Structure of the Mouse Peptide N-Glycanase-HR23 Complex Suggests Co-evolution of the Endoplasmic Reticulum-associated Degradation and DNA Repair Pathways*§

Received for publication, January 5, 2006, and in revised form, February 22, 2006 Published, JBC Papers in Press, February 24, 2006, DOI 10.1074/jbc.M600137200

Gang Zhao†§, Xiaoke Zhou†§, Liqun Wang†§, Guangtao Li†, Caroline Kisker‡¶, William J. Lennarz†, and Hermann Schindelin†§¶

From the †Center for Structural Biology, ‡Department of Biochemistry and Cell Biology, and §Department of Pharmacological Sciences, Stony Brook University, Stony Brook, New York 11794 and the ¶Rudolf Virchow Center for Experimental Biomedicine and Institute of Structural Biology, University of Würzburg, Versbacher Strasse 9, 97078 Würzburg, Germany

Peptide N-glycanase removes N-linked oligosaccharides from misfolded glycoproteins as part of the endoplasmic reticulum-associated degradation pathway. This process involves the formation of a tight complex of peptide N-glycanase with Rad23 in yeast and the orthologous HR23 proteins in mammals. In addition to its function in endoplasmic reticulum-associated degradation, HR23 is also involved in DNA repair, where it plays an important role in damage recognition in complex with the xeroderma pigmentosum group C protein. To characterize the dual role of HR23, we have determined the high resolution crystal structure of the mouse peptide N-glycanase catalytic core in complex with the xeroderma pigmentosum group C binding domain from HR23B. Peptide N-glycanase features a large cleft between its catalytic cysteine protease core and zinc binding domain. Opposite the zinc binding domain is the HR23B-interacting region, and surprisingly, the complex interface is fundamentally different from the orthologous yeast peptide N-glycanase-Rad23 complex. Different regions on both proteins are involved in complex formation, revealing an amazing degree of divergence in the interaction between two highly homologous proteins. Furthermore, the mouse peptide N-glycanase-HR23B complex mimics the interaction between xeroderma pigmentosum group C and HR23B, thereby providing a first structural model of how the two proteins interact within the nucleotide excision repair cascade in higher eukaryotes. The different interaction interfaces of the xeroderma pigmentosum group C binding domains in yeast and mammals suggest a co-evolution of the endoplasmic reticulum-associated degradation and DNA repair pathways.

Degradation of misfolded and misassembled proteins is essential for homeostasis and cell growth. In the endoplasmic reticulum (ER) a quality control system sorts proteins on the basis of their conformation (1). Proteins that fail to mature in the ER are retro-translocated into the cytosol, polyubiquitinated, and targeted to the proteasome for degradation as part of the ER-associated degradation pathway (2). Defects in the ER-associated degradation system have been associated with many diseases including α-1 antitrypsin deficiency, cystic fibrosis, neurodegenerative disorders, tyrosinase deficiency, and cancer (3–6).

Most of the misfolded glycoproteins are deglycosylated by a cytosolic peptide-N-glycanase (PNGase) activity before proteasomal degradation (7, 8). PNGase hydrolyzes the β-aspartylglycosylamine bond of N-linked glycoproteins, which facilitates their degradation by the proteasome (9). PNGase is highly conserved from yeast to human, supporting its functional significance. In addition to the catalytic core, in higher eukaryotes PNGase contains one additional domain each at the N and C terminus, which may function as binding partners for other proteins (10).

PNGase closely associates with the proteasome by either direct binding to components of the proteasome or through its interaction with HR23A/HR23B, the mammalian homologs of Rad23 in yeast (10, 11). HR23 participates in the ubiquitin-proteasome pathway by interacting with ubiquitin and components of the 26 S proteasome (12, 13). HR23 is also involved in the nucleotide excision repair (NER) pathway and forms a tight complex with xeroderma pigmentosum group C (XPC), the ortholog of yeast Rad4 in higher eukaryotes (14, 15). The XPC-HR23B complex is involved in damage recognition and binds to damaged DNA during the early stages of the NER pathway, thereby initiating global genome NER and recruiting TFIIH to the site of damage (16–18). The precise role of HR23 in NER is not clear, but it was shown to stabilize XPC and/or prevent its proteasomal destruction (14, 19). In yeast Rad4 interacts in an analogous fashion with Rad23 as part of the NER pathway. Interestingly, bioinformatics studies demonstrated that both XPC/Rad4 and PNGase belong to the transglutaminase family (20) and, hence, share a common fold. The interaction between PNGase and HR23 further strengthens the proposed relationship between the proteasome-dependent protein degradation and NER pathways (21, 22) and also provides first insights into how XPC interacts with HR23 to form the damage recognition complex.

