HHS Public Access

Author manuscript

Proteins. Author manuscript; available in PMC 2024 October 01.

Published in final edited form as:

Proteins. 2024 April ; 92(4): 554–566. doi:10.1002/prot.26647.

The N-terminal intrinsically disordered region of Ncb5or docks with the cytochrome b5 core to form a helical motif that is of ancient origin

David R. Benson[1], **Bin Deng**[4], **Maithri M. Kashipathy**[5], **Scott Lovell**[5], **Kevin P. Battaile**[7], **Anne Cooper**[6], **Philip Gao**[6], **Aron W. Fenton**[3], **Hao Zhu**[2,3,4]

[1]Department of Chemistry, University of Kansas, Lawrence, KS 66045, U.S.A.;

[2]Department of Clinical Laboratory Sciences, University of Kansas Medical Center, Kansas City, KS 66160, U.S.A.;

[3]Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, KS 66160, U.S.A.;

[4]Department of Physical Therapy and Rehabilitation Science, University of Kansas Medical Center, Kansas City, KS 66160, U.S.A.;

[5]Department of Protein Structure and X-ray Crystallography Laboratory, The University of Kansas, 2034 Becker Drive, Lawrence, KS 66047, USA;

[6]Department of Protein Production Group, The University of Kansas, 2034 Becker Drive, Lawrence, KS 66047, USA;

[7]Department of NYX, New York Structural Biology Center, Upton, NY, 11973, USA.

Abstract

Ncb5or (NADH cytochrome b5 oxidoreductase) is a cytosolic ferric reductase implicated in diabetes and neurological conditions. Ncb5or comprises cytochrome b5 (b5) and cytochrome b5 reductase (b5R) domains separated by a CHORD-Sgt1 (CS) linker domain. Ncb5or redox activity depends on proper interdomain interactions to mediate electron transfer from NADH or NADPH via FAD to heme. While full-length human Ncb5or has proven resistant to crystallization, we have succeeded in obtaining high-resolution atomic structures of the b5 domain and a construct containing the CS and b5R domains (CS/b5R). Ncb5or also contains an N-terminal intrinsically disordered region of 50 residues that has no homologs in other protein families in animals but features a distinctive, conserved $L^{34}MDWIRL^{40}$ motif also present in Root Lateral Root Formation protein (RLF) in rice and Increased Recombination Center 21 (IRC21) in baker's

Corresponding Authors: David R. Benson, Ph.D., 1567 Irving Hill Road, 1140 Gray-Little Hall, Lawrence, KS 66045, U.S.A., Phone: 785-864-4090, FAX: 785-864-5396, drb@ku.edu; Hao Zhu, Ph.D., 3901 Rainbow Blvd., MSN 4048G-Eaton, Kansas City, KS 66160, U.S.A., Phone: 913-588-2989, FAX: 913-588-5222, hzhu@kumc.edu.

Supporting Information Available.

This includes sequence alignments of Ncb5or in animals (N-terminal region and cytochrome b5 domain) and of the homologs in plants, fungi/yeasts, protists, and parasites, along with prediction plots of intrinsically disordered regions in human Ncb5or.

STRUCTURE DEPOSITION

Coordinates and structure factors for the root lateral formation protein (RLF) b5-domain structure were deposited to the Worldwide Protein Databank (wwPDB) with the accession code 8TGB.

yeast, all attaching to a b5 domain. After unsuccessful attempts at crystallizing a human Ncb5or construct comprising the N-terminal region naturally fused to the b5 domain, we were able to obtain a high-resolution atomic structure of a recombinant rice RLF construct corresponding to residues 25–129 of human Ncb5or (52% sequence identity; 74% similarity). The structure reveals Trp120 (corresponding to invariant Trp37 in Ncb5or) to be part of an 11-residue α-helix (S^{116}QMDWLKLTRT^{126}) packing against two of the four helices in the b5 domain that surround heme (α2 and α5). The Trp120 side chain forms a network of interactions with the side chains of four highly conserved residues corresponding to Tyr85 and Tyr88 (α2), Cys124 (α5), and Leu47 in Ncb5or. Circular dichroism (CD) measurements of human Ncb5or fragments further support a key role of Trp37 in nucleating the formation of the N-terminal helix, whose location in the N/b5 module suggests a role in regulating the function of this multidomain redox enzyme. This study revealed for the first time an ancient origin of a helical motif in the N/b5 module as reflected by its existence in a class of cytochrome b5 proteins from three kingdoms among eukaryotes.

Keywords

cytochrome b5; circular dichroism; crystallography; helix motif; intrinsically disordered region

INTRODUCTION

Ncb5or (NADH cytochrome b5 oxidoreductase, also called Cyb5R4, b5/b5R and b5+b5R) is a cytosolic ferric reductase implicated in diabetes, neurologic diseases and iron homeostasis.[1–5] Ncb5or is widely expressed in animal cells and localized to endoplasmic reticulum.[2, 6] Ncb5or contains two redox domains and one linkage domain that are homologous to cytochrome b5 (Cyb5), cytochrome b5 reductase (Cyb5R), and members of CHORD and SGT1 (CS) family, respectively.[6, 7] Ncb5or mediates electron transfer from NADH or NADPH to a redox substrate via FAD and heme cofactors in the b5R and b5 domains, respectively, and proper inter-domain interactions are essential to this process.[6] Although full-length Ncb5or from human has proven resistant to crystallization to date, we have succeeded in obtaining high-resolution structures of the b5 domain (PDB 3LF5) and a fragment containing the CS and b5R domains (CS/b5R) in complexes with NAD+ and NADP+ (PDB 6MV1 and 6MV2, respectively).[8, 9] Despite having the same general fold, the b5 domain exhibits substantially different heme ligation from that of microsomal cytochrome b5 (Cyb5A).[8] Similarly, the b5R domain differs from microsomal cytochrome b5 reductase (Cyb5R3) in having several multi-residue deletions and insertions that support extensive interactions with the CS domain and reflect a closer relationship to Cyb5R proteins from plants, fungi and protists than to Cyb5R3 from animals.[9]

The b5 domain of Ncb5or is preceded by a 50-residue N-terminal region that is unique among animal proteins but has homologs in Reduced Lateral Root Formation (RLF) proteins in plants and Increased Recombination Center 21 (IRC21) proteins in fungi. As in Ncb5or, the N-terminal region in the largely unexplored RLF and IRC21 proteins precedes a b5 domain. Herein we report the results of studies aimed at elucidating the structural properties of the N-terminal region of human Ncb5or both in isolation (hereafter N-term), and when natively fused to the b5 domain (hereafter N/b5). Circular dichroism (CD) spectroscopic

data revealed that N-term is intrinsically disordered but exhibits significant helical content in N/b5, with an invariant tryptophan residue (Trp37) playing a key role in inducing the helical structure. CD studies further suggested that Trp37, located in a highly conserved $L^{34}MDWIRL^{40}$ motif in mammalian orthologs, resides in a highly organized environment. While N/b5 constructs representing human Ncb5or failed to form crystals, a recombinant construct of rice RLF comprising residues K101-E218 crystallized readily. The resultant high-resolution X-ray crystallographic structure showed Trp^{120} (corresponding to Trp^{37} in Ncb5or) to be part of an 11-residue α-helix ($S^{116}QMDWLKLTRT^{126}$) packing against two of the four helices that surround heme (α2 and α5 in Ncb5or). The Trp^{120} side chain forms a network of interactions with the side chains of four highly conserved residues that are equivalent to Ncb5or residues Tyr^{85} and Tyr^{88} in α2, Cys^{124} in α5, and Leu^{47}. In addition to advancing our understanding of Ncb5or structure and drawing our attention to several surface-exposed side chains that may be functionally important, the studies also establish that the N/b5 unit of this multi-domain protein contains a helix motif that is of ancient origin and part of a well-conserved protein structural module.

MATERIALS AND METHODS

Molecular cloning and site-directed mutagenesis.

Human Ncb5or contains 521 amino acid residues that comprise the N-terminal region and the b5, b5R and CS domains as previously defined.[8] These segments of the protein contain 50, 113, 111, and 247 residues, respectively, that also include linker sequences in the middle two fragments (Figure 1A). Nomenclature of the constructs used herein is as follows: N/b5 (Met^{1} through Lys^{137}); N/b5-Δ21 (Gly^{22} through Lys^{137}); N/b5-Δ34 (Met^{35} through Lys^{137}); b5 (Lys^{51} through Lys^{137}); and N-term (Met^{1} through Leu^{50}). The cDNAs of all Ncb5or constructs (except N-term) and rice RLF (residues 101–218, XP_015647767) were expressed with no epitope tag (pET19b or pET22b). The cDNA of N-term was cloned into the pE-SUMOpro Kan vector (Life Sensors, Malvern, CA), which contains the 6XHis-SUMO gene immediately in front of the multiple cloning site previously described.[10] Missense point mutants, N/b5-LMAA (Leu^{34}Ala/Met^{35}Ala) and N/b5-W37A (Trp^{37}Ala) (Figure 1B), were generated using the QuikChange mutagenesis kit (Strategene, La Jolla, CA). All mutagenesis primers were synthesized by Integrated DNA Technology (Coralville, IA). All constructs were confirmed by DNA sequencing (ACGT, Inc., Wheeling, IL).

Protein preparation.

Except for N-term, all truncated constructs of human Ncb5or and rice RLF were expressed in *E.coli* BL21(DE3) or BL21(DE3)pLysS or BL21(DE3)pRARE cells at low temperature (<15°C) to increase solubility[2] for heme titration and purification as previously described.[8] Briefly, hemoproteins in whole cell lysate were reconstituted by hemin titration until saturation and then purified to homogeneity through ion-exchange chromatography followed by size exclusion on a Superdex 75 10/300 GL column. Ion-exchange chromatography was performed for b5 and N/b5-Δ34 on a Q HP column, N/b5 and variants (N/b5-W37A, N/b5-LMAA, and N/b5-Δ21) on a SP HP column, and rice RLF on both Q and SP HP columns. Full-length Ncb5or and its variant Ncb5or-Δ50 were prepared as previously described[2],

except that 0.5 mM IPTG was used for induction in TB media at 15°C. All purification steps were performed at 4°C. SDS-PAGE was used to estimate the purity of each protein used in spectroscopic analyses, kinetics studies and crystallization screening (all greater than 95%). All purified proteins were flash frozen in liquid nitrogen and stored as aliquots at −80°C until use, except for samples prepared for crystallization screening which were stored at 4°C. The size of each polypeptide product was confirmed by mass spectrometry. Expected sizes are N/b5, N/b5-W37A, N/b5-LMAA, ~15.6 kD; N/b5-Δ21 and RLF, ~ 13.6 kD; N/b5-Δ34, 11.9 kD; and Ncb5or-b5, 10.2 kD. The ratios of A_{413}/A_{280} for purified heme-containing constructs were as follows: b5 ≥ 4, N/b5 and variants, RLF ≥ 3.6. The heme content was determined by A_{413} for purified proteins where ~ 95% shown to contain heme as previously described.[2] As such, the heme concentration was used to represent protein concentrations in CD analyses (see below). Final protein yields (mg/L): 8 (b5), 5 (N/b5 and variants), ≤1 (RLF). A SUMOpro Expression System was used to prepare N-term.[10] Briefly, a 6XHis-SUMO-N-term fusion protein was expressed in BL21(DE3) cells and purified by Ni-NTA affinity resin (Qiagen, Valencia, CA). It was then cleaved by a 6XHis-SUMO protease[11] to release 6XHis-SUMO tag, both of which were removed by nickel resin. The purity and size of the N-term peptide were confirmed by SDS-PAGE and mass spectrometry (data not shown).

