PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The C-terminal extension of human RTEL1, mutated in Hoyeraal-Hreidarsson syndrome, contains Harmonin-N-like domains

Guilhem Faure,[1] Patrick Revy,[2] Michael Schertzer,[3] Arturo Londono-Vallejo,[3] and Isabelle Callebaut[1]*

[1] CNRS, UPMC University Paris 6, IMPMC, UMR7590-IUC, F-75005, Paris, France

[2] INSERM U768, Université Paris Descartes, Institut Imagine, Paris, France

[3] Telomeres and Cancer Laboratory (Labéllisé Ligue), Institut Curie, Paris, France

ABSTRACT

Several studies have recently shown that germline mutations in RTEL1, an essential DNA helicase involved in telomere regulation and DNA repair, cause Hoyeraal-Hreidarsson syndrome (HHS), a severe form of dyskeratosis congenita. Using original new softwares, facilitating the delineation of the different domains of the protein and the identification of remote relationships for orphan domains, we outline here that the C-terminal extension of RTEL1, downstream of its catalytic domain and including several HHS-associated mutations, contains a yet unidentified tandem of harmonin-N-like domains, which may serve as a hub for partner interaction. This finding highlights the potential critical role of this region for the function of RTEL1 and gives insights into the impact that the identified mutations would have on the structure and function of these domains.

Proteins 2014; 82:897–903.

Key words: Hoyeraal-Hreidarsson syndrome; dyskeratosis congenita; malcavernin; PAH domain; sequence analysis; remote relationship; hydrophobic cluster analysis.

INTRODUCTION

RTEL1 (regulator of telomere elongation helicase 1) is an essential DNA-helicase, involved in telomere length regulation and in DNA repair, processes that are critical for the maintenance of genome integrity. RTEL1 belongs, along with XPD (Xeroderma pigmentosum), FANCJ (Fanconi Anemia) and ChlR1 (DDX11) (Warsaw breakage syndrome), to the subclass of FeS cluster-containing DNA helicases within the SF2 DEAH subfamily.[1,2] Recently, several groups have performed whole genome linkage analyses and exome sequencing and have reported *RTEL1* germline mutations in patients with Hoyeraal-Hreidarsson syndrome (HHS) [MIM 300240], a severe variant of dyskeratosis congenita (DC) [MIM 305000] characterized by early onset bone marrow failure, immunodeficiency and developmental defects.[3–7] Some of the mutations identified in these studies are located outside the RTEL1 catalytic domain in a region lacking any known domain, as deduced from a search in the conserved domain database (CDD).[8]

We recently developed two original predictive tools for the identification and characterization of orphan sequences. The first one, called SEG-HCA, delineates foldable domains, that is, domains which may form stable 3D structures, from the only knowledge of a single amino acid sequence.[9] It is particularly useful to cut a protein sequence into individual domains and subsequently specifically search for similarities with each of them, thereby avoiding bias towards more conserved domains (in the case of RTEL1, the N-terminal helicase domain). Furthermore, we also developed another tool, called TREMOLO-HCA, for more easily detecting hidden relationships between remote sequences. To that aim, TREMOLO-HCA adds to the results of sequence

Grant sponsor: Agence Nationale de la Recherche (ANR TELO&DICENs), Institut du Cancer (INCA DI-REP).
*Correspondence to: Isabelle Callebaut, IMPMC, UMR7590, CNRS, Université Pierre et Marie Curie-Paris 6, Case 115, 4 Place Jussieu, 75252 Paris Cedex 05, France. E-mail: isabelle.callebaut@impmc.upmc.fr
Received 13 July 2013; Revised 13 September 2013; Accepted 26 September 2013
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24438