HR23 contains an N-terminal ubiquitin-like domain and two ubiquitin-associated domains. The stretch of residues responsible for XPC binding (XPCB) was mapped to a highly conserved region between the two ubiquitin-associated domains (23). The solution structure of the...
PNGase-HR23 Complex Crystal Structure

XPCB domain reveals that it consists of five amphipathic helices featuring hydrophobic patches on the surface that have been predicted to be involved in protein-protein interactions (24, 25). This agrees with previous studies indicating that the interaction between HR23 and XPC is predominantly hydrophobic (26). We present here the crystal structure of the mouse PNGase (mPNGase) core domain in complex with the XPCB domain of mouse HR23B (mHR23B) and propose a similar interaction between XPC and the HR23 proteins in mammals.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The DNA sequence encoding residues 165–450 (core domain) of mPNGase was amplified by PCR from full-length (residues 1–651) pET28a-mPNG1 (10) and cloned into the pET21b vector using the Ndel and XhoI restriction sites. The cDNA encoding mHR23B was cloned into the pET21b vector using the same sites (pET21b-mHR23B). The coding sequence of the XPCB domain (residues 273–332) of mHR23B was amplified from pET21b-mHR23B and inserted into the pTXB1 vector (Ndel/SapI) (New England Biolabs, Beverly, MA). Full-length and core domains of mPNGase and full-length mHR23B were overexpressed and purified as described previously (27). The XPCB domain was overexpressed in Escherichia coli BL21(DE3) cells (Novagen) at 15 °C overnight by inducing a mid-log culture with 1 mM isopropyl 1-thio-β-D-galactopyranoside. The protein was purified by chitin affinity chromatography according to the manufacturer’s instructions (New England Biolabs) as well as size exclusion chromatography on a Superdex 200 (Amersham Biosciences) column and was concentrated to 8 mg/ml.

Crystallization and Structure Determination—The mPNGase core domain and the XPCB domain were mixed in a 1:1 molar ratio, and a total protein concentration of 9.5 mg/ml was used for crystallization. Crystals of the complex were obtained by vapor diffusion against a reservoir containing 0.1 M Tris-HCl, pH 8.5, 28–32% polyethylene glycol 4000 and 0.2 M sodium acetate. The crystals reached their final size of 0.2 mm in each dimension within 5–10 days. The crystals belong to the monoclinic space group C2 with a = 97.3 Å, b = 52.4 Å, c = 81.3 Å, and β = 114° and contain one complex in the asymmetric unit. The structure was solved by multiple isomorphous replacement and anomalous scattering using mercury, platinum, and gold derivatives. Heavy atom derivatives were prepared by soaking crystals in crystallization buffer with 1 mM ethyl mercaptoethanol (4 h), 1 mM K3[PtCN4] (4 h), or 1 mM KAuCN5 (24 h). Cocrystallization of the protein complex with 5 mM concentrations of the inhibitor Z-VAD- fluoromethyl ketone (Sigma-Aldrich) was conducted under similar conditions. The inhibitor-bound crystals were isomorphous with the native crystals, and the structure was solved using isomorphous replacement. After transfer into a cryoprotectant containing 20% glycerol, the crystals were cryocoated and stored in liquid nitrogen. Diffraction data of the native, inhibitor-soaked and heavy atom-derivatized crystals were collected at 100 K on a Rigaku RU-HR3 rotating anode x-ray generator operating at 50 kV and 100 mA and equipped with confocal multilayer optics and a Rigaku RAXIS-IV++ detector. Due to its strong diffraction, a second high resolution dataset to 1.85 Å was collected for the gold derivative, which was used for refinement.

