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

Hao Zhu\textsuperscript{1}, David Benson\textsuperscript{2}, Bin Deng\textsuperscript{1}, Maithri M. Kashipathy\textsuperscript{2}, Scott Lovell\textsuperscript{2}, Kevin P. Battaile\textsuperscript{3}, Anne Cooper\textsuperscript{2}, Philip Gao\textsuperscript{2}, and Aron W. Fenton\textsuperscript{1}

\textsuperscript{1}The University of Kansas Medical Center
\textsuperscript{2}The University of Kansas
\textsuperscript{3}New York Structural Biology Center

July 22, 2023

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 with a distinctive, conserved L\textsuperscript{34}MDWIRL\textsuperscript{40} motif that has no homologs in animals but is present in root lateral formation protein (RLF) in rice and Increased Recombination Center 21 (IRC21) in baker’s yeast, and in these proteins, it is likewise attached 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 $\alpha$-helix ($S\textsuperscript{116}QMDWLKLTTRT\textsuperscript{126}$) packing against two of the four helices in the b5 domain that surround heme ($\alpha$2 and $\alpha$5). The Trp120 side chain forms a network of interactions with the side chains of four highly conserved residues corresponding to Tyr85 and Tyr88 ($\alpha$2), Cys124 ($\alpha$5), 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.

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,\textsuperscript{1} Bin Deng,\textsuperscript{4} Maithri M. Kashipathy,\textsuperscript{5} Scott Lovell,\textsuperscript{5} Kevin P. Battaile,\textsuperscript{7} Anne Cooper,\textsuperscript{6} Philip Gao,\textsuperscript{6} Aron W. Fenton,\textsuperscript{3} Hao Zhu\textsuperscript{2,3,4}

Department of \textsuperscript{1}Chemistry, University of Kansas, Lawrence, KS 66045, U.S.A.; Department of \textsuperscript{2}Clinical Laboratory Sciences, \textsuperscript{3}Biochemistry and Molecular Biology, \textsuperscript{4}Physical Therapy and Rehabilitation Science, University of Kansas Medical Center, Kansas City, KS 66160, U.S.A.; \textsuperscript{5}Protein Structure and X-ray Crystallography Laboratory, \textsuperscript{6}Protein Production Group, The University of Kansas, 2034 Becker Drive, Lawrence, KS 66047, USA; \textsuperscript{7}NYX, New York Structural Biology Center, Upton, NY, 11973, USA.
TITLE RUNNING HEAD: Intrinsically disordered region of Ncb5or

AUTHOR INFORMATION

Corresponding Authors:
drb@ku.edu (David R. Benson), hzhu@kumc.edu (Hao Zhu)

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, E-mail: drb@ku.edu;

Hao Zhu, Ph.D., 3001 Rainbow Blvd., MSN 4048G-Eaton, Kansas City, KS 66160, U.S.A., Phone: 913-588-2989, FAX: 913-588-5222, E-mail: hzhu@kumc.edu.

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.

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 with a distinctive, conserved L\textsuperscript{34}MDWIRL\textsuperscript{40} motif that has no homologs in animals but is present in root lateral formation protein (RLF) in rice and Increased Recombination Center 21 (IRC21) in baker’s yeast, and in these proteins, it is likewise attached 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 \textalpha-helix (S\textsuperscript{116}QMDWLKLTRT\textsuperscript{126}) packing against two of the four helices in the b5 domain that surround heme (\textalpha\textsubscript{2} and \textalpha\textsubscript{5}). The Trp120 side chain forms a network of interactions with the side chains of four highly conserved residues corresponding to Tyr85 and Tyr88 (\textalpha\textsubscript{2}, Cys124 (\textalpha\textsubscript{5}), 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.
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.

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. Ncb5or is widely expressed in animal cells and localized to endoplasmic reticulum. Ncb5or contains two redox domains and one linkage domain that are homologous to cytochrome b5 (Cyb5), cytochrome b5 reductase (Cyb5R), and members of C HORD and S GT1 (CS) family, respectively. 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. 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). Despite having the same general fold, the b5 domain exhibits substantially different heme ligation from that of microsomal cytochrome b5 (Cyb5A). 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.