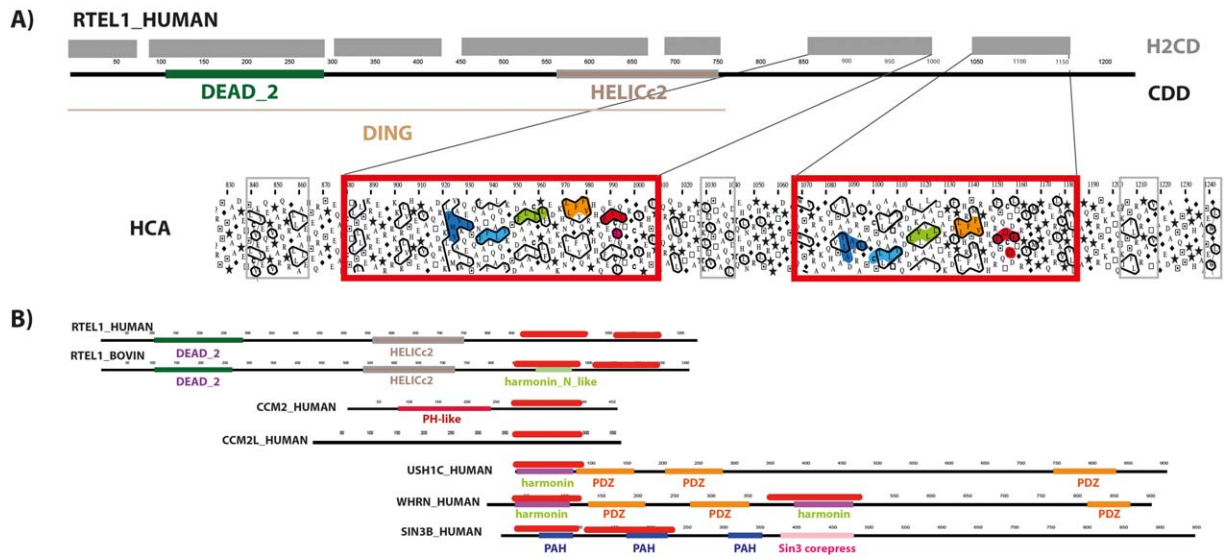

Figure 1

A hidden harmonin-N-like domain in the C-terminal region of human RTEL1. (**A**) Domain architecture of human RTEL1, as assigned by RPS-BLAST (CDD) and predicted by SEG-HCA (H2CD). Only the H2CD regions longer than 25 amino acids are reported. The HCA plots of the two C-terminal orphan H2CD regions are shown (boxed), highlighting the hydrophobic cluster similarities shared by these repeats. (B) Extract of the significant results from the TREMOLO-HCA output, obtained using the aa 856–954 fragment of human RTEL1 as query and highlighting the duplication and the relationship to harmonin-N-like and PAH domains. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

similarities searches information on domain architecture of the aligned sequences, as well as on the conservation of the hydrophobic core predicted for the query sequence.[10]

Here, using these sensitive tools, we show that the C-terminal region of RTEL1 contains a duplicated small domain, which is likely to adopt a harmonin-N-like/PAH helical fold and to serve as a hub for interaction with partners. This finding highlights the potential critical role of this region for the function of RTEL1 and gives insights into the impact that the identified mutations would have on the structure and function of these domains. In particular, a supersite, that is, a similar structural location involved in binding,[11] was identified between the two harmonin-N-like and PAH folds. This supersite includes one of the amino acids with identified HHS mutation, which thus might play a critical role in the yet unknown function of this domain.

MATERIAL AND METHODS

Sequence analysis

Sequence analysis was performed using two tools we recently developed, inspired from experience in deciphering hidden relationships within protein sequences. The first one, SEG-HCA, allows the automatic delineation of foldable regions from the only consideration of a single protein sequence.[9] SEG-HCA relies on the identification of regions with high densities in hydrophobic clusters (as defined through hydrophobic cluster analysis[12,13]), which mainly correspond to regular secondary structures forming together globular domains. These regions are called H2CD, after high hydrophobic cluster density. The reliability of the SEG-HCA predictions was assessed in comparison with known 3D structures.[9]

The second tool, TREMOLO-HCA,[10] facilitates the identification of hidden relationships for these globular domains. It adds to sequence similarity searches, performed using current tools (e.g., PSI-BLAST[14]), information on the domain architecture of the sequences aligned with the query and on the conservation rate of predicted core-forming residues.

The CDD[8] was searched for known domains. HHPRED was also used,[15] to assess further the observed relationships. Protein sequence alignments were rendered using ESPript.[16]

Molecular modeling and 3D structure visualization

A model of the 3D structure of the first harmonin-N-like domain of human RTEL1 has been constructed using Modeller (v9.10),[17] the alignment shown in Figure 2(A) and, as template, the 3D structure of human harmonin N-terminal domain (pdb 2kbr).