Diffraction data were processed using the HKL software (28). Phases at 3.0 Å were calculated with Solve (29) and Resolve (30), and the model was manually built in O (31). The coordinates of XPCB could easily be fitted into the initial maps together with helical segments of the mPNGase core domain. The partial model was refined in an iterative process, allowing the gradual incorporation of most residues in the resulting electron density maps. The model was refined at a resolution of 1.85 Å with REFMAC (32), and water molecules were added with ARP (33). All data between 20 Å and the respective high resolution limits were included. Additional calculations were done with the CCP4 suite (34).

Protein-Protein Interaction Studies—mPNGase and mHR23B proteins were mixed in buffer (20 mM Tris-HCl, pH 8.5, 250 mM NaCl, and 5% glycerol) and incubated for 30 min at 4 °C. A Superdex 200 HR 10/30 (Amersham Biosciences) analytical size exclusion chromatography column was calibrated with a molecular weight marker kit (MW-GF-200, Sigma). Individual proteins and the protein complexes were analyzed on the column, and the elution profiles were monitored by the absorbance at 280 nm, or 215 nm in the case of XPCB. The corresponding peak fractions were visualized either on 12% SDS-PAGE or 12% Tricine SDS-PAGE gels.

Calorimetric titrations were carried out with a VP-ITC calorimeter (MicroCal) at 25 °C. Proteins were dialyzed overnight at 4 °C against 2 liters of buffer C with 1 mM β-mercaptoethanol, and the solutions were thoroughly degassed before each titration. Purified PNGase (10–25 μM) in the sample cell was titrated with 100–250 μM concentrations of either full-length yRAD23 or full-length mHR23B or the mXPCB domain. The data were processed with the MicroCal Origin 7.0 software. By fitting the data to a single-site binding model, the binding stoichiometry, apparent Kd, and binding enthalpy ΔH were determined. Isothermal titration calorimetry (ITC) experiments were carried out in duplicate.

Sequence Alignment and Modeling Studies—The sequence alignment of the common RAD4-related sequences from mPNGase (305–445) and the XPC proteins from mouse (499–656), human (508–663), and Drosophila melanogaster (933–1091) was generated using ClustalW (35) and manually adjusted. The alignment and the mPNGase crystal structure were used to construct homology models of the corresponding region of human XPC using Modeler (36). Forty initial models were built, and the one with the lowest objective energy was selected.

RESULTS AND DISCUSSION

The mPNGase Core Domain Interacts with the XPCB Domain of mHR23B—The mPNGase protein contains a core domain, which is highly similar (42% sequence identity) to yeast PNGase (yPNGase), as well as N- and C-terminal extensions not present in yeast (Fig. 1A). The mPNGase core also shares similarities with the DNA repair proteins XPC/RAD4 (20); for example, PNGase and XPC from mouse are 29% identical within a conserved region, and both proteins form a complex with mHR23, the homolog of yRAD23 (10, 23). Due to the sequence homology between the core of PNGase and XPC, it is likely that the mPNGase core domain interacts with mHR23 in a fashion similar to that of XPC interacting with mHR23. The XPCB region in mHR23 has been mapped to a largely helical region (23), and we hypothesized that this region of mHR23 also interacts with the core domain of mPNGase. Based on secondary structure predictions and sequence alignments, the core domain of mPNGase (mPNGase-core, residues 165–450) and the XPCB fragment of mHR23B (mHR23B-XPCB, residues 273–332) were cloned, and the corresponding proteins were overexpressed and purified to homogeneity.

The in vitro interaction between mPNGase and mHR23B was investigated by analytical size exclusion chromatography (Fig. 1B). In agreement with previous studies (10), we found that full-length mPNGase interacts with full-length mHR23B in a 1:1 molar ratio. Additional chromatographic assays revealed that the core but neither the N-terminal (1–164) nor C-terminal (451–651) fragments of mPNGase, binds to mHR23B (data not shown). More significantly,
we were able to show that the mPNGase core interacts with the XPCB domain of mHR23B (Fig. 1). The complex eluted from the column at a higher apparent molecular mass than the mPNGase core, and when analyzed by SDS-PAGE, this peak contained both proteins (Fig. 1B and inset). The complex is very stable (see below) and was subsequently utilized in our structural studies.