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 L34MDWIRL 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 Trp120 (corresponding to Trp37 in Ncb5or) to be part of an 11-residue α-helix (S116QMDWKLKLR126) packing against two of the four helices that surround heme (α2 and α5 in Ncb5or). The Trp120 side chain forms a network of interactions with the side chains of four highly conserved residues that are equivalent to Ncb5or residues Tyr85 and Tyr88 in α2, Cys124 in α5, and Leu47. 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. 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 (Met1 through Lys37); N/b5-Δ21 (Gly22 through Lys137); N/b5-Δ34 (Met35 through Lys137); b5 (Lys51 through Lys137); and N-term (Met1 through Leu50). 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.
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 (Stratgene, 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 constructs of human Ncb5or and rice RLF were expressed in *E. coli* BL21(DE3) or BL21(DE3)pLysS or BL21(DE3)pRARE cells and purified as previously described.$^8$ In combination with size exclusion chromatography on a Superdex 75 10/300 GL column, ion-exchange chromatography was performed for b5 and N/b5-$\Delta$34 on a Q HP column, N/b5 and variants (N/b5-W37A, Nb5-LMAA, and N/b5-$\Delta$21) on a SP HP column, and rice RLF on both Q and SP HP columns. Full-length Ncb5or and its variant Ncb5or-$\Delta$50 were prepared as previously described$^2$, except that 0.5 mM IPTG was used for induction in TB media at 15°C. All purification steps were performed at 4°C. SDS-PAGE was used to estimate the purity of each protein used in spectroscopic analyses, kinetics studies and crystallization screening (all greater than 95%). All purified proteins were flash frozen in liquid nitrogen and stored as aliquots at -80°C until use, except for samples prepared for crystallization screening which were stored at 4°C. The size of each polypeptide product was confirmed by mass spectrometry. Expected sizes are: N/b5, N/b5-W37A, N/b5-LMAA, 15.6 kD; N/b5-$\Delta$21 and RLF, ~13.6 kD; N/b5-$\Delta$34, 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$^5$, Ncb5or$^1$, and Ncb5or-$\Delta$50$^1$.1. The FAD content of Ncb5or-b5R was determined by $A_{413}$ and used to represent enzyme concentrations as previously described.$^8$ Final protein yields (mg/L): 8 (b5), 5 (N/b5 and variants), 2 (b5R), $^1$ (Ncb5or, Ncb5or-$\Delta$50, RLF). Solubility in low-salt buffer: N/b5, RLF > b5 > Ncb5or > Ncb5or-$\Delta$50 > Ncb5or-b5R. Protease inhibitor (Sigma, St. Louis, MO) was used during protein preparation to prevent proteolytic cleavage at the b5-CS border of Ncb5or and Ncb5or-$\Delta$50. A SUMOpro Expression System was used to prepare N-term.$^{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 $\varepsilon_{413} = 130$ 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 $\mu$M. 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 ($\theta$, $\delta\varepsilon_{\theta}$ cm$^2$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 mM of protein and 100 mM of crystallization solution were dispensed and equilibrated against 50 $\mu$L of the latter. Crystals approximately 300-500 $\mu$m 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 $\mu$L) 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 $P\ 2_1\ 2_1$ or $P\ 2_1$. 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 $P\ 2_1$ 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.\textsuperscript{18} Further refinement and manual model building were conducted with Phenix\textsuperscript{19, 20} and Coot\textsuperscript{21} respectively. TLS parameters\textsuperscript{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\textsuperscript{24} and figures were prepared using the CCP4MG package\textsuperscript{25}. Polder omit electron density maps were calculated using Phenix.\textsuperscript{26} Structure superposition was carried out with GESAMT.\textsuperscript{27} Crystallographic data are described in Table S1, and the best data set was at 1.55 Å resolution.

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\textsuperscript{34}MDWIRL\textsuperscript{40} in mammalian orthologs.

The PONDR VL-XT program\textsuperscript{28-30} 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\textsuperscript{34}MDWIRL\textsuperscript{40} sequence. This included PROF\textsuperscript{31} and PSIPRED\textsuperscript{32, 33} (Figure S2B), as well as SPINE-X\textsuperscript{34} (not shown). The latter two \textit{ab initio} algorithms (\textit{i.e.}, those not utilizing homology modeling) have been shown to yield the most accurate predictions.\textsuperscript{35} AGADIR\textsuperscript{36}, an algorithm that predicts helical content of peptides on the basis of helix/coil transition theory, also predicts some helical tendency for the L\textsuperscript{34}MDWIRL\textsuperscript{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.\textsuperscript{37} 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.\textsuperscript{38} 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 (\textit{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).\textsuperscript{38} 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 L34MDWIRL^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.\(^{39, 40}\) 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 tested 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.