One thousand models were generated, which were evaluated and selected considering the DOPE[18] and ProQ[19] (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) scores. The local reliability was evaluated using VERIFY3D[20] (http://nihserver.mbi.ucla.edu/Verify_3D/), whereas the stereochemical quality was evaluated with PROCHECK.[21] 3D structures were manipulated using Chimera.[22]

RESULTS AND DISCUSSION

Identification of a duplicated domain in the C-terminal extension of RTEL1 and relationship to harmonin-N-like domains

Using SEG-HCA (see Material and Methods), we identified several segments in the human RTEL1 sequence predicted to consist in globular domains. Those found in the N-terminal region of human RTEL1 match the DEAD_2 (pfam06733) and Helicase_C_2 (pfam13307) profiles, as described in CDD. Both these regions are included in a multi-domain profile, referred to as DinG (COG1199, Rad3-related helicase) and correspond to the whole catalytic domain [Fig. 1(A)]. Downstream of this domain SEG-HCA highlighted the presence of two regions with high density in hydrophobic clusters (i.e., H2CD: 856–954 and H2CD: 1045–1159) within the RTEL1 isoform2 (UniProt Q9NZ71.2; 1219 amino acids), one of the two main isoform expressed in human cells.[4] Note several RTEL1 mRNA variants have been annotated in databases leading to differences in amino acid numbering. Here, we adopt the nomenclature found in UniProt (Q9NZ71.2: RTEL1 isoform 2, 1219 amino acids) used in Ref. 4, as opposed to the one reported in Refs. 3 and 5. Not surprisingly, these two putative domains do not match any profile in the CDD [Fig. 1(A)]. A further analysis of these two domains using TREMOLO-HCA, which adds to sequence similarity searches information on domain architecture, showed that: (i) these domains arose from a duplication and (ii) they can be related to harmonin-N-like domains (cd07357) found in harmonin, delphinin and whirlin, which are PDZ-containing scaffold proteins [Fig. 1(B)].

The TREMOLO-HCA analyses were performed on the results of PSI-BLAST run using as queries both domains predicted by SEG-HCA. The hidden relationships were detected in the significant hits obtained after eight PSIBLAST iterations. Thereby, we highlighted the duplication on the RTEL1 sequences from various species [duplicated red bar, one of which corresponding to the query sequence, Fig. 1(B) and 2(A)]. Moreover, the relationship to harmonin-N-like domains, as defined from the CDD, can be identified on the sequences of RTEL1 from *Equus caballus* and *Bos Taurus* [Figs. 1(B) and 2(A)]. For these sequences, only the first repeat is identified from CDD, while the second remains undetected.

The relationship to harmonin-N-like domains is clearly identified by the significant alignments of the query sequence with regions assigned as such from CDD. These are included in proteins such as harmonin (USH1C; PSI-BLAST *E*-values 5×10^{-8} (1st domain) and 7×10^{-5} (2nd domain), whirlin (WHRN; PSI-BLAST *E*-values 8×10^{-11} (1st domain) and 7×10^{-9} (2nd domain)) and delphilin (GRD2I; PSI-BLAST *E*-values 5×10^{-10} (1st domain) and 2×10^{-8} (2nd domain)) [Figs. 1(B) and 2(A)]. The harmonin-N-like domain adopts an alpha-helical fold, comprising five alpha-helices[23,24] [Fig. 2(B)]. A significant relationship is also detected with a region located in the C-terminal end of malcavernin, also known as CCM2 (after cerebral cavernous malformations protein 2; PSI-BLAST *E*-value 2×10^{-13} (1st domain) and 4×10^{-14} (2nd domain)) [Figs. 1(B) and 2(A)], whose 3D structure has just been solved and shown to correspond to a harmonin-N-like domain fold as well.[25]

Remarkably, positions in which hydrophobicity is conserved within the multiple alignments deduced from the PSI-BLAST significant pairwise alignments [squares in Fig. 2(A)] mainly correspond to amino acids that are buried in the harmonin-N-like domain 3D structures and constitute the hydrophobic core of the fold [Fig. 2(B)]. This further supports the observed relationships of RTEL1 with harmonin-N-like domains.