**FIGURE 1. Interaction between the mPNGase core domain and the XPCB domain of mHR23B.**

A, domain structures of mPNGase with the trans-glutaminase-like core in blue, the zinc binding domain (ZBD) in cyan, and mHR23 with the XPCB domain in yellow. UBA, ubiquitin-associated domain, UBL, ubiquitin-like domain. B, size exclusion chromatography of the mPNGase-XPCB domain complex and the individual components. Elution profiles are shown for the core domain (thin gray line) and the XPCB domain at 215 nm (dashed line) since this protein does not absorb at 280 nm. The protein complex eluted as a single peak as shown in the solid curve. The inset shows the 12% Tricine SDS-PAGE separation of the elution peaks of the individual proteins (C and X) and the complex (CX). mAU, milliabsorbance units. C, isothermal titration calorimetric analysis of the complex. The top panel shows the titration (6 µl per injection) of the mHR23B-XPCB domain at a concentration of 250 µM into a 25 µM solution of the mPNGase core domain. The solid line in the lower panel represents the best-fit according to a one-site binding model to the experimental data (open circles).
To further characterize the interaction between mPNGase and mHR23B, ITC experiments were performed (Fig. 1C) by titrating the mPNGase core with mHR23B-XPCB at 25 °C. Analysis of the exothermic binding curve showed that the best fit is obtained with a one site binding model, yielding a binding stoichiometry of 1:1 and an equilibrium dissociation constant (Kd) of 24 nM, which is 3-fold smaller than for the yeast proteins. Furthermore, ITC experiments with full-length yeast PNGase and all six C-terminal histidine residues from the affinity tag.

| TABLE 1 | Data collection and structure solution |
|----------|---------------------------------------|
| Rcryst  | 0.189 (0.239)                        |
| Rfree   | 0.220 (0.293)                        |
| Resolution limits (Å) | 20-1.85 |
| 20-2.2 | 20-2.2 |
| 50-2.3 | 50-2.3 |
| Platinum | Gold |
| Number of reflections | 29,636 |
| 14,551 | 14,551 |
| Number of protein/solvent atoms | 2,896/230 |
| 2,939/60 | 2,939/60 |
| Gold-HR | Z-VAD |
| Deviations from ideal values in | |
| Bond distances (Å) | 0.016 |
| 0.016 | 0.016 |
| 1.547 | 1.559 |
| 0.093 | 0.084 |
| Planar groups (Å²) | 0.006 |
| 0.003 | 0.005 |
| Torsion angles (°) | 6.909, 38.561, 17.370 |
| 8.003, 38.358, 18.514 | 8.003, 38.358, 18.514 |
| Ramachandran statistics | 0.921/0.070/0.006/0.003 |
| 0.840/0.135/0.019/0.006 | 0.840/0.135/0.019/0.006 |
| Average B factors (Å²) (mPNGase core/XPCB domain) | 47.4/46.6 |
| 65.8/65.0 | 65.8/65.0 |

Overall Structure of the mPNGase-XPCB Domain Complex—The crystal structure of the core domain of mPNGase in complex with the XPCB domain of mHR23B was determined by multiple isomorphous replacement (Table 1) and was refined at 1.85 Å resolution (Table 2) to an R-factor of 0.189 (Rfree = 0.239). The final model contains residues 273–333 of the mHR23B XPCB domain and residues 165–450 of mPNGase and all six C-terminal histidine residues from the affinity tag. The stereochemistry of the model is very good; 99% of the residues are located in the most favored and additionally allowed regions of the Ramachandran diagram. While these studies were in progress, the crystal structure of mPNGase in complex with the XPCB domain of yRAD23 was solved (37). Although the structures of the individual components are similar, complex formation is fundamentally different (see below), and thus, the structure of the complex studied by us represents the first detailed view into the interaction between these two proteins in avian and mammalian species.

The mPNGase core contains a catalytic transglutaminase-like fold, a zinc binding domain, and the XPCB association motif (Fig. 2). The transglutaminase-like domain harbors a central, strongly curved, six-stranded antiparallel β-sheet (S6-S11) surrounded by eight α-helices (H3 to H9 and H11) on the concave side and three α-helices (H10, H10*, and H12) on the convex side of the sheet. Secondary structure elements have been numbered according to yPNGase, with the N-terminal helix labeled as H3, since the first two helices of yPNGase are not present in the mouse protein. The zinc binding domain contains a three-stranded β-sheet (S2-S4) with the zinc located at the far end of this domain. In addition a third, very small β-sheet (S1 and S5) is located at the interface between the zinc binding and catalytic domains. The XPCB association motif is located on the opposite side of the zinc binding domain and comprises the C-terminal α-helices H11 and H12.