Spectroscopy.

UV/visible spectra were obtained on a Cary 100 Bio spectrophotometer (Varian). The concentrations of oxidized heme in all constructs were determined with ε_{413} = 130 $mM^{-1}cm^{-1}$ as previously described.[2] Circular dichroism (CD) analyses were performed on a JASCO-815 spectropolarimeter using protein concentrations of 4–10 μM. Far-UV (190–250 nm) spectra were recorded using a 1 mm cuvette, whereas a 1 cm cuvette was used for near-UV (250–350 nm) and visible (350–550 nm) CD spectra. All CD spectra are reported in units of molar ellipticity ([θ], $deg{\cdot}cm^2{\cdot}dmol^{-1}$).

RLF crystallization.

A purified RLF construct (Lys^{101}-Glu^{218}) was concentrated to 44.6 mg/mL in 20 mM Tris pH 8.0 for crystallization screening. All crystallization experiments were set up using an NT8 drop setting robot (Formulatrix Inc.) and UVXPO MRC (Molecular Dimensions) sitting drop vapor diffusion plates at 18°C. 100 nL of protein and 100 nL of crystallization solution were dispensed and equilibrated against 50 μL of the latter. Crystals approximately 300–500 μm long that displayed a needle morphology were observed within one day from the JCSG+ screen (Molecular Dimensions) condition A6 (20% (w/v) PEG 1000, 100 mM phosphate-citrate pH 4.2, 200 mM lithium sulfate). A cryoprotectant solution composed of 80% crystallization solution and 20% (w/v) glycerol was dispensed (2 μL) onto the drop, crystals were harvested with a cryoloop immediately and stored in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source beamline 17-ID using a Dectris Pilatus 6M pixel array detector.

Data collection and structure refinement.

Intensities were integrated using XDS[12, 13] via Autoproc[14] and the Laue class analysis and data scaling were performed with Aimless[15] which indicated that the crystals belong to the 2/*m* Laue class with possible space groups *P*2 or $P2_1$. Structure solution was conducted by molecular replacement with Phaser[16] using the Ncb5or b5-domain structure (PDB 3LF5[8]) as the search model. The top solution was obtained in the space group *P*2 and contained a dimer in the asymmetric unit. Initial refinement with Refmac[17] yielded R/R_{free} = 42.1%/42.5% and the model was improved by automated building with Arp/wARP.[18] Further refinement and manual model building were conducted with Phenix[19, 20] and Coot[21] respectively. TLS parameters[22, 23] were refined in the later stages to model anisotropic atomic displacement parameters. Disordered side chains were truncated to the point for which electron density could be observed. Structure validation was conducted with Molprobity[24] and figures were prepared using the CCP4MG package.[25] Polder omit electron density maps were calculated using Phenix.[26] Structure superposition was carried out with GESAMT.[27] Crystallographic data are described in Table S1, and the best data set was at 1.55 A resolution. Electron density maps (data not shown) for the heme molecules did not clearly indicate that the heme adopted two orientations related by rotation about the axis bisecting atoms CHA and CHC.

Protein sequence alignment.

Amino-acid sequences of human Ncb5or, rice RLF and yeast IRC21 were used as query in searches for homologous proteins against the current GenBank database with BLAST[28] to obtain initial alignments. Multiple sequence alignment across protein families was produced using ClustalW[29] followed by manual adjustments on the basis of structural alignments to minimize the number of gaps.

RESULTS AND DISCUSSION

The N-terminal region of Ncb5or is predicted to be intrinsically disordered.

The amino acid sequence of the N-terminal region of human Ncb5or is shown in Figure 1B, and an alignment of this 50-residue polypeptide with all known Ncb5or examples in animals is included in Figure S1. The alignment revealed an overall homology, which is weaker in the first 21 residues than the subsequent 29-residue stretch which is rich in Arg and Lys residues and contains five invariant residues (Lys25, Leu28, Gly31, Trp37 and Gly49). Trp37 is part of a highly conserved motif that has the sequence L^{34}MDWIRL[40] in mammalian orthologs.

The PONDR VL-XT program[30–32] predicts that the N-terminal region of Ncb5or is intrinsically disordered, in sharp contrast to the b5, b5R and CS domains (Figure S2A). Secondary structure prediction algorithms generally gave inconsistent predictions for the N-terminal region, but several of them indicated a preference for α-helical structure for the L^{34}MDWIRL[40] sequence. This included PROF[33] and PSIPRED[34, 35] (Figure S2B), as well as SPINE-X[36] (not shown). The latter two *ab initio* algorithms (*i.e.*, those not utilizing homology modeling) have been shown to yield the most accurate predictions.[37] AGADIR[38], an algorithm that predicts helical content of peptides on the basis of helix/coil transition

theory, also predicts some helical tendency for the L[34]MDWIRL[40] sequence, albeit with a population of less than 10% at pH 7 and 25°C (Figure S2C).

Interactions with the b5 domain induce helical structure in the N-terminal region.

To compare intrinsic structural properties of the N-terminal region of human Ncb5or with those it exhibits in the native protein, we performed CD spectroscopy studies of the 50-residue N-terminal polypeptide (hereafter N-term) and protein constructs representing the b5 domain with and without the N-terminal region attached (hereafter N/b5 and b5, respectively). Far-UV CD spectra of b5 and N/b5 are shown in Figure 2A (summarized in Table 1) and presented in molar ellipticity units to better reflect differences in total secondary structure content. The N/b5 spectrum differs from that of b5 in exhibiting enhanced negative ellipticity in the vicinity of 200 nm. The dominant feature in far-UV CD spectra of unstructured (random coil) polypeptides is a broad band of negative ellipticity centered near 200 nm.[39] The data in Figure 2A therefore suggests that N/b5 differs from b5 in having a larger number of disordered residues. This was confirmed by subtracting the b5 spectrum from the N/b5 spectrum to give the difference spectrum shown in Figure 2C. However, this difference spectrum also exhibits positive ellipticity near 190 nm and negative ellipticity in the vicinity of 220 nm, both suggestive of the presence of helical structure.[40] In contrast, the corresponding N-term spectrum (Figure 2C) is dominated by a negative band centered at 200 nm, revealing it to be largely unstructured. These results strongly suggest that the N-terminal region of Ncb5or is intrinsically disordered, consistent with the predictions described above, but that it adopts some helical structure when connected to the b5 domain.

The Trp37 side chain resides in a well-structured environment.

When located in highly asymmetric (*i.e.*, well-structured) environments, side chains of aromatic residues exhibit CD signals in overlapping regions of the near-UV (Phe, 250–270 nm; Tyr, 270–290 nm; Trp, 280–300 nm).[40] As shown in Figure 2B, b5 displays signals in the near-UV region (250–350 nm) attributable to the fact that it contains 2 Phe, 4 Tyr, and 2 Trp residues. Comparatively stronger signals, with significant fine structure, appear in the near-UV CD spectrum of N/b5 (Figure 2B). The N-terminal region contains only two aromatic residues, Phe9 and Trp37. Given that the strongest feature in the near-UV CD signature of N/b5 appears at 286 nm, we conclude that the major contributor is Trp37, which is part of the invariant mammalian Ncb5or L[34]MDWIRL[40] sequence noted above. In contrast, the corresponding spectrum of N-term exhibits no such signal at 286 nm (Figure 2D).

Aromatic residues in well-ordered regions of proteins often contribute to far-UV CD spectra as well. In the case of Trp, positive bands in the 225–250 nm region are a particularly common feature.[41, 42] Trp37 is therefore a likely source of the positive feature in the difference CD spectrum of N-term centered at 230 nm (Figure 2C), another conjecture which is supported by additional CD data described below. Notably, N-term exhibited no signals in the near-UV region (Figure 2D) and displays negative ellipticity in the vicinity of 230 nm (Figure 2C). The results described in this section allow us to conclude that interactions between the N-terminal region and the b5 domain induce secondary structure in

the former, and that this structural transition is accompanied by adoption of tertiary structure involving Trp37. This conclusion was further evaluated with truncation and point mutants of N/b5.

Trp37 plays a key role in the interactions between the N-terminal region and the b5 domain.

We generated two truncation mutants of N/b5 (Figure 1) and examined the effects of the truncations on CD spectra. The first mutant involved the removal of residues 1–21, the portion carrying the lowest conservation among animals, to yield N/b5-Δ21. Residues 1–34 were also removed to produce a domain fragment (N/b5-Δ34) analogous to the truncated Ncb5or that was initially cloned.[6]

Deleting the first 21 residues of N/b5 had no significant effect on the near-UV CD spectrum, or on the intensity of the positive far-UV CD band near 230 nm (Figures 2E–F, Table 1). The most notable effect on CD spectra caused by removing residues 1–21 was a decrease in negative ellipticity of the lowest wavelength far-UV band, along with a slight red shift of that band. These observations suggest that residues 1–21 are almost exclusively disordered and show that they play no role in the secondary and tertiary structure formation resulting from the interactions between the N-terminal region and the b5 core. They also support our conclusion that the near-UV CD signals and the positive far-UV feature near 230 nm arise from Trp37, with no contribution from Phe9.

Comparison of the far-UV CD difference spectra for N/b5-Δ21 and N/b5-Δ34 (Figure 2E, Table 1) shows that deleting the additional residues 22–34 decreased, but did not abolish, secondary structure content. In addition, Figure 2F (also Table 1) reveals diminished intensity for the near-UV CD signals attributable to Trp37 in N/b5-Δ34, but the signals in this region are much more like those in the spectrum of N/b5 than of b5. This suggests that Trp37 is involved in tertiary structure in both truncation mutants, but that removal of residues 22–34 causes subtle disruption of that tertiary structure.

The evidence obtained with the truncation mutants strongly indicates a key role for Trp37 in interactions between the N-terminal region and the b5 domain of Ncb5or. We therefore examined the effect of mutating Trp37 to Ala on CD spectra of N/b5. Subtracting the far-UV CD spectrum of b5 from that of N/b5-W37A yielded a difference spectrum similar to the spectrum of N-term (Figure 2G). Moreover, the W37A mutation in N/b5 resulted in loss of the near-UV CD features present in the N/b5 spectrum (Figure 2H). It can therefore be concluded that the W37A mutation in N/b5 abolishes both secondary and tertiary structure.