Relationship to PAH domains: A common fold and a shared hydrophobic core between harmonin-N-like and PAH domains

Our TREMOLO-HCA analysis also demonstrated that these harmonin-N-like domains are related to paired amphipathic helix (PAH) domains found in yeast Sin3 (PSI-BLAST *E*-values 1×10^{-14} (1st domain) and 7×10^{-15} (2nd domain)) and rodent Gon4L (PSI-BLAST *E*-value 5×10^{-4} (2nd domain)) [Figs. 1(B) and 2(A)]. The four first helices of harmonin-N-like domains are indeed well superimposed with those constituting the entire PAH domain fold.[26] This can be observed from a structural alignment perfectly matching the sequence alignment observed in the PSI-BLAST results [Fig. 2(A,B)]. The structural relationship between PAH and harmonin-N-like domain has been already noticed when harmonin-N-like domain was searched for related fold using DALI, in spite of the fact that no significant sequence similarity was detected at that time.[23] Again, positions in which hydrophobicity is conserved within the multiple alignments deduced from the PSI-BLAST significant pairwise alignments [squares in Fig. 2(A)] mainly correspond to amino acids that are buried in the two 3D structures and constitute the hydrophobic core of the folds [Fig. 2(B)]. Thus, not only significant statistical values clearly assess the relationship of RTEL1 with harmonin-N-like and PAH domains, but also the observation of noticeable