The transglutaminase-like domain shares high homologies with the catalytic cores of the cysteine protease superfAMILY members, which include N-acetyltransferases, factor XIII transglutaminase, deubiquitinating enzymes, and the papain-like proteases (38, 39). Those enzymes feature a catalytic triad composed of cysteine, histidine, and aspartate residues (Cys-306, His-333, and Asp-350 in mPNGase) in the “cysteine protease” motif consisting of an α-helix, H8 in mPNGase, and four antiparallel
The catalytic core of mPNGase more closely resembles that of N-acetyltransferase and transglutaminase, although they catalyze reactions in the opposite direction, with N-acetyltransferase and transglutaminase resulting in amide formation and PNGase acting as an amidase.

The Active Site—Recently, Z-VAD, a potential caspase inhibitor, was reported to inhibit PNGase by covalently attacking the active site cysteine (40). Z-VAD features a Val-Ala-Asp tripeptide with a fluoromethyl ketone group at the C-terminal end. To illustrate the inhibition mechanism and possibly the catalytic mechanism of PNGase, we determined the structure of the Z-VAD-modified mPNGase-XPCB domain complex at 2.3 Å resolution (Figs. 2, B and C and supplemental Fig. S1). The overall structure of the complex was not perturbed after binding of the inhibitor as evidenced by a root mean square (r.m.s.) deviation of the Cα atoms by only 0.24 Å compared with the complex in the absence of the inhibitor. The complete structure of the inhibitor was defined based on the electron density maps (Fig. 2C). As expected, the reactive end of the inhibitor is covalently linked to the active site cysteine (Cys-306). In addition, the inhibitor is bound to the enzyme via two hydrogen bonds and ionic interactions involving residues Arg-291 and Arg-305 of PNGase and the aspartate side chain of the inhibitor, which is in good agreement with the inhibitor-bound structure of yeast PNGase (37).

Zinc Binding Domain—The zinc binding domain of mPNGase (Fig. 3, left) features a βββ (S2-S4) zinc-ribbon-fold (41, 42). A CXXC motif is located in each of the two extended loops before the first and the third β strands that form the zinc-chelating structure. The CXXC motif is present in all PNGase homologs, and mutagenesis studies revealed that the cysteines are essential for γPNGase activity (43). Zinc binding was suggested to stabilize the enzyme conformation (37). However, the precise role of zinc binding with respect to PNGase activity is not clear. The zinc binding motif shows a scaffold similar to the zinc-ribbon structures of Bacillus stearothermophilus adenylate kinase (Fig. 3, center) and transcriptional elongation factor S-II (Fig. 3, right). By forming a “lid” over the active site of the enzyme, the zinc ribbon of adenylate kinase was
proposed to regulate the thermodynamics of catalysis by stabilizing the intermediate state and promoting product release (42). Similarly, the zinc-ribbon of PNGase may play a role in modulating substrate binding and/or the thermodynamics of the reaction. This may explain how PNGase catalyzes reactions in the reverse direction of transglutaminases, in which the zinc ribbon is absent. The zinc ribbon forms half of the catalytic cleft, and Arg-291 at the end of β4 (Arg-176 in yPNGase) together with Arg-305 (Arg-190 in yPNGase) were observed to form hydrogen bonds and salt bridges with the aspartate of the inhibitor molecule in our structure and in the yeast complex (37). Arg-291/Arg-305 may therefore participate in substrate binding by recognizing the oligosaccharide-attached Asn residue and promote deglycosylation through electrostatic interactions with the Asp residue in the product, which is derived from the Asn to which the carbohydrate was attached.