Two other conserved residues near Trp37 in Ncb5or (Leu34 and Met35) have hydrophobic side chains. To investigate the possibility that they may also be involved in N-terminal helix induction, we generated a L34A,M35A double mutant (N/b5-LMAA). The difference spectrum obtained by subtracting the far-UV CD spectrum of N/b5-LMAA from that of N/b5 exhibits modestly increased intensity of the negative band near 220 nm (Figure 2G, Table 1) and of the positive band near 190, suggesting a small increase in helix content. This may be attributable to the fact that Ala has a much higher helix propensity than do Leu and Met.[43] More notably, the LM→AA mutation does not significantly alter the shape

or intensity of the near-UV and far-UV CD features attributable to Trp37 (Figure 2G–H). Moreover, the CD Soret band signal for the LMAA mutant is virtually identical to that of Ncb5or-b5 (Figure 2H). These observations clearly show that neither Leu34 nor Met35 are essential for N-terminal helix induction, further suggesting a key role for Trp37 in seeding this structural motif.

The results described in this section indicate that residues spanning Met35 through Leu50 are sufficient to induce secondary structure (likely helical) and accompanying tertiary structure in the N-terminal region when it docks with the b5 core. The results also show that Trp37 plays an essential role in inducing this structural transition, suggesting a specific recognition site on the b5 domain surface. These results further indicate that the induced secondary structure is propagated toward the N-terminus but does not extend beyond Gly22. This additional secondary structure appears to stabilize the tertiary structure.

Plant and fungal proteins identified as potential models for N/b5.

Efforts to confirm the results of the CD studies described above via X-ray crystallographic studies of the N/b5 constructs were unfortunately not successful, prompting us to seek potential homologs that might crystallize more readily. BLAST searches using human Ncb5or residues 1–50 as the query revealed only a few other proteins that contain a homologous polypeptide, all from plants and fungi. In all cases the homologous polypeptide precedes a b5 domain having significant homology to that of Ncb5or and contains a Trp residue analogous to Trp37 of Ncb5or but no additional domains or regions (Figures S1, S3, S4). No structural characterization has been performed for these homologues, and only two have been characterized functionally: RLF protein from plants (e.g., rice, *Arabidopsis*, and glycine) and IRC21 protein from yeast (e.g., *Saccharomyces cerevisiae* and *S. pombe*). RLF appears to regulate lateral root formation independently of ARF7/19-mediated auxin signaling[44], whereas IRC21 functions in chromatin remodeling[45] and DNA repair[45–47] likely through its binding to lipid and protein phosphatase.[48, 49]

The N-terminal regions of plant RLF proteins are approximately 130-residues in length (Figure 3A–B) and contain a sequence spanning residues 115–121 (Q^{115}MDWLKL^{121}) that is strikingly similar to residues 34–40 of human Ncb5or which encompass Trp37 (L^{34}MDWIRL^{40}). The N-terminal regions of IRC21 proteins are approximately 117 residues in length (Figure 3A–B), and in that respect are more like Ncb5or than the plant RLF proteins. However, there is weaker homology for residues 61–67 (A^{61}LDWHSL^{67}) that correspond to residues 34–40 of Ncb5or). In addition, the sequence separating this signature from the b5 core is approximately 40 residues longer in yeast IRC21 protein than in human Ncb5or and rice RLF (Figure 3B), an insertion that may have arisen subsequent to divergence of the plant, fungi and animal kingdoms.[50] For these reasons we selected rice RLF protein for our structural studies, specifically residues 101–218, corresponding to residues 22–137 of human Ncb5or.

The b5 cores of Ncb5or and the rice RLF protein have identical folds.

The rice RLF construct crystallized readily as a non-crystallographic dimer and could be modeled from Phe^{111}-Glu^{218} in subunit A and Lys^{113}-Phe^{213} in subunit B. The fold

comprises six α-helices and four β-sheets (Figure 4A–B) that are arranged as follows: α1 (Ser^{116}-Thr^{126}), β1 (Arg^{140}-Ile^{142}), α2 (Leu^{144}-Lys^{148}), β2 (Trp^{157}-Leu^{160}), β3 (Arg^{163}-Asn^{166}), α3 (Ala^{168}-Phe^{173}), α4 (Val^{178}-Met^{182}), α5 (Thr^{190}-His^{197}), α6 (Phe^{202}-Leu^{205}), β4 (Leu^{210}-Leu^{214}). Note that the numbering is based on the NCBI reference sequence XP_015647767.1. The b5 core of RLF is identical to that of the Ncb5or b5 domain, with superposition yielding a root mean square deviation (RMSD) of 1.01 Å (80 residues) between Cα atoms (Figure 5).

The heme molecules are positioned within a cleft formed by helices α3-α6. It is important to note that the corresponding helices in the published structure of the Ncb5or b5 domain, and in structures of single domain b5 family members, are numbered α2-α5. In the comparison of the b5 and N/b5 structures we will refer to the relevant helices exclusively as α3-α6. Like all cytochrome b5 family members, the heme iron is coordinated by the side chains of two histidine residues (His174 and His197). His174 (His89 in Ncb5or) is located in the loop separating α3 and α4, while His197 (His112 in Ncb5or) constitutes the C-terminal residue in helix α5. As observed for His89 and His112 in the crystal structure of the Ncb5or b5 domain, the imidazolyl moieties of His174 and His197 in the rice RLF structure are nearly orthogonal to one another (Figure 6B). The angle between the mean planes defined by the indole rings is 77.6° (subunit A) and 77.2° (subunit B) as calculated using Mercury.[51] The corresponding angles for His89 and His112 in the two subunits of the Ncb5or b5 domain structure (3LF5) are 83.2° and 81.3°.[8] This structural feature distinguishes these proteins (and likely IRC21 as well), from the better known microsomal and mitochondrial isoforms of cytochrome b5, in which angles between the His ligand side chains are closer to 20°.[8]

We reported that the b5 core of Ncb5or differs from microsomal and mitochondrial isoforms of cytochrome b5 in having a strictly conserved tryptophan residue (Trp114) at the mouth of the heme binding pocket, two residues removed from heme ligand His112.[8] The intervening residue, Arg113, is invariant among mammalian Ncb5or isoforms. In the crystal structure of the b5 domain of Ncb5or the side chain of Trp114 is located on the protein exterior and is substantially solvent-exposed (Figure 6A), with its only inter-protein interactions involving the side chain of adjacent residue Arg113. The rice RLF residues corresponding to the Ncb5or His^{112}ArgTrp^{114} sequence (His197, Ala198, and Trp199) are strictly conserved among known members of this protein family. In the RLF crystal structure the Trp199 side chain projects into solvent, but as shown in Figure 6A its side chain conformation differs from that of Trp114 in the Ncb5or b5 domain structure, enabling hydrogen bonding between its Nε-H and one of the heme propionate groups. Manually changing the side chain torsional angles of Trp114 in the Ncb5or b5 domain crystal structure ($\chi^\circ = -53.9$; $\chi 2 = 115.7°$) to match those of Trp199 in the new rice RLF structure ($\chi^\circ = 58.9$; $\chi 2 = 85.4°$), using PyMol[52], showed that Trp114 could form an analogous hydrogen bond with heme without introducing unfavorable steric interactions. Given that this solvent exposed Trp residue is strictly conserved among all known Ncb5or and RLF proteins (and in IRC21 proteins as well), its ability to form a hydrogen bond with heme suggests that it serves an essential functional role.

Yet another distinguishing feature shared by the b5 cores of Ncb5or and rice RLF is an irregular helix (α6) in the heme-binding pocket, featuring a central kink that leaves Leu205 (Met120 in Ncb5or) without an intra-helical hydrogen bond. As will be noted in the following section, this kink plays an important role in interactions with the N-terminal region.

While conducting the studies reported herein, we became aware of reports of b5 family members in *Giardia* and some other parasitic organisms with strong homology to the b5 domains of Ncb5or, RLF and IRC21 proteins. As shown by some representative examples in Figure S5, these proteins feature an N-terminal region, albeit generally shorter than those in Ncb5or, RLF and IRC21 and without a Trp residue analogous to Trp37 in Ncb5or.[53] Many of them have also maintained a surface tryptophan residue analogous to Trp114 of Ncb5or. The *Giardia* proteins were shown to exhibit redox potentials[54] similar to those determined for Ncb5or,[2] which are considerably more negative than for microsomal b5s. The *Giardia* proteins and Ncb5or also share EPR spectroscopic signatures that are characteristic of orthogonal His ligands.[54] It is reasonable to conclude that these proteins have evolved from the common N/b5 ancestor of Ncb5or, RLF, and IRC21, perhaps with divergent functions.[53, 54] This notion is supported by the observation of cytochrome b5 proteins in protists (*Trypanosoma* and *Dictyostelium*) that share the same b5 core[53] and the N-terminal motif as that in Ncb5or, including the Trp114 and Trp37 residues, respectively (Figure S5).

The rice RLF structure confirms N-terminal helix formation involving the Trp120 side chain.

Trp120 in RLF (corresponding to Trp37 in Ncb5or) is part of an 11-residue α-helix, designated α1 (S^{116}QMDWLKLTRT^{126}), that packs against the side of the heme-binding pocket defined by helices α3 and α6 (Figure 7). In this and subsequent sections, all corresponding residues in Ncb5or are listed in parentheses. A stretch of approximately 13 residues connects α1 to β1, the first element of defined secondary structure in the "classic" b5 core. The corresponding stretch in Ncb5or is two residues shorter, with the gap appearing in the vicinity of β1. There are multiple conserved residues in the RLF and Ncb5or loops, and in the RLF structure several of them exhibit specific backbone and side chain interactions both within the loop and with the b5 core. These include two three-residue turns featuring a single backbone hydrogen bond, the first between Asp129 and Gly132 (Asp46 and Gly49), and the second between Leu133 and Gln136 (Leu 50 and Arg53). The Asp129 (Asp46) side chain in the first turn extends this motif via hydrogen bonds to the backbone amide N-H groups of Leu133 and Lys134 (Leu50 and Lys51), residues which can be considered to represent the transition between N-term and the b5 core. The side chain of Leu133 (Leu50) is buried and engages in hydrophobic interactions with several nearby residues, namely Cβ of Ala131 (Thr48), Cβ of Gln136 (Arg53), and Cα of Asn138 (Ile55). It also packs against the side chain of Val211 (Val126) which is in strand β4 of the b5 core. All the residues in this hydrogen bonded array other than solvent exposed Gln136 (Arg53) are conserved in RLF and Ncb5or proteins. Likely because of these interactions, this polypeptide segment is well-ordered. Packing between the N-terminal helix and the b5 core involves primarily hydrophobic interactions and does not noticeably impact the b5 core fold. Consistent with the results of far-UV CD studies described above, the Trp120 side chain is situated in an environment with well-defined tertiary structure that includes

hydrophobic packing interactions with the side chains of four residues in α3 and α6, each corresponding to a residue that is highly conserved in Ncb5or: Tyr170 and Phe173 in α3 (Tyr85 and Tyr88); and Leu206 and Cys209 in α6 (Leu121 and Cys124). The Trp120 side chain engages in additional hydrophobic interactions with the nearby side chains of Leu121 and Thr124 in α1 (Ile38 and Thr41) and with the side chain of Leu130 (Leu47), located in the first 3-residue turn in the 17-residue loop connecting α1 to α2.