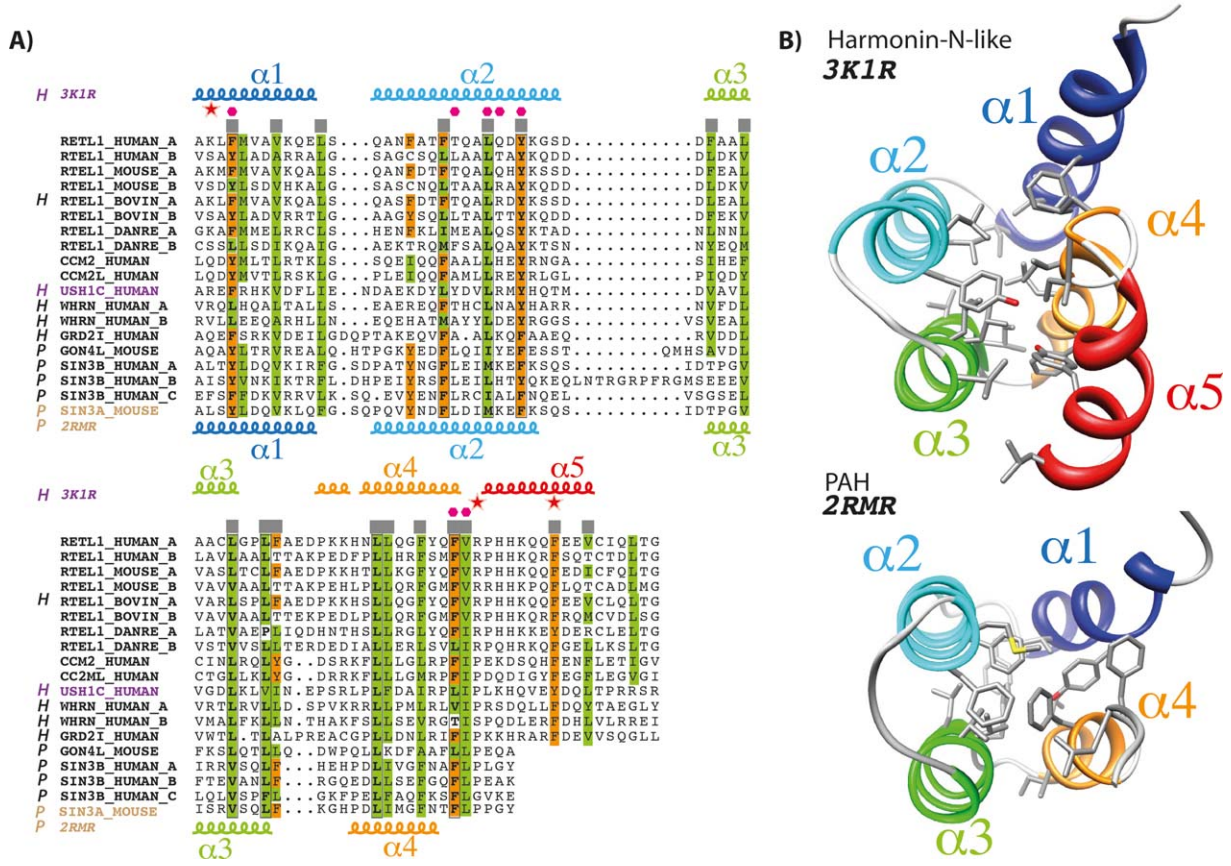

Figure 2

Sequence and structure relationship between RTEL1, harmonin-N-like domains and PAH domains. (**A**) Sequence alignment of the hidden harmonin-N-like domains from human RTEL1 and human malcavernin with cognate harmonin-N-like (H) and PAH (P) domains. Positions in which hydrophobicity is conserved are colored green (general case) or orange when the aromatic character is prevalent. UniProt identifiers and limits of domains: human RTEL1: Q9NZ71.2 (896–973 and 1060–1137), mouse RTEL1: Q0VGM9 (896–973 and 1044–1121), bovine RTEL1: A4K436 (899–976 and 1063–1140), *Danio rerio* RTEL1: P0C928 (907–984 and 907–1141), human CCM2 (malcavernin): Q9BSQ5 (299–374), human CCM2L: Q9NUG4 (415–490), human USH1C (harmonin): Q9Y6N9 (6–73; pdb: 3k1r, 2kbr), human WHRN (whirlin): Q9P202 (49–115 and 423–500), human GRD2I (dephilin): A4D2P6 (137–215), mouse GON4L: Q9DB00 (1655–1722), human Sin3B: O75182 (44–108, 168–239, and 307–370), mouse Sin3A: Q60520 (126–190; pdb: 2rmr, 2ld7, 2l9s). The positions of the three amino acids mutated in HHS are indicated with stars above the alignment, whereas amino acids involved in cadherin 23 binding in harmonin are highlighted with hexagons. (**B**) 3D structures of the harmonin-N-like domain from human harmonin (pdb 3k1r) and the PAH domain from mouse sin3A (pdb 2rmr), highlighting their relatedness. Are shown in atomic details the amino acids occupying positions for which hydrophobicity is conserved in the alignment shown in panel A (grey squares). The top of the structure displays a hydrophobic groove, interacting with ligands (see left views on Fig. 3). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

conservation of the residues participating in the hydrophobic cores. The relationship of RTEL1 to the harmonin-N-like and PAH families can also be highlighted using profile-profile comparisons by HH-Pred. Indeed, probabilities of 99.7 and 97.5–94.9 were observed for the first harmonin-N-like domain of RTEL1 with malcavernin (pdb 4fqn) and PAH domains (pdb 2czy, 2cr7, 1g&e, 1e91, 2ld7), respectively, whereas the probability associated with harmonin-N-like domain is slightly lower (74.5 with pdb 3klr). The second harmonin-N-like domain of RTEL1 has similar *P*-values, even higher for the link with harmonin-N-like domains (malcavernin: 99.9, PAH domains: 98.3–97.3, harmonin-N-like domain: 94.6).

Molecular modeling of the RTEL1 harmonin-N-like domains

The conserved hydrophobic core found in the 3D structures of harmonin-N-like domains supports a modeling approach, even sequence identities are low.[27] Models of the harmonin-N-like domains of human RTEL1 were generated, and their quality assessed using current tools (see Material and Methods). As the two RTEL1 concerned segments have, as harmonin-like domains, five alpha-helices, we used as template the harmonin-N-like domain of harmonin (pdb 2kbr), sharing with the RTEL1 domains 16.25 and 15.00% of sequence identity. LGscore and MaxSub, as calculated by ProQ, are 1.97