Structure of the XPCB Domain—The mHR23B-XPCB domain is a compact structure with five highly amphipathic helices (Figs. 2 and 4). It displays varying degrees of structural similarity to the XPCB domain of yRAD23 as present in complex with yPNGase and the solution NMR structures of the apo-XPCB domains of hHR23A (PDB code 1TP4) and hHR23B (PDB code 1PVE) with respective r.m.s. deviations of 2.62 Å for 44 Ca atoms, 3.54 Å for 57 Ca atoms, and 1.17 Å for 54 Ca atoms. The XPCB domain of mHR23B is identical in sequence to its human counterpart (hHR23B) and is very similar to that of hHR23A (90% sequence identity). A comparison of the bound mHR23B-XPCB domain from our structure with the structure of apo-hHR23B reveals only very minor conformational changes in the backbone and surface features, thus suggesting that binding of mPNGase does not induce significant conformational changes in the XPCB domain of mHR23B. In contrast, the XPCB domain of hHR23A adopts a flatter structure compared with that of HR23B. H1 and H4 in HR23B-XPCB each form a 45° angle with the H2/H3 plane, whereas in hHR23A the angles are 0° and 30°, respectively. In addition, the angle between H0 and H1 is 90° in HR23B versus 120° in hHR23A. These architectural differences contribute to the significantly larger r.m.s. deviation between the XPCB...
domain of mHR23B and the corresponding domain of hHR23A. These structural differences are real despite the high level of sequence identity (supplemental Fig. S2) and appear to be of biological relevance and are not the result of different experimental techniques (x-ray crystallography versus NMR), as similar differences were also observed between the NMR models of the two XPCB domains (25).

More importantly, the yRAD23-XPCB and mHR23B-XPCB structures display striking differences in their molecular shapes and the distribution of hydrophobic residues (Fig. 4A) despite the fact that the main chain traces of the yRAD23-XPCP and mHR23B-XPCP structures are quite similar (32% sequence identity). The yRAD23-XPCB domain consists of four helices instead of five with the extra helix (H0) in the mouse protein located at the N terminus (Fig. 4B). The helix boundaries are similar, but the angles between the helices are different. As a result, the XPCB domain of yRAD23 is less globular and the hydrophobic groove observed in mHR23B-XPCP is not present in the yeast protein. The main difference between the mouse and yeast XPCB domains is a 10° increase in the angle between helices H2 and H3 in the mouse protein. As a result, one of the “hydrophobic trails,” HT1 (25), forms a deep and wide hydrophobic groove into which H12 of the mPNGase core is deeply inserted. As seen already in the NMR structures (25), a flatter and smaller hydrophobic trail, HT2, is present on the opposite side of the hydrophobic groove (Fig. 4A). At the position corresponding to the smaller hydrophobic trail (HT2) of mHR23B-XPCP, a deep and extended cleft is formed in the yeast structure that partially buries the N-terminal helix (H1) of yPNGase (37). In general, the surface of yRAD23-XPCP is more hydrophobic, and the hydrophobic residues are more widely dispersed. The differences in the distribution of hydrophobic residues are significant, since the interactions between PNGase and HR23/RAD23 are mainly hydrophobic (26, 37), and therefore, the location of these residues determines the sites for protein-protein interactions. This distinction is also noteworthy with respect to the interaction between XPC and RAD23, which is also driven by hydrophobic interactions (23).

mPNGase-XPCB Domain Interaction—A comparison with the crystal structure of yPNGase (37) in complex with the XPCB domain of yRAD23 reveals that the interactions between mPNGase and the XPCB domain of mHR23B are drastically different (Fig. 5) despite the fact that the catalytic core and zinc binding domains of PNGase in the two complexes can be superimposed with an r.m.s. deviation of 1.27 Å for 224 CO atoms in the mPNGase core domain. As mentioned earlier, most of the secondary structure elements are conserved with the exception of the first two α-helices of the yeast enzyme, which are missing in the mouse ortholog. In addition, there is an extra helix in mPNGase after H10 that has been labeled H10′ (Fig. 2).