The Trp120 (Trp37) side chain has one additional contact, a hydrogen bond between its N-H group and the backbone carbonyl of Leu205 in α6 (Met120). As noted in the previous section, the backbone carbonyl of that residue does not form an intra-helix H-bond due to a kink in α6. The present results suggest that the α6 kink may have evolved to serve a distinct functional role in these proteins.

The only other polar contacts between α1 and the b5 core comprise a network of hydrogen bonds involving the side chains of the first two α1 residues Ser116 and Gln117 (Ser33 and Leu34), the phenolic side chain of Tyr170 in α3 (Tyr85), and the backbone carbonyl of Phe173 in α3 (Tyr88). Gln117 is the hub of this network in the rice RLF structure, its polar side chain amide group forming hydrogen bonds with all the other participating residues while also making hydrophobic contact with a heme methyl group via its gamma carbon. The apolar side chain of the corresponding Ncb5or residue (Leu34) can be reasonably expected to make analogous (and potentially more extensive) van der Waals contact with heme but is unable to engage in a hydrogen bonding network that would involve Ser33, Tyr85 and Tyr88. Its invariance among mammalian Ncb5or proteins suggests it is important, despite the observation from CD studies of the LMAA double mutant that it does not play a key role in N-terminal helix induction.

Tyr170 is perhaps the most notable residue in the preceding comparison, given that in the rice RLF structure its polar phenolic OH group makes van der Waals contact with one of the heme *meso* positions. Tyr85 in the structure of human Ncb5or b5 core (PDB 3LF5) exhibits identical interactions with heme. This counterintuitive positioning of a polar group suggests a potential functional role. An intriguing possibility is that oxidation of ferrous heme to ferric heme during reduction of the downstream substrate is accompanied by deprotonation of the phenolic group. This could conceivably facilitate a coupled electron/proton transfer process during substrate reduction. Alternatively, it could allow for electrostatic stabilization of the b5 domain when heme is in the ferric oxidation state and bears a formal positive charge.

AlphaFold predicts an ST N-cap role for Ser116.

The AlphaFold artificial intelligence system[55] successfully predicted N-terminal helices corresponding to α1 in both rice RLF (AF-B9FWA5-F1) and human Ncb5or (AF-Q7L1T6-F1), albeit with distinct and potentially significant differences (Figure S6A). Plots of Cα and overall RMSD differences between the predicted and experimental rice RLF structures reveal greatest similarity for α1 in the vicinity of Trp120 (Figure S6B). AlphaFold even accurately predicted the H-bonding interaction between the side chain of Trp120 and the backbone carbonyl of Leu205 in α6. Similarity is weaker at the N- and C-termini of α1 both in terms of backbone RMSD and of interactions involving conserved residues. For example,

in both molecules in the asymmetric unit of the RLF crystal structure the α-carbonyl of Ser116 forms an *i*, *i* +4 hydrogen bond with the α-amide of Trp120 (Figure S7A). And as noted above, the Ser116 side chain hydroxyl group forms a hydrogen bond with the amide side chain of Gln34, part of a network that also involves hydrogen bonds between the Gln34 side chain, the Tyr170 side chain, and the Phe173 backbone carbonyl. In the predicted structure (Figure S7B), Ser116 acts as an N-terminal helix cap through a hydrogen bond between its side chain hydroxyl group and the backbone amide of Asp119, a motif commonly referred to as an ST N-cap.[56] AlphaFold likewise predicted an N-cap role for Ser33 in Ncb5or. We consider these predictions to be quite compelling, given the ubiquity of serine N-cap interactions in proteins and the conserved nature of serine in that location of RLF and Ncb5or proteins. It is of more than glancing interest given that the disposition of Ser116 will affect preceding residues. Indeed, the Ser116 N-cap in the predicted structure of the rice RLF protein enables the side chain of the preceding residue, Tyr115, to form an edge-to-face interaction with the side chain of Phe173, which itself is edge-to-face with the Trp120 side chain. In the crystal structure the side chain of Tyr115 projects into solvent. Crystal packing may contribute to this difference in structure at the N-terminus of α1.

Another significant difference between the predicted and experimental structures of the rice RLF protein seen in the RMSD plots (Figure S6B) occurs in the vicinity of Pro128, the second residue beyond α1. It arises due to substantially different backbone torsional angles for Pro128 and Asp129. Similar divergence in backbone conformations in this region is observed between the predicted structures of RLF (Pro128, Asp129) and Ncb5or (Gly44, Lys45), suggestive of significant local conformational flexibility.

Several conserved residues in the N-terminal helix have solvent-exposed side chains.

The side chains of several other highly conserved residues in the N-terminal helix project into solvent (Figure 8), including three in the vicinity of Trp120: Met118, Asp119, and Lys122 (corresponding to Met35, Asp36, and Arg39 in Ncb5or). Particularly notable in this group is Met118 (Met35). The thioether side chain of methionine is considered nonpolar, and indeed methionine is often subject to evolutionarily conservative replacement by residues with aliphatic side chains via single codon mutations (leucine, isoleucine, and valine). Only a few examples of such a mutation, specifically Met to Leu, exist among known Ncb5or proteins, and only in non-mammalian organisms (Figure S1). Methionine differs from the aliphatic amino acids it often replaces in that its thioether functional group can serve functional roles (for example, as a ligand to heme iron in cytochrome *c*). The highly conserved nature of Met35 in Ncb5or, considered in the context of its solvent-exposed location, is suggestive of an essential functional role.

CONCLUSION

This study represents the first *in vitro* characterization of the N-terminal region of Ncb5or, a polypeptide segment not found in any other animal proteins, but which shares a common ancestor with RLF proteins in plants and IRC21 proteins in fungi. The N-terminal regions in all three of these protein families precede highly homologous cytochrome b5 domains and share sequence similarities that suggest that the results we have obtained using N/b5

constructs derived from human Ncb5or and rice RLF protein will pertain to the IRC21 proteins as well. Specifically, we have shown that the N-terminal region is intrinsically disordered but that a 11-residue motif adopts a helical conformation when natively attached to the b5 core, a structural transition that is nucleated by a strictly conserved tryptophane residue (Trp37 in human Ncb5or). Such docking-induced folding is a common theme for intrinsically disordered regions of mammalian proteins that are often involved in DNA or RNA binding, substrate binding, and protein-protein interactions[57–60], of which the specificity is generally mediated by motif structures.[61, 62] In this context, it is worth investigating whether IRC21 utilizes the helical motif in the N/b5 module to activate protein phosphatase 2A for DNA damage response in yeasts.[49] Intrinsically disordered regions appear to be enriched in disease-related proteins, such as the N-terminal transactivation domain of tumor suppressor p53,[63] alpha-synuclein in Parkinson's disease[64], tau protein in Alzheimer's disease[65], and in a number of proteins associated with cardiovascular disease.[66] Mutation of the intrinsically disordered region of carboxyl ester lipase in pancreatic acinar cells can cause lean diabetes, likely as a result of protein mis-folding.[67] Functional contribution of the N-terminal intrinsically disordered region and the helical motif to Ncb5or's cellular function and related disease pathways is being investigated.

The presence of the helical motif within the N/b5 module in three kingdoms of eukaryotes that diverged about 1.6 billion years ago[50, 68] suggests an ancient heritage of Ncb5or. This is consistent with our previous observation of closer similarities of its b5R domain to monomeric b5R proteins from plants, fungi and protists than to mammalian Cyb5R3 proteins.[9] The nucleated N-terminal helix lies between two of the four helices that comprise the heme binding pocket and is connected to what we have previously referred to as the b5 core by a well-structured polypeptide. The N-terminal region in Ncb5or and the plant and fungal proteins could conceivably comprise an integral part of their b5 cores, which would thus be expanded relative to other modern cytochrome b5 superfamily members. The N-terminal regions of these proteins could alternatively function to regulate their enzymatic functions through reversible docking, an intriguing possibility that will be the focus of future mechanistic studies. Suggestive of this latter scenario is the observation that the 13 residues separating the N-terminal helix α1 from β1 in rice RLF protein exhibit few interactions with the "classic" portion of the b5 core, in contrast to the 11-residue helix that is nucleated by the Trp37-dependent docking interactions. The low homology in the N-terminal region of cytochrome b5 proteins of parasites, including the absence of the helical motif, is likely a result of diminishing selection pressure for the above-mentioned motif-mediated function due to the unique parasite-host interaction and pathophysiology.

Future studies will also probe potential structural and functional roles of other strictly conserved residues emphasized above. These include Tyr85, Try88 and Cys120 in the b5 core, whose side chains pack against the side chain of Trp37, as well as Met35 in the α1 and Trp114 which is located two residues from heme ligand His112. The side chains of these latter residues project conspicuously into solvent, locations that would subject them to random mutation were they not essential for protein function, especially since they are encoded by a single codon.[69] Functional roles could include modulating heme reduction by the b5R domain or reduction by heme of the still unknown substrate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Christopher D. Lima (Sloan-Kettering Institute, New York) for the SUMO protease as a gift and the anonymous reviewer of the journal for the constructive comments.