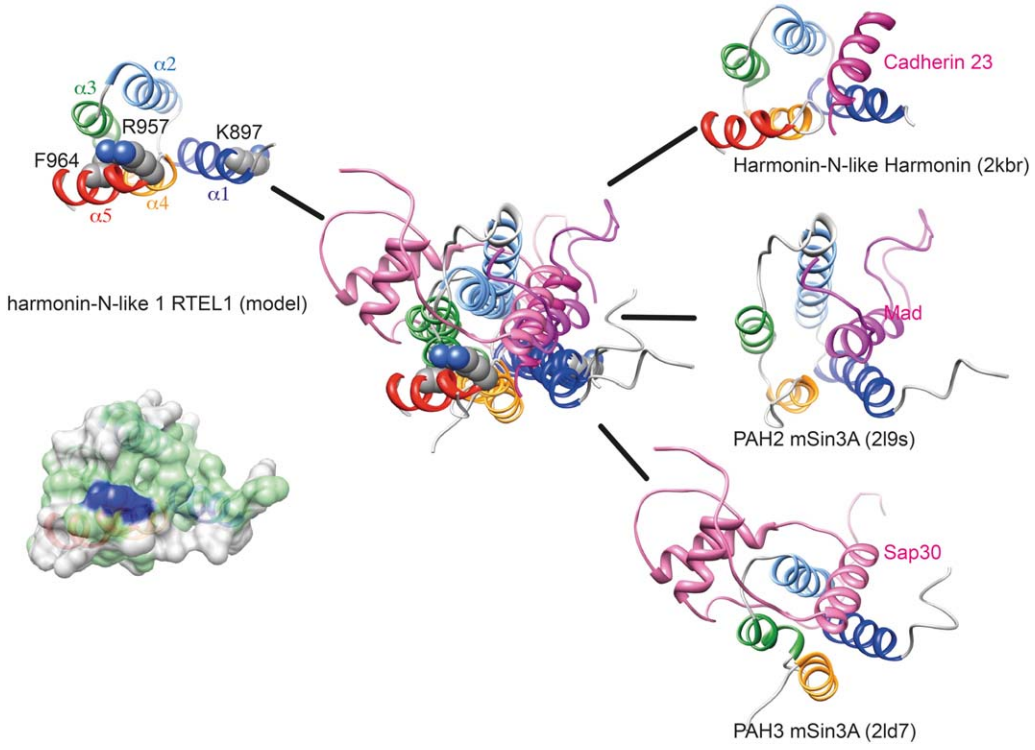

Figure 3

Positions of the HHS-causing mutations on the model of the 3D structure of the first RTEL1 harmonin-N-like domain, compared to experimental structures in complex with partners. The model of the 3D structure of the first harmonin-N-like domain of human RTEL1 is shown on the left (ribbon representation—top, solvent accessible surface—bottom). The model has been superimposed (middle panel) to the experimental 3D structures of the human harmonin-N-terminal domain in complex with a cadherin 23 peptide (pdb 2kbr[23]), of the PAH2 domain of mouse Sin3A in complex with Pf1 (pdb 2l9s[33]) and of the PAH3 domain of mouse Sin3A in complex with Sap30 (pdb 2l7d[34]) (at right). The regular secondary structures of the harmonin-N-like and PAH domains are colored according to Figs. 2(A,B), the ligands being colored in pink. The positions of the three amino acids mutated in HHS are reported on the 3D structure at left. R957 is visible in blue on the solvent accessible surface of the first RTEL1 harmonin-N-like domain (bottom left), whereas hydrophobic amino acids are depicted in green.

and 0.44 for the first RTEL1 harmonin-N-like domain (16.25% sequence identity with 2kbr) and 3.04 and 0.70 for the second RTEL1 harmonin-N-like domain (15% sequence identity with 2kbr), a range of scores actually observed for correct models.

For further assessment, we modeled in the same conditions the malcavernin using as template the same 3D structure (harmonin-N-like domain of harmonin, pdb 2kbr). LGscore (2.0) and Maxsub (0.44) were in the same range than those observed for the RTEL1 model made with the same experimental 3D structure, and the RMSD between the modeled and experimental structure (pdb 4fqn) of the malcaverin harmonin-N-like domain is 0.97 Å (15% sequence identity), thus providing support to our modeling approach for RTEL1, even though performed at low level of sequence identity.

Finally, we also modeled the first RTEL1 harmonin domain on the malcavernin template (pdb 4fqn, 20% sequence identity, LGscore = 3.31 and Maxsub 0.71); the resulting RTEL1 models (made on the two different templates) showed 1.22 Å RMSD (63 C-alpha atoms). Thus, convergence towards similar models is observed, starting from templates sharing low levels of sequence identity.