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. In comparison, 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.\(^{41}\) 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). However, the CD Soret band signal for the LMAA mutant is virtually identical to that of Ncb5or-b5 (Figure 2H). These observations suggest that the LM-AA mutation does not alter the ability of N-term to dock with the b5 core, with associated adoption of tertiary structure involving Trp37, but that it alters (or perhaps weakens) interactions between the N-terminal region and the b5 core.
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 (R educed L eteral root F ormation protein) from plants (e.g., rice, *Arabidopsis* and glycine) and IRC21 (*I* ncreased *R* ecombination *C* enter 21) protein from yeast (e.g., *Saccharomyces cerevisiae* and *S. pombe*). RLF appears to regulate lateral root formation independently of ARF7/19-mediated auxin signaling, whereas IRC21 functions in chromatin remodeling and DNA repair likely through its binding to lipid and protein phosphatase.

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 (Q15MDWLKI121) that is strikingly similar to residues 34-40 of Ncb5or which encompass Trp37 (L34MDWIRL40). The N-terminal regions of IRC21 proteins are approximately 77 residues in length, and in that respect are more like Ncb5or than are the plant RLF proteins. However, there is weaker homology for residues 61-67 (A61LDSHSL67) that correspond to residues 34-40 of Ncb5or (Figure 3A-B). In addition, the sequence separating this signature from the b5 core is approximately 40 residues longer in IRC21 proteins than in Ncb5or and RLF, an insertion that may have arisen subsequent to divergence of the plant, fungi and animal kingdoms.

**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 the published structure of the Ncb5or b5 domain, with superposition yielding an RMSD deviation 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 His177). His174 (His89 in Ncb5or) is located in the loop separating α3 and α4, while His177 (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 His177 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. The corresponding angles for His89 and His112 in the two subunits of the Ncb5or b5 domain structure (3LF5) are 83.2° and 81.3°. 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 the better known 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. 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 S5), with its only inter-protein interactions involving the side chain of adjacent residue Arg113. The rice RLF residues corresponding to the Ncb5or His112-ArgTrp114 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 Ne-H and one of the heme propionate groups. Manually changing the side chain torsional angles of Trp114 in the Ncb5or b5 domain crystal structure (χ1 = -53.9; χ2 = 115.7°) to match those of Trp199 in the new rice RLF structure (χ1 = 58.9; χ2 = 85.4°), using PyMol, 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 carrying out 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 Trp114 of Ncb5or. The Giardia proteins were shown to exhibit redox potentials similar to those determined for Ncb5or,6 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.51 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.50, 51 This notion is supported by the observation of cytochrome b5 proteins in protists (Trypanosoma and Dictostelium) that share the same b5 core 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 (S126QMDWLKTRT126), 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 17-residue stretch connects α1 to α2, the first element of defined secondary structure in the “classic” b5 core (α2 is the only helix in the b5 core that does not comprise part of the heme binding pocket). The corresponding stretch in Ncb5or is two residues shorter, with the gap appearing close to the junction with α2. 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 (Leu50 and Arg53). The Asp129 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 Cβ3 of Ala131 (Thr48), Cβ3 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.

It is worth noting that the interaction between the Trp120 and Cys209 side chains (Trp37 and Cys124) involves the cysteine thiol (SH group) packing against the tryptophan π system. Such interactions are relatively common and are thought to play a stabilizing role. Both residues are invariant in Ncb5or proteins.

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 (Ser31 and Leu32), the phenolic side chain of Tyr170 in α3 (Tyr85), and the backbone carbonyl of Phe173 in α3 (Tyr88). While this network of hydrogen bonds may be important in stabilizing α1, it is noteworthy that the polar OH group of the Tyr170 side chain also makes van der Waals contact with one of the heme meso positions. 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.