Funding Sources:

This work was supported by the National Institute of General Medical Sciences, National Institutes of Health (P30GM110761; R.P. Hanzlik, University of Kansas). Additional support came from the University of Kansas Medical Center School of Health Professions (H.Z.) and University of Kansas College of Liberal Arts and Sciences (D.R.B.). Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Hauptman-Woodward Medical Research Institute. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

ABBREVIATIONS

CD	circular dichroism
Cyb5A	type A cytochrome b5
Cyb5R3	NADH cytochrome b5 reductase isoform 3
IRC21	Increased Recombination Center 21
NADH	nicotinamide adenine dinucleotide (reduced)
Ncb5or	NADH cytochrome b5 oxidoreductase
RLF	reduced lateral root formation

REFERENCES

[1]. Xie J, Zhu H, Larade K, Ladoux A, Seguritan A, Chu M, Ito S, Bronson RT, Leiter EH, Zhang CY, Rosen ED, Bunn HF: Absence of a reductase, NCB5OR, causes insulin-deficient diabetes. Proc Natl Acad Sci U S A 2004, 101:10750–5. [PubMed: 15247412]

[2]. Zhu H, Larade K, Jackson TA, Xie J, Ladoux A, Acker H, Berchner-Pfannschmidt U, Fandrey J, Cross AR, Lukat-Rodgers GS, Rodgers KR, Bunn HF: NCB5OR is a novel soluble NAD(P)H reductase localized in the endoplasmic reticulum. J Biol Chem 2004, 279:30316–25. [PubMed: 15131110]

[3]. Wang WF, Guo Y, Xu M, Huang H-H, Novikova L, Larade K, Jiang Z-G, Thayer TC, Frontera JR, Aires D, Ding H, Turk J, Mathews CE, Bunn HF, Stehno-Bittel L, Zhu H: Development of diabetes in Ncb5or-null mice is associated with manifestations of endoplasmic reticulum and oxidative stress in beta cells. Biochimica et Biophysica Acta 2011, 1812:1532–41. [PubMed: 21839170]

[4]. Stroh MA, Winter MK, Swerdlow RH, McCarson KE, Zhu H: Loss of NCB5OR in the cerebellum disturbs iron pathways, potentiates behavioral abnormalities, and exacerbates harmaline-induced tremor in mice. Metabolic Brain Disease 2016, 31:951–64. [PubMed: 27188291]

[5]. Stroh MA, Winter MK, McCarson KE, Thyfault JP, Zhu H: NCB5OR Deficiency in the Cerebellum and Midbrain Leads to Dehydration and Alterations in Thirst Response, Fasted Feeding Behavior, and Voluntary Exercise in Mice. Cerebellum 2018, 17:152–64. [PubMed: 28887630]

[6]. Zhu H, Qiu H, Yoon HW, Huang S, Bunn HF: Identification of a cytochrome b-type NAD(P)H oxidoreductase ubiquitously expressed in human cells. Proc Natl Acad Sci U S A 1999, 96:14742–7. [PubMed: 10611283]

[7]. Garcia-Ranea JA, Mirey G, Camonis J, Valencia A: p23 and HSP20/alpha-crystallin proteins define a conserved sequence domain present in other eukaryotic protein families. FEBS Lett 2002, 529:162–7. [PubMed: 12372593]

[8]. Deng B, Parthasarathy S, Wang W, Gibney BR, Battaile KP, Lovell SW, Benson DR, Zhu H: Study of the individual cytochrome b5 and cytochrome b5 reductase domains of Ncb5or reveals a unique heme pocket and a possible role of the CS domain. J Biol Chem 2010, 285:30181–91. [PubMed: 20630863]

[9]. Benson DR, Lovell S, Mehzabeen N, Galeva N, Cooper A, Gao P, Battaile KP, Zhu H: Crystal structures of the naturally fused CS and cytochrome b5 reductase (b5R) domains of Ncb5or reveal an expanded CS fold, extensive CS-b5R interactions and productive binding of the NAD(P)(+) nicotinamide ring. Acta Crystallogr D Struct Biol 2019, 75:628–38. [PubMed: 31282472]

[10]. Prasannan CB, Tang Q, Fenton AW: Allosteric regulation of human liver pyruvate kinase by peptides that mimic the phosphorylated/dephosphorylated N-terminus. Methods Mol Biol 2012, 796:335–49. [PubMed: 22052499]

[11]. Reverter D, Lima CD: Preparation of SUMO proteases and kinetic analysis using endogenous substrates. Methods Mol Biol 2009, 497:225–39. [PubMed: 19107421]

[12]. Kabsch W: Automatic indexing of rotation diffraction patterns. J Applied Crystallography 1988, 21:67–72.

[13]. Kabsch W: Xds. Acta Crystallogr D Biol Crystallogr 2010, 66:125–32. [PubMed: 20124692]

[14]. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T, Bricogne G: Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr 2011, 67:293–302. [PubMed: 21460447]

[15]. Evans PR: An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr 2011, 67:282–92. [PubMed: 21460446]

[16]. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: *Phaser* crystallographic software. J Appl Cryst 2007, 40:658–74. [PubMed: 19461840]

[17]. Murshudov GN, Vagin AA, Dodson EJ: Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997, 53:240–55. [PubMed: 15299926]

[18]. Langer G, Cohen SX, Lamzin VS, Perrakis A: Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 2008, 3:1171–9. [PubMed: 18600222]

[19]. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD: Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 2012, 68:352–67. [PubMed: 22505256]

[20]. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC: PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 2002, 58:1948–54. [PubMed: 12393927]

[21]. Emsley P, Lohkamp B, Scott WG, Cowtan K: Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010, 66:486–501. [PubMed: 20383002]

[22]. Painter J, Merritt EA: Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 2006, 62:439–50. [PubMed: 16552146]

[23]. Winn MD, Isupov MN, Murshudov GN: Use of TLS Parameters to Model Anisotropic Displacements in Macromolecular Refinement. Acta Crystallogr D Biol Crystallography 2001, 57:122–33.

[24]. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010, 66:12–21. [PubMed: 20057044]

[25]. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M: Developments in the CCP4 molecular-graphics project. Acta Crystallogr D Biol Crystallogr 2004, 60:2288–94. [PubMed: 15572783]

[26]. Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD: Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr D Struct Biol 2017, 73:148–57. [PubMed: 28177311]

[27]. Krissinel E: Enhanced fold recognition using efficient short fragment clustering. J Mol Biochem 2012, 1:76–85. [PubMed: 27882309]

[28]. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389–402. [PubMed: 9254694]

[29]. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R: Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 2013, 41:W597–600. [PubMed: 23671338]

[30]. Li X, Romero P, Rani M, Dunker AK, Obradovic Z: Predicting Protein Disorder for N-, C-, and Internal Regions. Genome Inform Ser Workshop Genome Inform 1999, 10:30–40.

[31]. Romero, Obradovic, Dunker K: Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome Inform Ser Workshop Genome Inform 1997, 8:110–24.

[32]. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK: Sequence complexity of disordered protein. Proteins 2001, 42:38–48. [PubMed: 11093259]

[33]. Rost B, Sander C: Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993, 232:584–99. [PubMed: 8345525]

[34]. Jones DT: Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999, 292:195–202. [PubMed: 10493868]

[35]. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT: Protein annotation and modelling servers at University College London. Nucleic Acids Res 2010, 38:W563–8. [PubMed: 20507913]

[36]. Faraggi E, Yang Y, Zhang S, Zhou Y: Predicting continuous local structure and the effect of its substitution for secondary structure in fragment-free protein structure prediction. Structure 2009, 17:1515–27. [PubMed: 19913486]

[37]. Zhang H, Zhang T, Chen K, Kedarisetti KD, Mizianty MJ, Bao Q, Stach W, Kurgan L: Critical assessment of high-throughput standalone methods for secondary structure prediction. Brief Bioinform 2011, 12:672–88. [PubMed: 21252072]

[38]. Munoz V, Serrano L: Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol 1994, 1:399–409. [PubMed: 7664054]

[39]. Greenfield N, Fasman GD: Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 1969, 8:4108–16. [PubMed: 5346390]

[40]. Kelly SM, Jess TJ, Price NC: How to study proteins by circular dichroism. Biochim Biophys Acta 2005, 1751:119–39. [PubMed: 16027053]

[41]. Woody RW: Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins. Biopolymers 1978, 17:1451–67.

[42]. Freskgard PO, Martensson LG, Jonasson P, Jonsson BH, Carlsson U: Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II. Biochemistry 1994, 33:14281–8. [PubMed: 7947839]

[43]. Pace CN, Scholtz JM: A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 1998, 75:422–7. [PubMed: 9649402]

[44]. Ikeyama Y, Tasaka M, Fukaki H: RLF, a cytochrome b(5)-like heme/steroid binding domain protein, controls lateral root formation independently of ARF7/19-mediated auxin signaling in Arabidopsis thaliana. Plant J 2010, 62:865–75. [PubMed: 20230485]

[45]. Alvaro D, Lisby M, Rothstein R: Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. PLoS Genet 2007, 3:e228. [PubMed: 18085829]

[46]. Lee W, St Onge RP, Proctor M, Flaherty P, Jordan MI, Arkin AP, Davis RW, Nislow C, Giaever G: Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. PLoS Genet 2005, 1:e24. [PubMed: 16121259]

[47]. Guenole A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, Ideker T, van Attikum H: Dissection of DNA damage responses using multiconditional genetic interaction maps. Mol Cell 2013, 49:346–58. [PubMed: 23273983]

[48]. Gallego O, Betts MJ, Gvozdenovic-Jeremic J, Maeda K, Matetzki C, Aguilar-Gurrieri C, Beltran-Alvarez P, Bonn S, Fernandez-Tornero C, Jensen LJ, Kuhn M, Trott J, Rybin V, Muller CW, Bork P, Kaksonen M, Russell RB, Gavin AC: A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. Mol Syst Biol 2010, 6:430. [PubMed: 21119626]

[49]. Ferrari E, Bruhn C, Peretti M, Cassani C, Carotenuto WV, Elgendy M, Shubassi G, Lucca C, Bermejo R, Varasi M, Minucci S, Longhese MP, Foiani M: PP2A Controls Genome Integrity by Integrating Nutrient-Sensing and Metabolic Pathways with the DNA Damage Response. Mol Cell 2017, 67:266–81 e4. [PubMed: 28648781]

[50]. Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB: Molecular evidence for the early colonization of land by fungi and plants. Science 2001, 293:1129–33. [PubMed: 11498589]

[51]. Macrae CF, Bruno IJ, Chisholm JA, Edgington PR, McCabe P, Pidcock E, Rodriguez-Monge L, Taylor R, van de Streek J, Wood PA: Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. Journal of Applied Crystallography 2008, 41:466–70.

[52]. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[53]. Pyrih J, Harant K, Martincova E, Sutak R, Lesuisse E, Hrdy I, Tachezy J: Giardia intestinalis Incorporates Heme into Cytosolic Cytochrome b(5). Eukaryotic Cell 2014, 13:231–9. [PubMed: 24297440]

[54]. Pazdzior R, Yang ZA, Mesbahuddin MS, Yee J, van der Est A, Rafferty S: Low reduction potential cytochrome b5 isotypes of Giardia intestinalis. Exp Parasitol 2015, 157:197–201. [PubMed: 26299244]

[55]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D: Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596:583–9. [PubMed: 34265844]

[56]. Wan WY, Milner-White EJ: A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at α-helical N termini and in other situations. Journal of Molecular Biology 1999, 286:1651–62. [PubMed: 10064721]

[57]. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z: Intrinsic disorder and protein function. Biochemistry 2002, 41:6573–82. [PubMed: 12022860]

[58]. Dyson HJ, Wright PE: Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005, 6:197–208. [PubMed: 15738986]

[59]. Sugase K, Dyson HJ, Wright PE: Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007, 447:1021–5. [PubMed: 17522630]

[60]. Liu Y, Matthews KS, Bondos SE: Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. J Biol Chem 2008, 283:20874–87. [PubMed: 18508761]

[61]. Wang X, Bajaj R, Bollen M, Peti W, Page R: Expanding the PP2A Interactome by Defining a B56-Specific SLiM. Structure 2016, 24:2174–81. [PubMed: 27998540]

[62]. Singh JP, Li Y, Chen YY, Hsu SD, Page R, Peti W, Meng TC: The catalytic activity of TCPTP is auto-regulated by its intrinsically disordered tail and activated by Integrin alpha-1. Nat Commun 2022, 13:94. [PubMed: 35013194]

[63]. Joerger AC, Fersht AR: Structural biology of the tumor suppressor p53. Annu Rev Biochem 2008, 77:557–82. [PubMed: 18410249]

[64]. Breydo L, Wu JW, Uversky VN: alpha-Synuclein misfolding and Parkinson's disease. Biochim Biophys Acta 2012, 1822:261–85. [PubMed: 22024360]

[65]. Uversky VN, Oldfield CJ, Dunker AK: Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008, 37:215–46. [PubMed: 18573080]

[66]. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN: Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 2006, 45:10448–60. [PubMed: 16939197]

[67]. Johansson BB, Torsvik J, Bjorkhaug L, Vesterhus M, Ragvin A, Tjora E, Fjeld K, Hoem D, Johansson S, Raeder H, Lindquist S, Hernell O, Cnop M, Saraste J, Flatmark T, Molven A, Njolstad PR: Diabetes and pancreatic exocrine dysfunction due to mutations in the carboxyl ester lipase gene-maturity onset diabetes of the young (CEL-MODY): a protein misfolding disease. J Biol Chem 2011, 286:34593–605. [PubMed: 21784842]

[68]. Woese CR: Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A 2000, 97:8392–6. [PubMed: 10900003]

[69]. Barik S: The Uniqueness of Tryptophan in Biology: Properties, Metabolism, Interactions and Localization in Proteins. Int J Mol Sci 2020, 21.