The harmonin-N-like domains of RTEL1 are likely involved in intermolecular interactions

The common theme shared by harmonin-N-like and PAH domains is their inclusion in scaffold proteins and their ability to act as protein–protein interaction modules. The harmonin-N-like domain of Harmonin, which belongs to the master usher protein complex scaffold, is indeed able to bind other usher proteins, including Cadherin 23 and SANS.[23] In CCM2 (malcavernin), a crucial regulator of heart and vessel formation and integrity[28,29] mutated in human vascular malformations,[30,31] the role of the harmonin-N-like domain is not yet understood at the molecular level, but it appears to be critical for inducing cell death through a pathway involving the TrkA receptor tyrosine kinase.[32] The PAH domains have been largely characterized in the two closely related Sin3 proteins, which belong to a large

corepressor complex involved in (i) transcription regulation, (ii) heterochromatin silencing, and (iii) some DNA repair pathways.[26] The PAH domains of the Sin3 proteins are involved in the assembly of the corepressor complex and in the recruitment of diverse transcription factors.

Interestingly, examination of the whole set of 3D structures from both families of domains revealed that their modes of interaction with partners are strikingly similar, involving a deep hydrophobic cleft of the surface of the domain, which bind helical segments (Fig. 3). This hydrophobic cleft may thus constitute a "supersite," within a superfold,[11] consisting of the common four helices of harmonin-N-like and PAH domains. The observation reported here of hidden harmonin-N-like domains in RTEL1 sheds thus light into the roles that these domains may play in this protein as molecular scaffolds and provides clues to understand the molecular basis of their interactions with partners.

Possible impact of RTEL1 *germline mutations*

Five *RTEL1* germline mutations linked to HHS[3,5] affect amino acids belonging to the harmonin-N-like domain repeat of human RTEL1. Three of them [p.K897E, p.R957W, and p.F964L, marked with stars in Fig. 1(C)] belong to the first harmonin-N-like domain of RTEL1, whereas the two other ones (p.R974Ter and p.R986Ter) are included in a nonstructured segment linking the two RTEL1 harmonin-N-like domains. The two last mutations lead to a premature termination of the polypeptidic chain, which then would lack both the second harmonin-N-like domain and the PCNA-interacting protein motif (PIP). p.F964 corresponds to a buried position of the first harmonin-N-like domain, for which the hydrophobicity is conserved among sequences of the family. Its substitution to leucine may directly affect the folding or the fold stability, even though the hydrophobic character is not lost. In contrast, p.K897, substituted by E, and p.R957, substituted by W, are exposed to solvent and may thus be available for protein–protein interaction. In support of this hypothesis is that p.R957 is included in the super-binding site, defined from the observation of complexes of harmonin-N-like or PAH domains with partners [see above; Fig. 2(A) (the amino acids of harmonin involved in cadherin 23 binding are indicated with hexagons) and Fig. 3]. Moreover, p.R957 belongs to a region which is well conserved when comparing the sequences of the two tandem domains of RTEL1 [Fig. 2(A)] and is included in a large hydrophobic area on the surface of the 3D structure (Fig. 3), thus constituting a potential binding site for the RTEL1's partners. It is worth noting that most of the amino acids, especially hydrophobic ones, participating in ligand binding in harmonin [hexagons in Fig. 2(A)], and, by extension, in PAH domains, are well conserved in the RTEL1 harmonin-N-like domains.

CONCLUSION

The only information available today for the C-terminal part of RTEL1 is the presence of a PCNA (proliferating cell nuclear antigen) interacting protein (PIP) box, located between the tandem of harmonin-N-like domains, and a RING-finger that we discovered in the C-terminal extension of the RTEL1 1300 variant.[4] Both motifs (PIP box and RING finger) are also expected, as the harmonin-N-like tandem, to bind specific partners, which remain to be identified experimentally. Interestingly, a HHS mutation (p.C1244R) has also been found in one of the zinc-binding amino acids of the RTEL1, suggesting that this mutation would have a profound impact on the functional properties of this domain.[4] The C-terminal region of RTEL1 thus includes multiple domains, which may regulate the helicase activity through the binding of different partners. The present works offers thus clues for further investigations aiming at defining the identity of binding partners of these newly identified RTEL1 harmonin-N-like domains and their role in the RTEL1 function. While this manuscript was in press, an article, published by Vannier and colleagues in Science (2013)342:239–242, established that RTEL1 associates with PCNA through its PIP box.

ACKNOWLEGMENTS

Patrick Revy is a scientist from CNRS. The “Telomere and Cancer Lab” is labellisé Ligue.