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\textsuperscript{54-57}, of which the specificity is generally mediated by motif structures.\textsuperscript{58, 59} In this context, it is worth investigating whether IRC21 utilizes the helical motif in the Nb5 module to activate protein phosphatase 2A for DNA damage response in yeasts.\textsuperscript{47} Intrinsically disordered regions appear to be enriched in disease-related proteins, such as the N-terminal transactivation domain of tumor suppressor p53,\textsuperscript{60} alpha-synuclein in Parkinson’s disease\textsuperscript{61}, tau protein in Alzheimer’s disease\textsuperscript{62}, and in a number of proteins associated with cardiovascular disease.\textsuperscript{63} 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.\textsuperscript{64} 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\textsuperscript{48, 65} 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.\textsuperscript{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. Alternatively, the N-terminal regions of these proteins could function to regulate their enzymatic functions, through reversible docking. Suggestive of this latter scenario is the observation that the 17 residues separating the N-terminal helix α1 from α2 in rice RLF protein exhibit few interactions with the “classic” portion of the b5 core, in contrast to the 17-residue helix that is nucleated by the Trp37-dependent docking interactions. These possibilities will be the subject of future studies. A 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 be aimed at probing potential functional roles of other strictly conserved residues in Ncb5or emphasized above. These include Met35 in the N-terminal helix, and Trp114 which is located two residues from heme ligand His112. The side chains of both these 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.\textsuperscript{66} Functional roles could include modulating heme reduction by the b5R domain or reduction by heme of the still unknown substrate.

ACKNOWLEDGMENT

We thank Christopher D. Lima (Sloan-Kettering Institute, New York) for the SUMO protease as a gift.

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.

<table>
<thead>
<tr>
<th>Molar ellipticity</th>
<th>Far-UV (deg*cm^2/dmol)(minus b5)</th>
<th>Far-UV (minus b5) Helix</th>
<th>Near-UV</th>
<th>Tertiary structure</th>
<th>Soret</th>
<th>Heme contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>190 nm</td>
<td>545453</td>
<td>-310424</td>
<td>Yes</td>
<td>-35041</td>
<td>-82501</td>
<td>Yes</td>
</tr>
<tr>
<td>N/b5</td>
<td>778363</td>
<td>-312167</td>
<td>Yes</td>
<td>-36453</td>
<td>-84650</td>
<td>Yes</td>
</tr>
<tr>
<td>N/b5-Δ21</td>
<td>764345</td>
<td>-396791</td>
<td>Yes</td>
<td>-34688</td>
<td>-69546</td>
<td>No</td>
</tr>
<tr>
<td>LMAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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*cm²/dmol. N/A: not applicable. Table S1. Crystallographic data for RLF.

<table>
<thead>
<tr>
<th>Molar ellipticity</th>
<th>Far-UV (deg*cm²/dmol minus b5)</th>
<th>Far-UV (minus b5)</th>
<th>Helix</th>
<th>Near-UV</th>
<th>Tertiary structure</th>
<th>Soret</th>
<th>Heme contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/b5-Δ34</td>
<td>636442</td>
<td>-136573</td>
<td>Yes</td>
<td>-27727</td>
<td>Partial</td>
<td>-70516</td>
<td>No</td>
</tr>
<tr>
<td>N/b5-</td>
<td>-132660</td>
<td>-245259</td>
<td>No</td>
<td>-346</td>
<td>No</td>
<td>-72577</td>
<td>No</td>
</tr>
<tr>
<td>W37A b5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>2262</td>
<td>No</td>
<td>-71620</td>
<td>No</td>
</tr>
</tbody>
</table>

RLF

Data Collection

Unit-cell parameters (Å, °)    a=58.95, b=26.14, c=82.87, β=95.5
Space group                 P2
Resolution (Å)               45.79-1.55 (1.58-1.55)
Wavelength (Å)               1.0000
Temperature (K)              100
Observed reflections         150,129
Unique reflections           36,463
<1/σ(I)>1                   11.0 (2.0)
Completeness (%)1           97.6 (98.1)
Multiplicity1                4.1 (4.0)

Refinement

Resolution (Å)1            41.25-1.55
Reflections (working/test)1 34,634/1,819
Rfactor / Rfree (1.3)        17.5/20.0
No. of atoms (Protein/Heme/Water) 1,686 / 86 / 164
Model Quality

R.m.s deviations
Bond lengths (Å)            0.008
Bond angles (°)             0.961
Mean B-factor (Å²)
All Atoms                   23.5
Protein                     22.9
Heme                        20.2
Water                       31.3
Coordinate error
(maximum likelihood) (Å)    0.12
Ramachandran Plot
Most favored (%)            97.6
Additionally allowed (%)     2.4
Values in parenthesis are for the highest resolution shell.

\[ R_{\text{merge}} = \frac{\Sigma_{i} |I_i(hkl) - <I(hkl)>|}{\Sigma_{i} \Sigma_{hkl} I_i(hkl)}, \]

where \( I_i(hkl) \) is the intensity measured for the \( i \)th reflection and \( <I(hkl)> \) is the average intensity of all reflections with indices \( hkl \).

\[ R_{\text{factor}} = \frac{\Sigma_{hkl} |F_{\text{obs}}(hkl) - |F_{\text{calc}}(hkl)|||}{\Sigma_{hkl} |F_{\text{obs}}(hkl)||}; \]

\( R_{\text{merge}} \) is calculated using identical manner using 5% of randomly selected reflections that were not included in the refinement.