[70]. Evans P: Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 2006, 62:72–82. [PubMed: 16369096]

[71]. Diederichs K, Karplus PA: Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 1997, 4:269–75. [PubMed: 9095194]

[72]. Weiss MS: Global indicators of X-ray data quality. Journal of Applied Crystallography 2001, 34.

[73]. Karplus PA, Diederichs K: Linking crystallographic model and data quality. Science 2012, 336:1030–3. [PubMed: 22628654]

[74]. Evans P: Biochemistry. Resolving some old problems in protein crystallography. Science 2012, 336:986–7. [PubMed: 22628641]

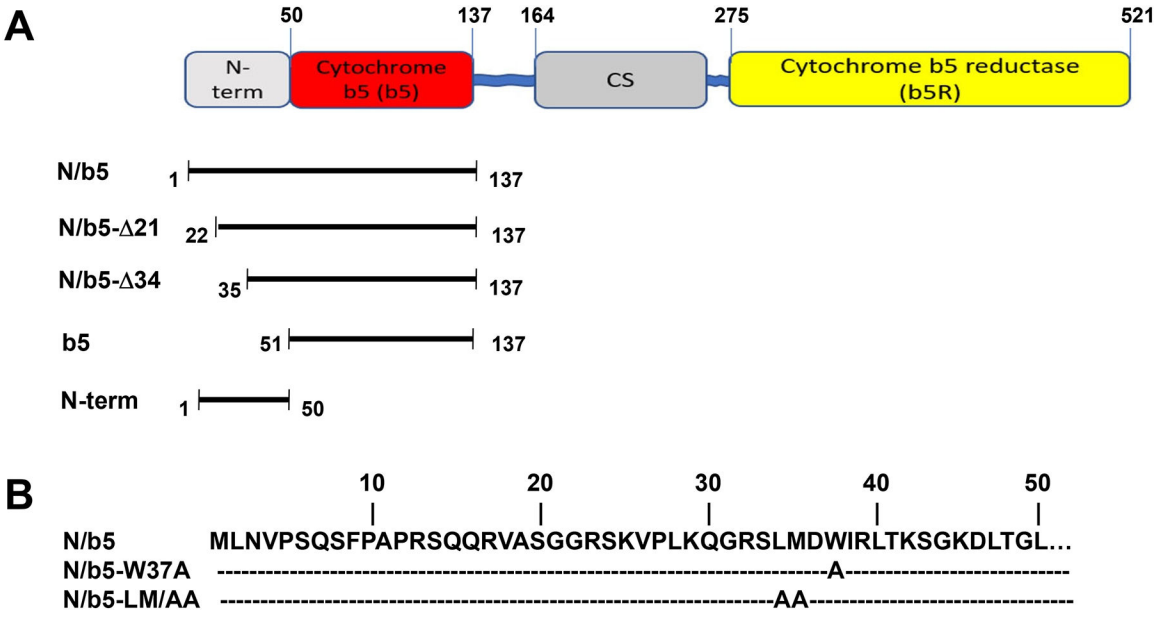
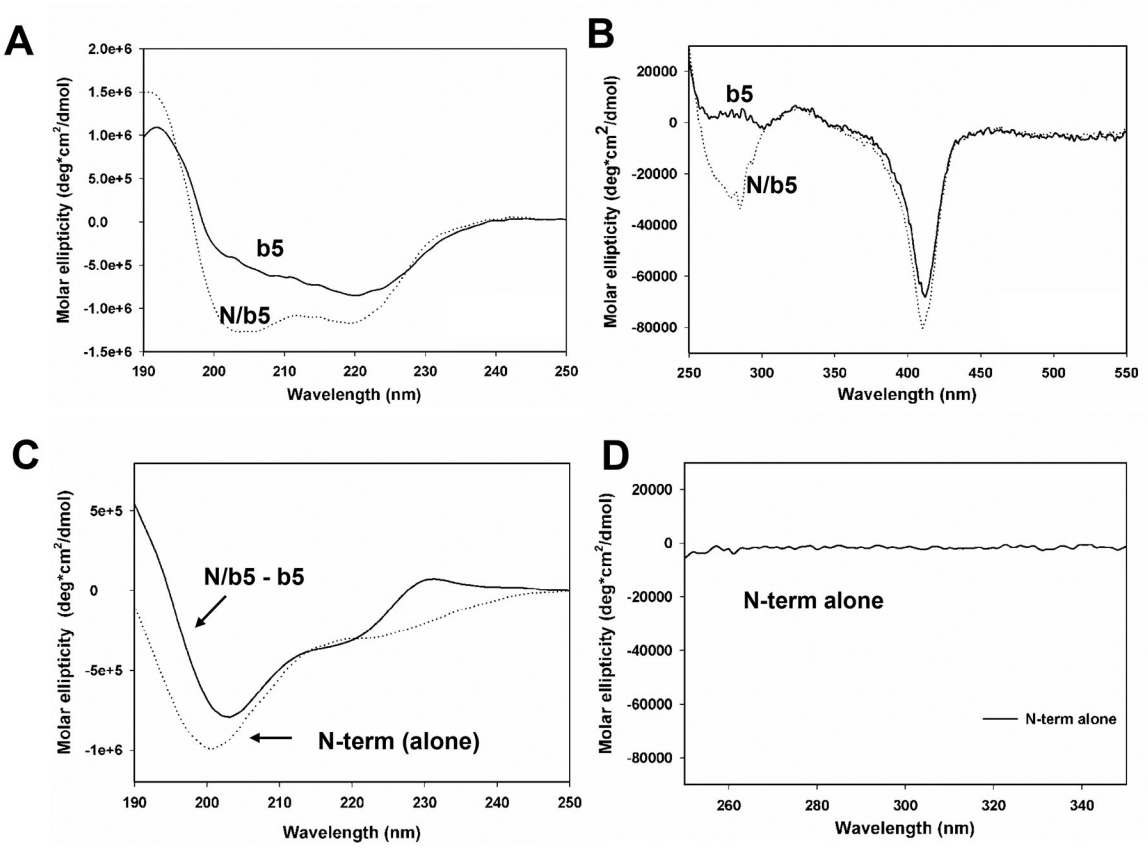


Figure 1.

(***A***) Schematic diagram of Ncb5or constructs. Except Ncb5or-Δ34, all are used in the current study. (***B***) Amino acid sequences of the N-terminal region in human Ncb5or and its point-mutation variant.



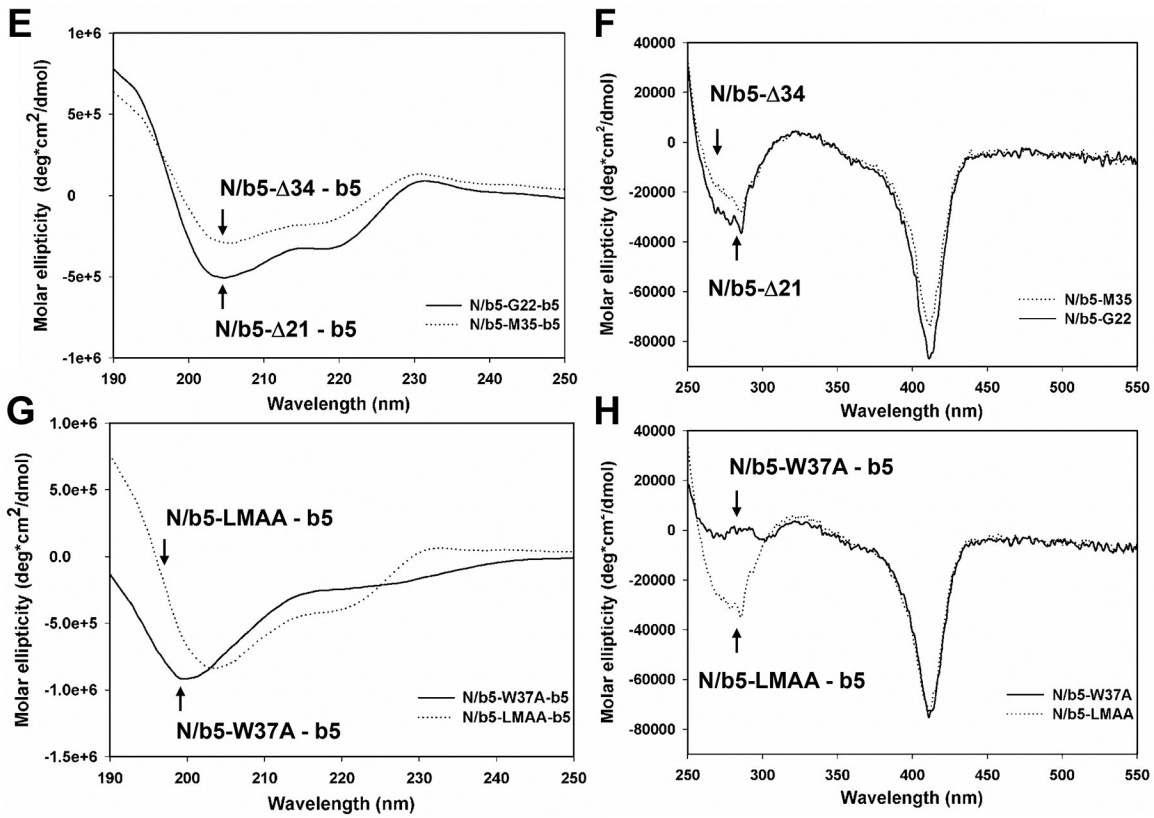


Figure 2.