REFERENCES

1. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem 2007;76:23–50.
2. Uringa EJ, Youds JL, Lisaingo K, Lansdorp PM, Boulton SJ. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. Nucleic Acids Res 2011;39:1647–1655.
3. Ballew BJ, Yeager M, Jacobs K, Giri N, Boland J, Burdett L, Alter BP, Savage SA. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in dyskeratosis congenita. Hum Genet 2013; 132:473–480.
4. Le Guen T, Jullien L, Touzot F, Schertzer M, Gaillard L, et al. Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. Hum Mol Genet 2013; 22:3239–3949.
5. Walne AJ, Vulliamy T, Kirwan M, Plagnol V, Dokal I. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. Am J Hum Genet 2013;92:448–453.
6. Ballew BJ, Joseph V, De S, Sarek G, Vannier JB, Stracker T, Schrader KA, Small TN, O'Reilly R, Manschreck C, Harlan Fleischut MM, Zhang L, Sullivan J, Stratton K, Yeager M, Jacobs K, Giri N, Alter BP, Boland J, Burdett L, Offit K, Boulton SJ, Savage SA, Petrini JH. A recessive founder mutation in regulator of telomere elongation helicase 1, RTEL1, underlies severe immunodeficiency and features of Hoyeraal Hreidarsson syndrome. PLoS Genet 2013;9:e1003695.

7. Deng Z, Glousker G, Molczan A, Fox AJ, Lamm N, Dheekollu J, Weizman OE, Schertzer M, Wang Z, Vladimirova O, Schug J, Aker M, Londoño-Vallejo A, Kaestner KH, Lieberman PM, Tzfati Y. Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. Proc Natl Acad Sci USA 2013;110:E3408–E3416.
8. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 2013;41:D348–D352.
9. Faure G, Callebaut I. A comprehensive repertoire of foldable segments within genomes. PLoS Comput Biol, in press.
10. Faure G, Callebaut I. Identification of hidden relationships from the coupling of hydrophobic cluster analysis and domain architecture information. Bioinformatics 2013;29:1726–1733.
11. Russell RB, Sasieni PD, Sternberg MJ. Supersites within superfolds. Binding site similarity in the absence of homology. J Mol Biol 1998; 282:903–918.
12. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 1997;53:621–645.
13. Gaboriaud C, Bissery V, Benchetrit T, Mornon JP. Hydrophobic cluster analysis: an efficient new way to compare and analyze amino acid sequences. FEBS Lett 1987;224:149–155.
14. 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;25:3389–3402.
15. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33:W244–W248.
16. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 2003;31:3320–3323.
17. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 2000;29:291–325.
18. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–2524.
19. Wallner B, Elofsson A. Can correct protein models be identified? Protein Sci 2003;12:1073–1086.
20. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK—a program to check the stereochemical quality of protein structures. J App Cryst 1993;26:283–291.
22. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
23. Pan L, Yan J, Wu L, Zhang M. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. Proc Natl Acad Sci USA 2009;106:5575–5580.
24. Yan J, Pan L, Chen X, Wu L, Zhang M. The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins. Proc Natl Acad Sci USA 2010;107: 4040–4045.
25. Fisher OS, Zhang R, Li X, Murphy JW, Demeler B, Boggon TJ. Structural studies of cerebral cavernous malformations 2 (CCM2) reveal a folded helical domain at its C-terminus. FEBS Lett 2013; 587:272–277.
26. Sahu SC, Swanson KA, Kang RS, Huang K, Brubaker K, Ratcliff K, Radhakrishnan I. Conserved themes in target recognition by the PAH1 and PAH2 domains of the Sin3 transcriptional corepressor. J Mol Biol 2008;375:1444–1456.
27. Gao J, Li Z. Uncover the conserved property underlying sequence-distant and structure-similar proteins. Biopolymers 2010;93:340–347.
28. Kleaveland B, Zheng X, Liu JJ, Blum Y, Tung JJ, Zou Z, Sweeney SM, Chen M, Guo L, Lu MM, Zhou D, Kitajewski J, Affolter M, Ginsberg MH, Kahn ML. Regulation of cardiovascular development and integrity by the heart of glass-cerebral cavernous malformation protein pathway. Nat Med 2009;15:169–176.
29. Whitehead KJ, Chan AC, Navankasattusas S, Koh W, London NR, Ling J, Mayo AH, Drakos SG, Jones CA, Zhu W, Marchuk DA, Davis GE, Li DY. The cerebral cavernous malformation signaling pathway promotes vascular integrity via Rho GTPases. Nat Med 2009;15:177–184.
30. