changes within loop 1 between wild-type and mutant are small but significant. The root-mean-square deviation to the mean within loop 1 is about 0.2 Å for wild type and P333E. However, the difference in the positions of the amide nitrogens of residues 333 and 334 are about 1.2 and 0.8 Å respectively. This clearly indicates that the structural changes observed are significant within the accuracy of the NMR data. The largest change in main chain conformation occurs at the site of mutation. The backbone torsion angles for the residue Phe332 preceding the mutated residue are different for the wild-type and the mutant protein. The difference of the φ angle for Phe332 angle is small, about −152° for the wild-type and −168° for the mutant. In contrast, the ψ angle for Phe332 is drastically altered from about −138° for the wild type to −77° for the mutant P333E. This results in a different main chain conformation for the loop 1 for wild type and mutant from residue Phe332 to residue Ser335. Starting with residue Val337 the main chain and side chain coordinates can again be superimposed within the error limits of the NMR data. The differences in main chain torsion angle also lead to different packing for the residues in contact with residue 333. Pro333 of the wild-type UBA(2) shows contacts to other residues participating in the hydrophobic core, namely, residues Phe332 and Leu336. The χ_1 torsion angle of residue Phe332 is changed from −60° to −52°, and the χ_2 torsion angle is altered from

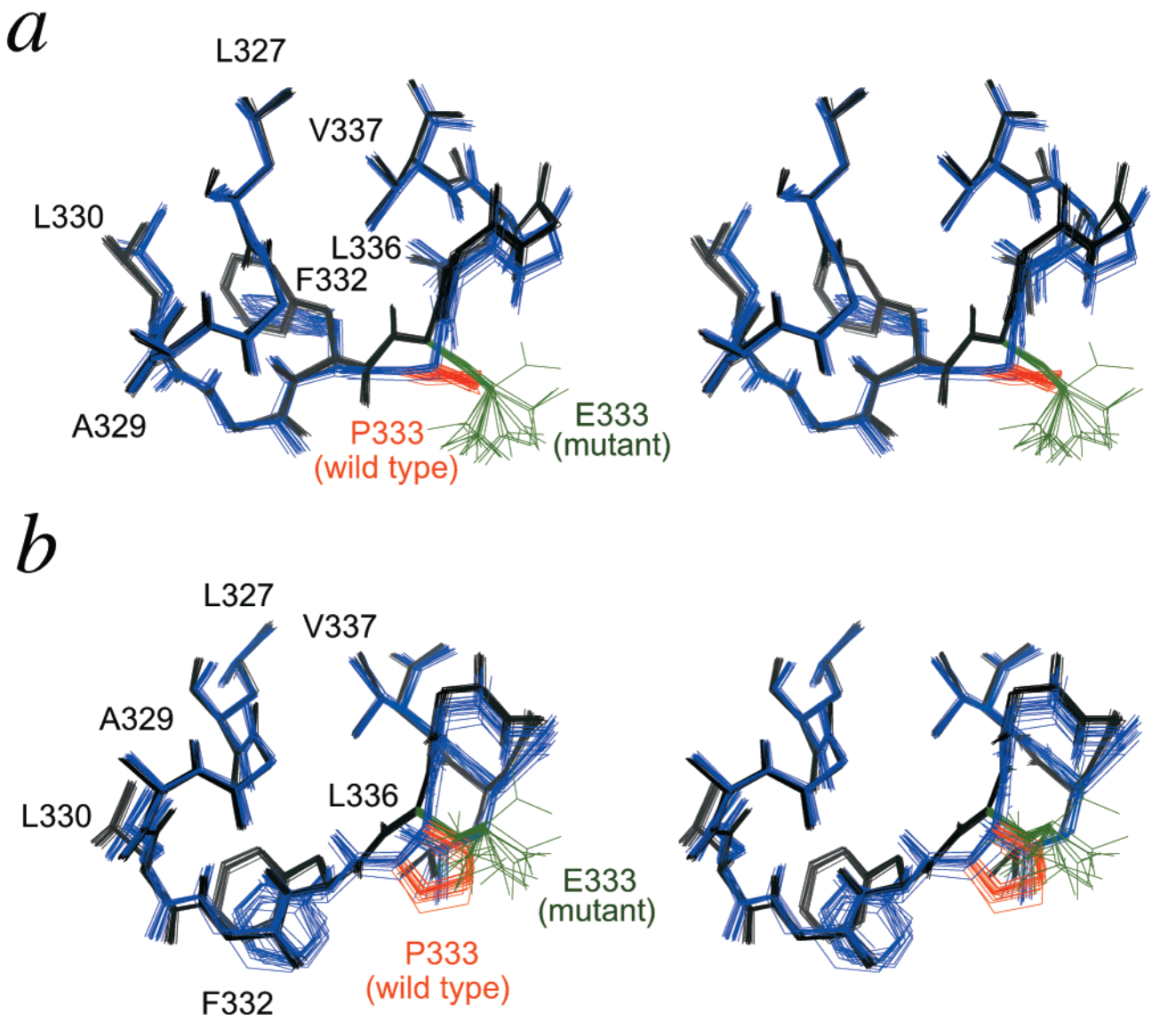


FIGURE 6: (A) Comparison of the structure of loop 1 (residues Leu327 to Val337) of UBA(2) (black) and UBA(2)P333E (blue). (B) Same as in panel A but rotated around the *x*-axis by approximately 90°. The site of mutation is illustrated by different colors for the respective side chains, red for the wild-type Pro333 and green for the mutant Glu333. The superposition of the structures reveals that the differences between the loop conformations of the wild-type and mutant proteins are larger than the rms deviation within their structure ensembles. Residues farther away from the mutation, e.g., Leu327 and Val337, have identical side chain conformations in the wild-type and mutant protein domain structures. The major structural change occurs at the backbone ψ torsion angle for residue Phe332, which leads to a different side chain conformation of the Phe residue and a different orientation of the mutated residue 333.

107° to 152° for the wild-type vs P333E mutant proteins, respectively. Differences of a similar order of magnitude are observed for residue Leu336 as well. The torsion angles χ_1 and χ_2 are changed by about 16° (from 68° for wild-type to 84° for P333E) and 54° (from −172° for the wild-type to −118° for P333E), respectively. Residue Leu330, which has no direct contact to residue 333, exhibits only small changes for the side chain orientation as an effect of the packing against residue Phe332. The side chain torsion angles for residues Leu327 and Val337, both of which are packed in the hydrophobic core, are identical within the error limits of the NMR data.

Comparison of Structural Data with the Results of S. cer. Rad23A Mutagenesis. A NOESY spectrum of the *S. cerevisiae* Rad23A UBA(2) domain showed a very similar NOE pattern to the wild-type UBA(2) domain (data not shown). This indicates that, like the P333E mutant, it exhibits the same overall fold and secondary structure as the HHR23A UBA(2). The *S. cerevisiae* Rad23A UBA(2) domain does not bind to Vpr, as discussed. In comparison to the UBA(2) domain of HHR23A, the UBA(2) domain of *S. cerevisiae* differs in the amino acid sequence in the vicinity of the GFP-loop, specifically Ala329 is substituted by a negatively charged Glu, Pro333 is exchanged by a Glu, the subsequent amino acid Glu334 is charge-reversed with an Arg residue, and the Ser335 residue is replaced by a negatively charged Asp. Analysis of a UBA(2) chimera derived from HHR23A and *S. cerevisiae* demonstrated that the Pro residue within the GFP-loop is insufficient for binding to Vpr (Figure 3, panel C). Replacement of the Glu residue found at position 370 in the UBA domain of *S. cerevisiae* [equivalent to position 333 in HHR23A UBA(2)] with a Pro does not restore binding to HIV-1 Vpr, indicating that there are additional determinants for the binding of Vpr. Using the sequence alignment of Figure 3, panel A, we can predict the equivalent hydrophobic epitope 1 for the UBA domain

of *S. cerevisiae*. This analysis shows that in addition to the change to Glu at position 333, two other residues that are part of the hydrophobic patch are different between the two UBA domains, i.e., Ala329 and Leu356, which are replaced by Glu and Phe, respectively. This change of two small hydrophobic residues for a charged and much larger amino acid, respectively, is likely to have a significant effect on the structure of the proposed interaction surface between UBA(2) and Vpr.

SUMMARY AND CONCLUSIONS

The UBA domains are found in a number of proteins besides HHR23A with diverse functions involved in ubiquitination, DNA repair, and signaling pathways via protein kinases (*16*). One of the unusual features of the UBA(2) domain of HHR23A is the presence of large hydrophobic patches at the protein surface. The existence of these large hydrophobic patches strongly suggests that these may serve as binding sites for other proteins. For HIV-1 Vpr, we confirmed by site-directed mutagenesis that one of the hydrophobic patches encompassing Pro333 is critical for the interaction between Vpr and UBA(2) (*18*). To understand the structural basis of the loss in binding upon mutation of Pro 333 to Glu, we determined the structures of the wild-type UBA(2) domain and the mutant P333E to high resolution. A comparison shows that while the overall fold of the UBA domain is not affected by the substitution, there are significant changes that occur locally around the site of mutation. The high accuracy of the structures revealed that substitution of a Pro residue 333 with a Glu resulted in changes of the shape and charge of the hydrophobic patch predicted to be the Vpr-interacting domain. These changes account for the loss of the ability to bind Vpr.

The structural and genetic information developed in the study should facilitate the identification and characterization of cellular proteins that bind through UBA. One hypothesis that can be addressed with the structural knowledge obtained here is that Vpr may compete or alter the interaction between UBA(2) and a normal cellular protein involved in DNA repair signaling processes. To date, despite their ubiquitous occurrence, only two cellular binding partners for UBA domains have been identified. Ubiquitin has been shown to bind to the UBA domain of p62, a ligand of the SH2 domain of $p56^{lck}$ (*37*). However, the very low sequence homology between the UBA domain of p62 and UBA(2) of HHR23A does not suggest a similar interaction between UBA(2) and ubiquitin. Significantly for this work, HHR23A and B have been shown to interact with the 3-methyladenine DNA glycosylase (MPG) protein and have been proposed to serve as universal DNA damage recognition accessory proteins. Furthermore, MPG binds specifically to the UBA(2) of HHR23A. It is therefore possible that HIV-1 Vpr may compete with MPG for binding to HHR23A, thus affecting DNA nucleotide excision repair. This provides support for our hypothesis that the interaction of HIV-1 Vpr with this domain may be important to its functional role in mediating cell cycle arrest.

ACKNOWLEDGMENT

We thank Dr. Thorsten Dieckmann for advice in the early stages of this work. We thank Dr. Mark A. Jarosinski and the Peptide Synthesis group at AmGen (Boulder, Co) for synthesis of the peptides used in the binding assays and the UBA(2)P333E peptide used in the structural studies and Mr. Evan Feinstein for help in manuscript preparation.

REFERENCES

1. Lombaerts, M., Goeloe, J. I., den Dulk, H., Brandsma, J. A., and Brouwer, J. (2000) *Biochem. Biophys. Res. Commun. 268*, 210−215.
2. Watkins, J. F., Sung, P., Prakash, L., and Prakash, S. (1993) *Mol. Cell. Biol. 13*, 7757−7765.
3. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., Vanderspek, P. J., Bootsma, D., Hoeijmakers, J. H. J., and Hanaoka, F. (1994) *EMBO J. 13*, 1831−1843.
4. Madura, K., and Prakash, S. (1990) *Nucleic Acids Res. 18*, 4737−4742.
5. Guzder, S. N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1995) *J. Biol. Chem. 270*, 12973−12976.
6. Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., Vanderspek, P. J., Bootsma, D., Hoeijmakers, J. H. J., and Hanaoka, F. (1996) *Mol. Cell. Biol. 16*, 4852−4861.
7. Masutani, C., Araki, M., Sugasawa, K., van der Spek, P. J., Yamada, A., Uchida, A., Maekawa, T., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1997) *Mol. Cell Biol. 17*, 6915−6923.
8. Li, L., Lu, X. Y., Peterson, C., and Legerski, R. (1997) *Mutat. Res. 383*, 197−203.
9. Wang, Z. G., Wei, S. G., Reed, S. H., Wu, X. H., Svejstrup, J. Q., Feaver, W. J., Kornberg, R. D., and Friedberg, E. C. (1997) *Mol. Cell. Biol. 17*, 635−643.
10. Jansen, L. E. T., Verhage, R. A., and Brouwer, J. A. (1998) *J. Biol. Chem. 273*, 33111−33114.
11. Guzder, S. N., Sung, P., Prakash, L., and Prakash, S. (1998) *J. Biol. Chem. 273*, 31541−31546.
12. Mueller, J. P., and Smerdon, M. J. (1996) *Mol. Cell. Biol. 16*, 2361−2368.
13. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W., and Madura, K. (1998) *Nature 391*, 715−718.
14. Lambertson, D., Chen, L., and Madura, K. (1999) *Genetics 153*, 69−79.
15. Hiyama, H., Yokoi, M., Masutani, C., Sugasawa, K., Maekawa, T., Tanaka, K., Hoeijmakers, J. H. J., and Hanaoka, F. (1999) *J. Biol. Chem. 274*, 28019−28025.
16. Hofmann, K., and Bucher, P. (1996) *Trends Biochem. Sci. 21*, 172−173.
17. Miao, F., Bouziane, M., Dammann, R., Masutani, C., Hanaoka, F., Pfeifer, G. P., and O'Connor, T. R. (2000) *J. Biol. Chem. 275*, 28433−28438.
18. Withers-Ward, E. S., Jowett, J. B. M., Stewart, S. A., Xie, Y. M., Garfinkel, A., Shibagaki, Y., Chow, S. A., Shah, N., Hanaoka, F., Sawitz, D. G., Armstrong, R. W., Souza, L. M., and Chen, I. S. Y. (1997) *J. Virol. 71*, 9732−9742.
19. Emerman, M. (1996) *Curr. Biol. 6*, 1096−1103.
20. Jowett, J. B. M., Planelles, V., Poon, B., Shah, N. P., Chen, M. L., and Chen, I. S. Y. (1995) *J. Virol. 69*, 6304−6313.
21. He, J. L., Choe, S., Walker, R., Dimarzio, P., Morgan, D. O., and Landau, N. R. (1995) *J. Virol. 69*, 6705−6711.
22. Re, F., Braaten, D., Franke, E. K., and Luban, J. (1995) *J. Virol. 69*, 6859−6864.
23. Rogel, M. E., Wu, L. I., and Emerman, M. (1995) *J. Virol. 69*, 882−888.
24. Stewart, S. A., Poon, B., Jowett, J. B. M., and Chen, I. S. Y. (1997) *J. Virol. 71*, 5579−5592.
25. Mahalingam, S., Ayyavoo, V., Patel, M., KieberEmmons, T., and Weiner, D. B. (1997) *J. Virol. 71*, 6339−6347.
26. Di Marzio, P., Choe, S., Ebright, M., Knoblauch, R., and Landau, N. R. (1995) *J. Virol. 69*, 7909−7916.
27. Goh, W. C., Rogel, M. E., Kinsey, C. M., Michael, S. F., Fultz, P. N., Nowak, M. A., Hahn, B. H., and Emerman, M. (1998) *Nat. Med. 4*, 65−71.

28. Poon, B., Grovit-Ferbas, K., Stewart, S. A., and Chen, I. S. Y. (1998) *Science 281*, 266−269.
29. Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. Y., and Feigon, J. (1998) *Nat. Struct. Biol. 5*, 1042−1047.
30. Blake, T. J., Shapiro, M., Morse, H. C., and Langdon, W. Y. (1991) *Oncogene 6*, 653−657.
31. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*: *Principles and Practice*, Academic Press, Inc., San Diego.
32. Bartels, C., Xia, T. H., Billeter, M., Guntert, P., and Wuthrich, K. (1995) *J. Biomol. NMR 6*, 1−10.
33. Glaser, R. W., and Wüthrich, K. (1997) http://www.mol.biol.ethz.ch/wuthrich/software/spscan/.
34. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
35. Brünger, A. T. (1992) *X-PLOR* (Version 3.1) Manual, Yale University Press, New Haven and London.
36. Carson, M. (1991) *J. App. Crystal. 24*, 958−961.
37. Vadlamudi, R. K., Joung, I., Strominger, J. L., and Shin, J. (1996) *J. Biol. Chem. 271*, 20235−20237.

BI0017071