Both complexes form with similar affinities when full-length Rad23 is used as a ligand in ITC experiments (Kd of 22 nM for the mPNGase core and 24 nM for full-length yPNGase). Analyses of the structures show small differences in the buried surface areas (930 Å² in the mouse complex versus 1160 Å² in the yeast complex) and a different number of direct hydrogen bonds (4 in the mouse versus 3 in the yeast complex plus 2 extra salt bridges). The similar binding affinities between XPCB and PNGase in yeast and mouse are also reflected in their similar surface complementarities of 0.676 and 0.692, respectively (44). The XPCB domain of mHR23B interacts extensively with H12 of mPNGase and to a small degree also with H11. 20% of the solvent-accessible surface area of the XPCB domain is buried in the binding interface. The character of the interaction between the mPNGase core and mHR23B-XPCB is largely hydrophobic, with several hydrogen bonds formed at the periphery (Fig. 5A). Specifically, Leu-431, Ile-435, Val-436, Leu-438, Val-439, and Ile-442 of mPNGase form the hydrophobic face of H12, which deeply inserts into the hydrophobic trail HT1 of XPCB containing Leu-279, Phe-285, Ile-291, Leu-298, Leu-302, Ile-306, and Leu-313. The hydrophobic residues in PNGase are highly conserved in vertebrates (Table 3) and as the hydrophobic residues in the XPCB domain of HR23 from these organisms (supplemental Fig. S2).

In contrast to the mouse complex, the interaction between yPNGase and the XPCB domain of yeast Rad23 primarily involves the N-terminal helix H1 and only to a smaller degree, the C-terminal helix H12. Residues located in H1 are engaged in hydrophobic interactions and a series of hydrogen bonds with the XPCB domain (37), yet this helix is absent in the mouse protein as are the residues involved in the binding interactions (Table 3). The participation of the first helix of yeast PNGase was already predicted by biochemical studies (27) prior to structure determination of the yeast complex. H12 of yPNGase is significantly different from the corresponding helix in the mouse enzyme in terms of hydrophobicity and amphipathicity (Table 3). Specifically, five of the seven yPNGase residues in H12 that contact yRAD23 are not conserved in vertebrates (Table 3). Because the presence or absence of H1 of PNGase only slightly affects the binding affinity of the PNGase-HR23 interaction in mouse (see above), it is unlikely that the observed differences between the yeast and mouse complexes are due to conformational changes or multiple interaction modes. In addition, no gene duplication of PNGase in mammals and yeast has been observed, which rules out that the yeast and mouse PNGases are in fact paralogs. Instead, the differences in the interactions between PNGase and RAD23/HR23 in yeast and mammals may be explained by differences in the XPCB domains and the XPCB interaction motifs of PNGase. In fact, when the sequences of H12 from various species are compared, an evolutionary trend is observed; the XPCB-interacting residues in yeast and lower species are gradually exchanged with a different set in birds and mammals. More interestingly, this trend closely correlates with the emergence of the N- and C-terminal extensions of PNGase in higher species (Table 3), suggesting that H12 and the extensions co-evolve in such a way that the primary XPCB-interacting function migrates from H1 to H12.

Homologous Proteins Interacting Differently—The divergence between the yeast and mouse PNGase-Rad23/HR23B interfaces is all the more surprising since it has been shown that residues involved in protein interactions are more conserved than surface residues (45, 46). To the best of our knowledge, the differences in the XPCB/PNGase interaction between yeast and mouse represent one of the most striking changes in protein complex formation (47), and as such, the new interaction mode observed in the mouse complex is akin to a new protein fold (48). A detailed study of this system will therefore enrich our understanding of protein-protein interactions and molecular evolution.

Homologous proteins interacting differently can be classified into three classes. The first consists of homologous proteins with different quaternary structures, such as globins (49) and lectins (50). These proteins have conserved tertiary structures; however, utilizing different types of protein contacts, they can assemble into a variety of quaternary states, from dimers to oligomers containing as many as 144 subunits as in the case of Lumbricus terrestris erythrocruorin (51). The second category involves transient or reversible domain interactions. For example, the SH2 and SH3 domains of protein kinases can have at least three different types of interactions (47). Conformational changes such as those observed in molecular motors and molecular switches also fit in this category. In both of these classes, the interacting proteins usually either involve a single gene or a family of related genes; thus, they are not under very stringent evolutionary constraints. In the third and, in the context of this work, most interesting class are those complexes in which structurally homologous proteins interact differently with their
partners. A drastic example prior to this study is found in a bacterial chemotaxis pathway (52). In response to environmental changes, the sensor kinase CheA interacts and phosphorylates the response regulator CheY. Both CheA and CheY are structurally conserved in *E. coli* and *Thermotoga maritima*; however, the orientation of the two CheAs in their complexes with CheY differ by a 90° rotation (52). Even though the
interacting residues between CheA and CheY are different in the two complexes, the interaction sites on both CheA and CheY are largely preserved as reflected in an interaction iRMSD (47) of 11.6 Å.

In contrast the interaction between XPCB and PNGase is unique in several ways. First and most importantly, the interacting inter- faces on these proteins are different in both partners; yeast XPCB uses one side to interact with helix H1 of yeast PNGase, whereas mouse XPCB uses the opposite side to interact with helix H12 of mouse PNGase. As a consequence, the iRMSD (47) between the yeast and mouse complex is 28.0 Å, more than double the value of the CheA and CheY complexes. These differences are all the more remarkable since the interacting proteins are highly similar with 42% identity between yeast and mouse PNGase and 32% identity between the yeast and mouse XPCB domains. Second, despite these differences, both the yeast and mammalian complexes are characterized by equally high binding affinities and surface complementarities.
Furthermore, the different binding interactions of yeast and mouse XPCB with the respective PNGases are very likely also preserved in the RAD4/XPC proteins (see below). Thus, the apparently coupled evolution of the ternary RAD23/HR23-PNGase-RAD4/XPC system is intriguing and provides an excellent model system for studying molecular evolution in eukaryotes. Investigations into the XPCB/PNGase interactions, especially in species that are neither closely related to mouse nor yeast (Table 3, Fig. 6 and S2) such as Caenorhabditis elegans and D. melanogaster, will provide valuable insights into the evolution of protein–protein interactions.

Insights into the Interaction between HR23 and XPC—Based on the high identity sequence between the core of PNGase and XPC, this structure provides the first structural insights into the interaction between RAD4/XPC and the RAD23/HR23 proteins in the NER pathway. Structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified. In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified. In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.

In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.

In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.

In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.

In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.

In contrast to the previous proposal that XPC embraces a hydrophobic loop of HR23-XPCB (24), our study suggests that the amphipathic helix of XPC projects into the hydrophobic groove (hydrophobic trail 1) formed by H3 and H4 of HR23-XPCB. It has been observed that XPC by itself is sufficient for damage recognition in global genome repair but is more effective in complex with HR23 (14, 44). Lawrence and colleagues (98) identified a HR23 binding motif in human XPC encompassing residues 607–742 (53). Homology modeling and structural studies of XPC have been hampered because XPC is intrinsically unstable and behaves poorly in several protein expression systems (14, 24). Homology modeling with mPNGase identified a HR23 binding motif in human XPC encompassing residues 620–663 (Fig. 6) that could not be identified in the previous alignment (20). This segment is located in the HR23B binding region (amino acid 496–734) of human XPC (53). Homology modeling and secondary structure predictions suggested that the HR23 binding motif consists of two helices (Fig. 6 and data not shown), corresponding to H11 and H12 of mPNGase. The amphipathic character of the second predicted helix is easily identified.
47. Aloy, P., Ceulemans, H., Stark, A., and Russell, R. B. (2003) J. Mol. Biol. 332, 989–998
48. Aloy, P., and Russell, R. B. (2004) Nat. Biotechnol. 22, 1317–1321
49. Royer, W. E., Jr., Knapp, J. E., Strand, K., and Heaslet, H. A. (2001) Trends Biochem. Sci. 26, 297–304
50. Brinda, K. V., Surolia, A., and Vishveshwara, S. (2005) Biochem. J. 391, 1–15
51. Royer, W. E., Jr., Strand, K., van Heel, M., and Hendrickson, W. A. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7107–7111
52. Park, S. Y., Beel, B. D., Simon, M. I., Bilwes, A. M., and Crane, B. R. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 11646–11651
53. Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y., and Hanaoka, F. (2002) DNA Repair 1, 449–461
54. Guzder, S. N., Sung, P., Prakash, L., and Prakash, S. (1998) J. Biol. Chem. 273, 31541–31546
55. Sugasawa, K., Shimizu, Y., Iwai, S., and Hanaoka, F. (2002) DNA Repair 1, 95–107
56. Xie, Z., Liu, S., Zhang, Y., and Wang, Z. (2004) Nucleic Acids Res. 32, 5981–5990
57. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Proteins 12, 345–364