\( R_{\text{free}} \) is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

\( R_{\text{meas}} = \frac{\Sigma_{hkl} |F_{\text{obs}}(hkl)|}{\Sigma_{hkl} |F_{\text{calc}}(hkl)|}; \]

\( R_{\text{merge}} \) is redundancy-independent (multiplicity-weighted) \( R_{\text{merge}} \).

\( R_{\text{pim}} \) is precision-indicating (multiplicity-weighted) \( R_{\text{merge}} \).

CC\(_{1/2}\) is the correlation coefficient of the mean intensities between two random half-sets of data.

**FIGURE LEGENDS**

**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 Ncb5or and RLF. (B) Amino acid sequence alignment of human Ncb5or (residues 1 – 137), rice RLF (Oryza sativa, residues 72 – 218), and baker’s yeast IRC21 (Succharomyces cerevisiae, residues 17-201). Identical and similar residues are coded 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 deviation between Cα 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 \( \alpha \)1 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\(^{120}\).

**Figure S1.** Sequence alignment of N-terminal region and cytochrome b5 domain of known Ncb5or proteins of a diverse background. Asterisk and "_" indicate conserved and chemically similar residues, among mammals, respectively. Non-conserved residues at key positions are indicated in red. Species (GenBank accession number) are as follows: (mammals in yellow) human (NM_016230), monkey (XP_518614.3), monkey Old World (XP_011920356.1), guinea pig (XP_005006001), mouse (NM_024195), rat (NM_133427), cow (NM_001038159), horse (XM_001499913), sheep (XP_027818354), pig (XP_013848110), dog (XM_532219),

12
minke whale (XM_007165020.1), ferret (XP_004773650), bat/Molossus (XP_011363184), brown bat (XM_008145048), opossum (XM1375853), platpus (XP_028903113); (birds, reptiles and amphibians in green) Zebra Finch (XP_030123394), golden eagle (XP_029856007), peregrine falcon (XP_013152060), chicken (XM_002879520), Dalmatian pelican (KFR09350), crested ibis (KFX1198870), turtle/Terrapene (XP_026510156) turtle/Pelodiscus (XP_006133410), garter snake (XP_032072321), Xenopus (XP_031757575), Coelacanth (XP_014343350); (fishes in blue) Cyprinidae (XM_016550377), Zebrafish (NP_001018461), Pimephales (XP_039507272), Salmon (XP_020361560), bony fish (XM_019263747), Fugu (TNN02071), Pufferfish/Japan (XM_003965751), tire track eel (XR_003294861), swamp eel (XM_020621659), Gasterosteus (XP_040020873), rice fish (AM140533), rice fish India (XP_024117733), Lamprey (XP_032829495), Skate (XP_032877801), seahorse (XM_019880106), (invertebrates in grey) Lancelet (XM_002603916), Trichoplax (XM_002112883), tunicate (XR_053035), Hydra (XM_002165807), sea anemone (XP_048577331), coral (EZ038676, Acropora), coral/Exaiptasia (KXJ05517), sea urchin/Lytechinus (XP_041463305), sea urchin (XP_1198870), Teladorsagia (PIO73582), hookworm/Necator (RCN51358), C.elegans (NM_001026613), pig worm (KFD50571c, Trichuris suis), Trichuris (KJH45328); (insects in purple) Drosophila (NM_137575), Mosquito/Anopheles (XP_041181397), Mosquito/Aedes (KXJ83737), Honeybee (XM_394412), Louse (XM_002428283), Jewl wasp (XM_001601866), Beetle (XM_963153), Pea aphid (XM_001949299), Deer tick (XM_002401084),

Figure S2. Prediction of disordered domains in Ncb5or by PONDR VL-XT (A ). Prediction of secondary structure in the 50 N-terminal residues of Ncb5or by PSIPRED and PROF (B ) and by AGADIR (C ). AA (amino acid): Target sequence, Pred: Predicted secondary structure (H=helix, E=strand, C=coil), Conf: Confidence (0=low, 9=high).

Figure S3. Sequence alignment of the N-terminal region and b5 core (partial) of human Ncb5or against RLF proteins in plants. Asterisk (*) indicates conserved residues, whereas “_” represents chemically similar residues.

Figure S4. Sequence alignment of the N-terminal region of human Ncb5or against IRC21 proteins in fungi. Asterisk (*) indicates conserved residues, whereas “_” represents chemically similar residues.

Figure S5. Sequence alignment of the N-terminal region and cytochrome b5 domain of human Ncb5or against cytochrome b5 proteins in protists (Dictyostelium and Trypanosoma ) and parasites (all others). Asterisk (*) indicates conserved residues, whereas “_” represents chemically similar residues. The conserved “HPGG” heme binding site is highlighted, and the N-terminal motif in human and protists is indicated by “——”.

REFERENCES


Figure 1
Figure 2
Figure 3

A

Ncb5or

RLF

IRC21

B

[Sequence alignments and annotations]

human Ncb5or
rice RLF
yeast IRC21

Figure 3
Figure 4

A

B

108 118 128 138

...............

MGKKPAARAKVPFEKGYSQMDWLKLTRTHPDLAGLKGQLN

148 158 168 178

RRLLSLTEVKQHKTGDSIWTVLKGRVYNIAPYMKFHPGGV

188 198 208 218

DMLMKAAGKDSTALFNKYHAVNFETFLEKCLVGFLDPNE
Figure 8
Sequence alignment of N-terminal region and cytochrome b5 domain of known Ncb5or proteins.
Figure S2

A

B

AA: MLNVPSSFPAPRSSQVRQVGGRSKVPLQGRSLMDWIRLTSGKDLTGLKRLIEVTEELKHKHNKD

PSIPRED results:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCCC
Conf: 988899999998755558999643105899985359987428999889888421039999976518889

PROF results:
Pred: ___________________________________________________________________________HHHHHHH____________________EEE_HHHHHH_______

Figure S2
Sequence alignment of N-terminal region of human Ncb5or against RLF in plants.
Sequence alignment of N-terminal region of human Ncb5or against IRC21 in fungi.
Human, Ncb5or
NP_013789, yeast (IRC21)
XP_002416862,Candida
EEQ45925,Candida
XP_002550312,Candida
XP_002616018,Claviscora
CAQ6857,Dubymycetes

NP_013789, yeast (IRC21)
XP_002416862,Candida
EEQ45925,Candida
XP_002550312,Candida
XP_002616018,Claviscora
CAQ6857,Dubymycetes

...KPNVRFAAP...QRLSVHAISSPVLHPMSSKRRKPLIVKIRDLPGHSLDWSLTASRFQMRGV.
...SNRGLFPAVNSAQRA...GSAVQKQKVKLGPHSLDWAHLNRTPQRHKLGRV.
...SNRGLFPAVNSAQRAANATTT...SSAAANGQKVKLGPHSLDWAHLNRTPQRHKLGRV.
...SNRLPLLNQSAQRA...SSASERTKVKHGPHSLDWAHLNRTPQRHKLGRV.
...KPFTRPSFPSFAQRAAPPIIA...SSSSTPQAARRAQVKLGPHSLDWAHLNRAPKHKLGRV.
...TLGLTPPSAQRAVQ...SLPGTKNARKVRKVLAPGPHSLDWAHLNRTPQRHKLGRV.
...TLGLTPPSAQRAVQ...SLPGTKNARKVRKVLAPGPHSLDWAHLNRTPQRHKLGRV.
...IFPANPSQRS...RLGIPNGNRSRQVKVLGPHSLDWAHLNRTPQRHKLGRV.
...APGFPAPSSLLRPSPPPVTQVTGVC...LTQPQAGASKEKVKLGPHSLDWAHLNRTPQRHKLGRV.
...SLLRAPPCTTPGKSLSSSTSLPLG...LPQKPSQQRPQVILEPGHSLDWAHLNRTPQRHKLGRV.
...SLLRAPPCTTPGKSSSLSTSLPLG...LPQKPSQQRPQVILEPGHSLDWAHLNRTPQRHKLGRV.
...STLRVPPGMRPSTTNSNSSLSPNS...VTLPKPVKASQRAVILEPGHSLDWAHLNRTPQRHKLGRV.
...ASLRVPP...SRPASNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
...ASLRVPP...SRPAPNNLTAPI...LTLFQSKAKAARVILEPGHSLDWAHNLNRTPQRHKLGRV.
Sequence alignment of N-terminal region and cytochrome b5 domain of human Ncb5or against cytochrome b5s in parasites and protists.