Far-UV (***A***) and near-UV/visible (***B***) CD spectra of Ncb5or-N/b5 and Ncb5or-b5. Far UV (***C***) and near-UV/visible (***D***) spectra of Ncb5or-N as a stand-alone peptide (alone) or when fused to Ncb5or-b5 (difference, N/b5 – b5). Difference far UV (***E, G***) and near-UV/visible (***F, H***) CD spectra of Ncb5or-N/b5 variants. Difference spectra were presented by smoothing raw data, which are shown in Figure S3, with the use of negative exponential program in SigmaPlot 10. Molar ellipticity is used in all CD spectra. Values of signature bands are summarized in Table 1.

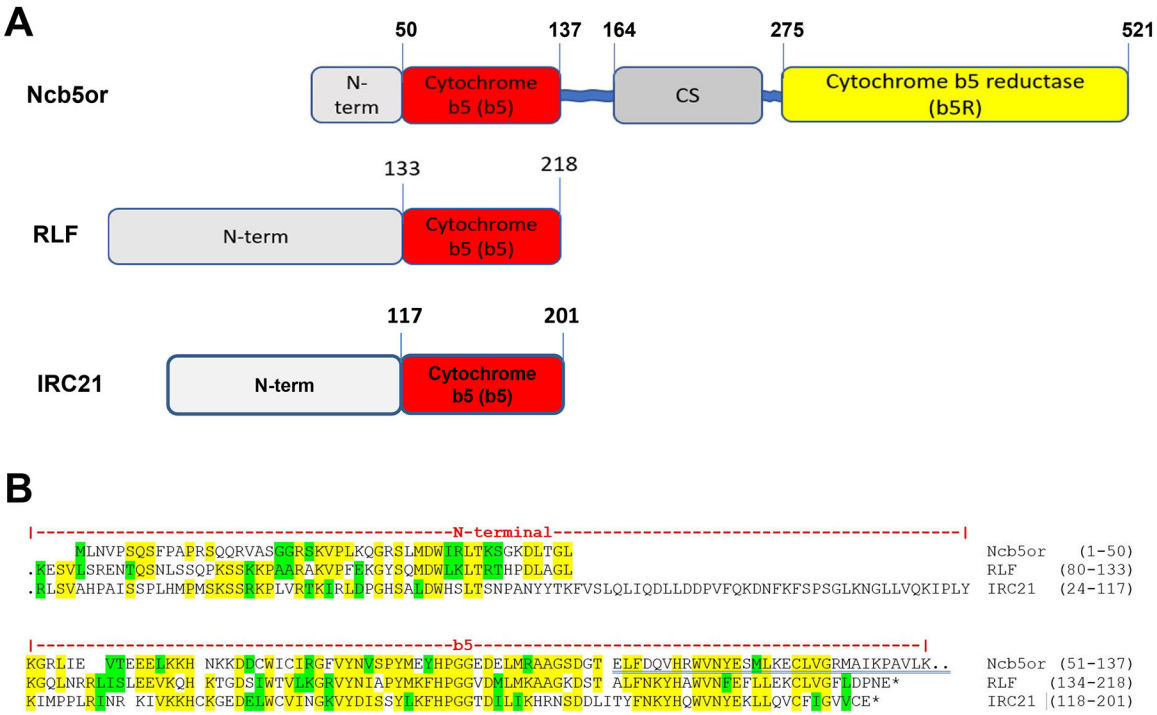


Figure 3.

(***A***) Comparison of the domain arrangements in human Ncb5or, rice RLF and baker's yeast IRC21 proteins. (***B***) Amino-acid sequence alignment of Ncb5or (residues 1–137, GenBank accession # NP_057314.2), RLF (residues 80–218, XP_015647768.1), and IRC21 (residues 24–201, NP_013789.1). Identical and similar residues are indicated in yellow and green, respectively.

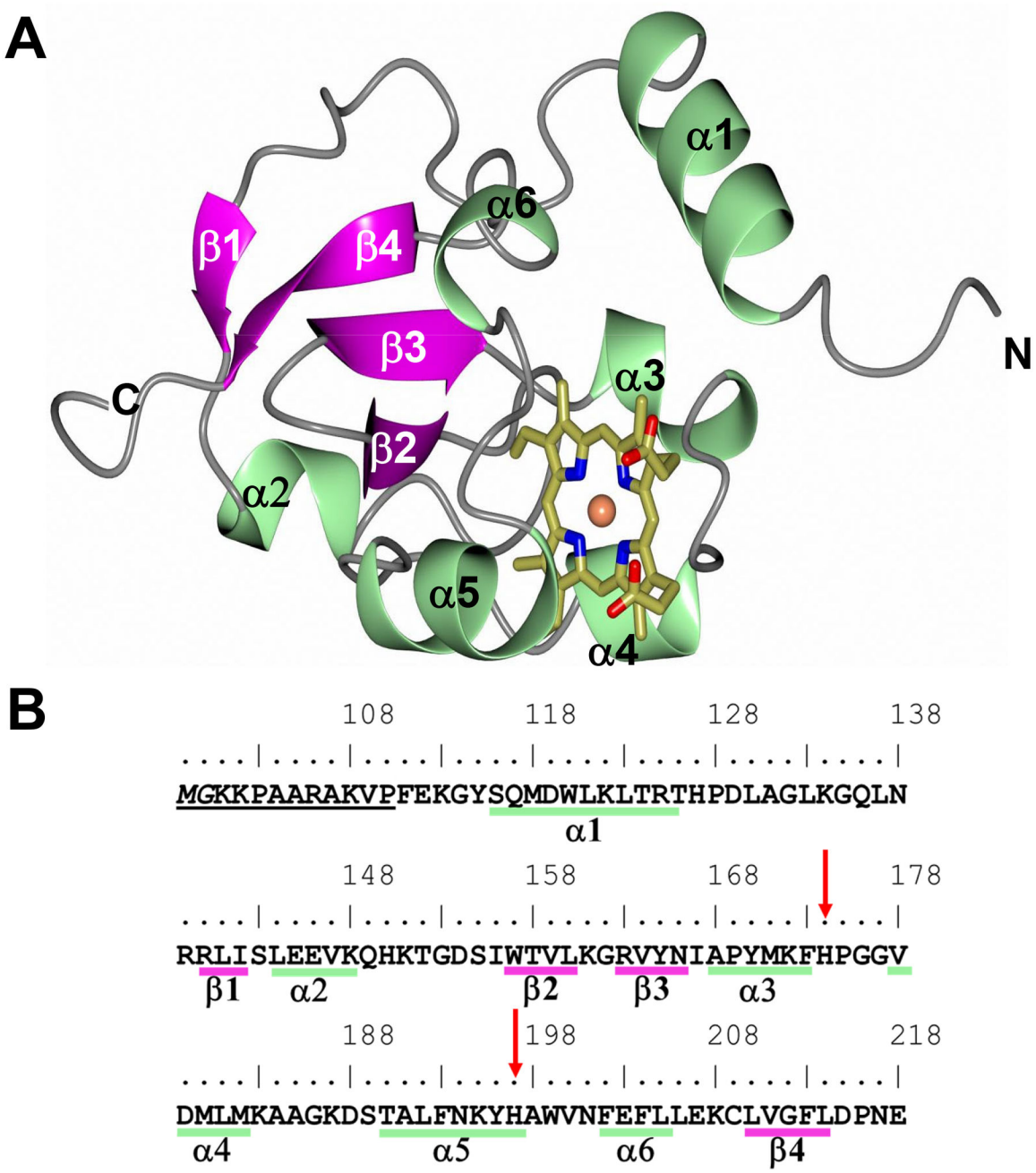


Figure 4.

(*A*) Structure of RLF (subunit A) with the secondary structure elements annotated. The heme molecule is rendered as gold cylinders. **(*B*)** RLF sequence with the secondary structure elements annotated. The red arrows indicate the His residues that coordinate the heme.

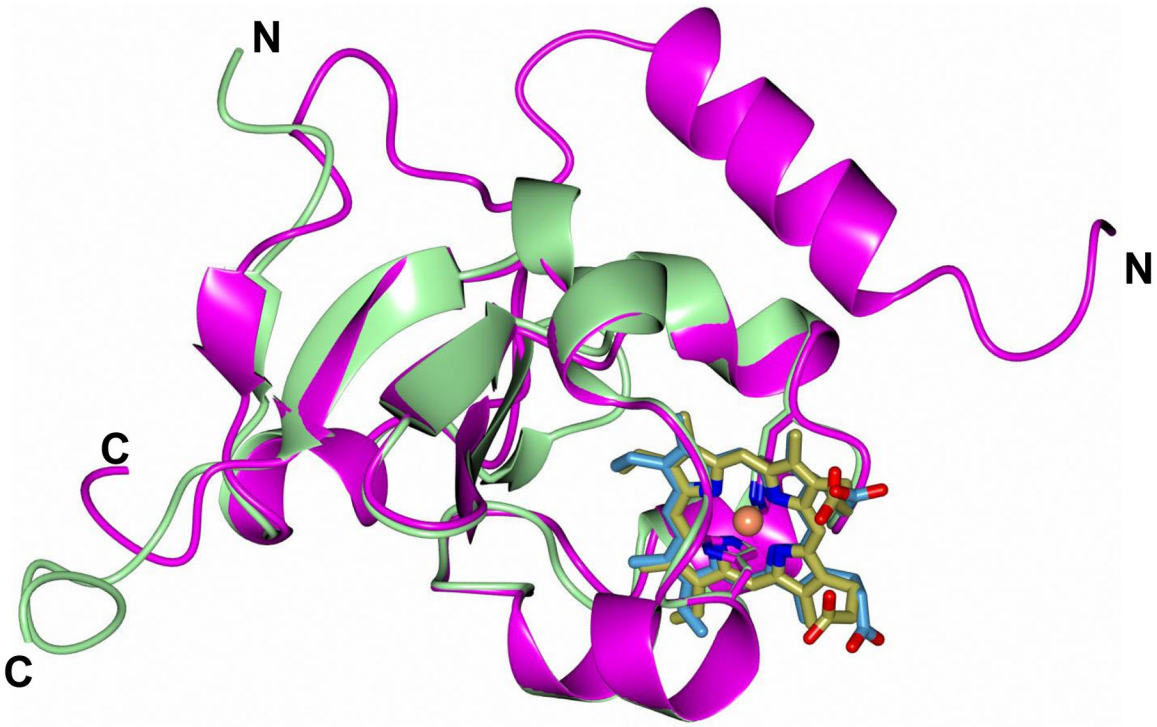


Figure 5.

Superposition of rice RLF (magenta) with the Ncb5or b5-domain (PDB 3LF5, green). RMSD between Ca for b5 core residues (see Figure 4) is 1.01 Å.

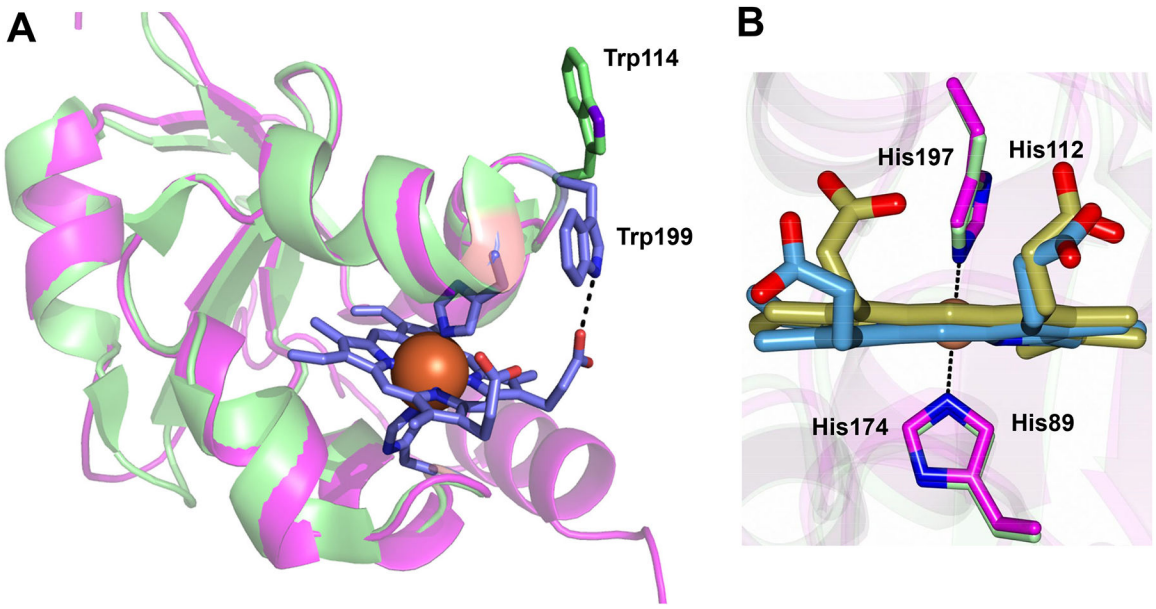


Figure 6.

(*A*) Comparison of conserved, surface exposed tryptophan residue in Ncb5or (green) and RLF (magenta). Trp^{199} in the RLF structure forms a hydrogen bond with a heme propionate group, while Trp^{114} in Ncb5or extends into solvent. This flexibility suggests a possible role in modulating electron flow. **(*B*)** Comparison of the heme orientation and the coordinating His residues. His^{89}/His^{112} are colored green for 3LF5 and His^{174}/His^{197} are colored magenta for RLF.

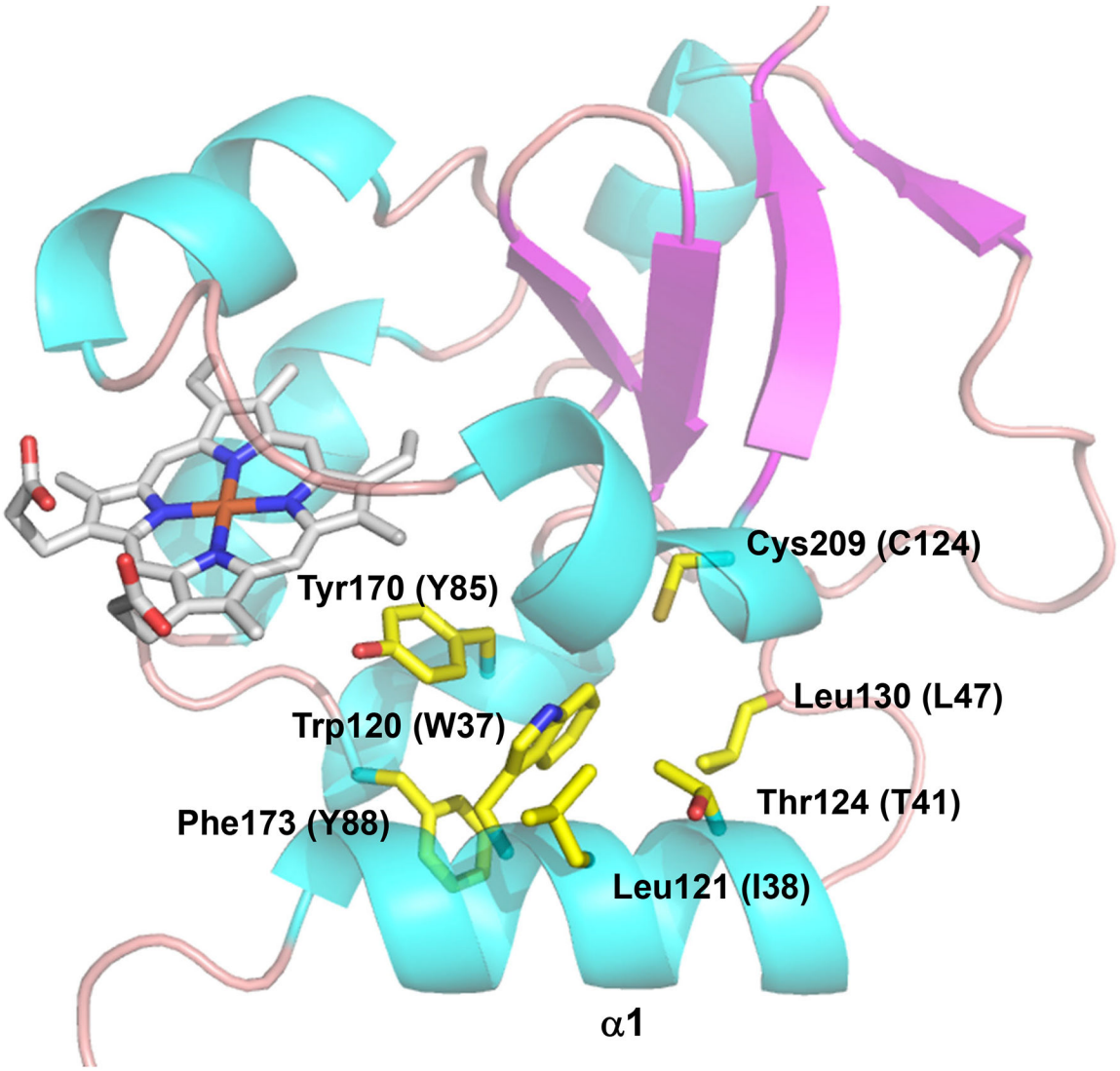


Figure 7.

View of the RLF structure highlighting helix α1 and locations/interactions of residues. Corresponding Ncb5or residues are shown in parentheses. Black dotted lines indicate hydrogen bonds.

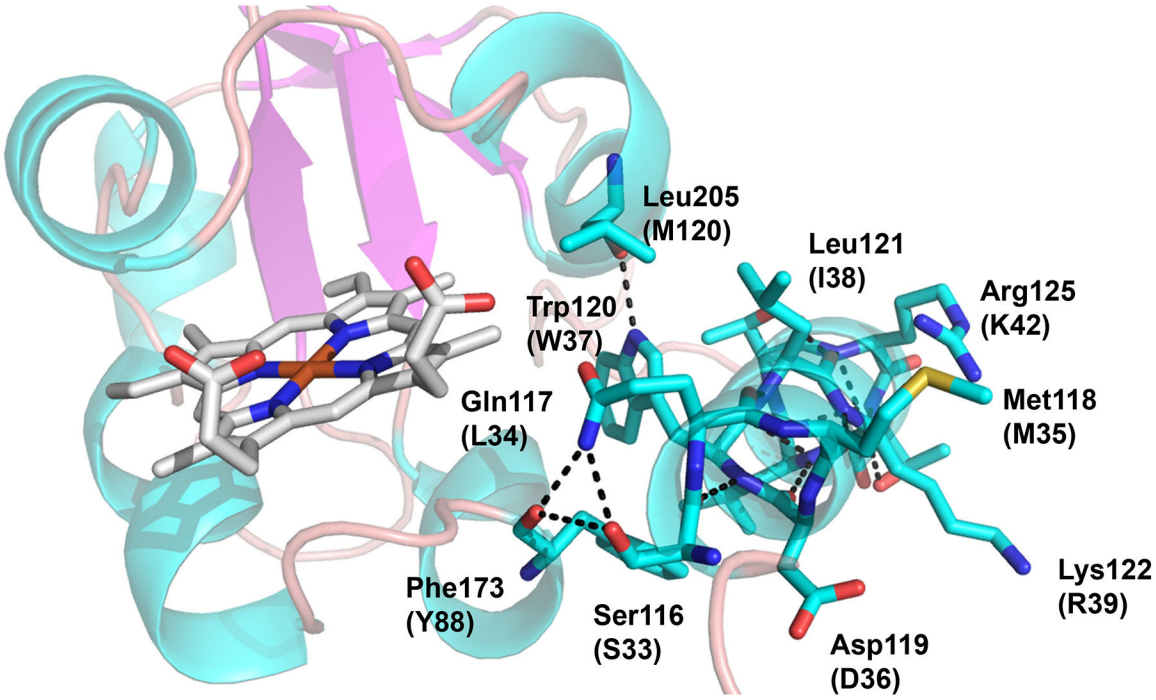


Figure 8.

Residue Trp^{120} in RLF, corresponding to Trp^{37} in Ncb5or, forms a network of hydrophobic packing interactions with other conserved residues (shown with carbon atoms rendered in yellow). The hydroxyl group of Tyr^{170} (Tyr^{85} in Ncb5or) makes van der Waals contact with the heme, and the thiol group of Cys^{209} (Cys^{124} in Ncb5or) makes van der Waals contact with Trp^{120}.

Table 1.

Signature signals of CD spectra for N/b5, its variants, and b5 of human Ncb5or.

Molar ellipticity (deg*cm^2/dmol, 10^{-3})	Far-UV (minus b5)		Helix	Near-UV	Tertiary structure	Soret	Heme contact
	190 nm	220 nm		286 nm		413 nm	
N/b5	545	−310	Yes	−35.0	Yes	−82.5	Yes
N/b5-Δ21	778	−312	Yes	−36.5	Yes	−84.7	Yes
N/b5-LMAA	764	−397	Yes	−34.7	Yes	−69.5	No
N/b5-Δ34	636	−137	Yes	−27.8	Partial	−70.5	No
N/b5-W37A	−133	−245	No	−0.35	No	−72.6	No
b5	N/A	N/A	N/A	2.26	No	−71.6	No

Molar ellipticity values of far-UV signals (190 and 220 nm) are obtained from difference CD spectra (Figures 2C,E,G) and indicate secondary structure in the N-terminal domain. Its tertiary structure and impact on the heme center are suggested by molar ellipticity values of near-UV (286 nm) and Soret (413 nm) signals, respectively, from CD spectra (Figures 2B,F,H). All Molar ellipticity values are presented in deg*cm^2/dmol. N/A: not applicable.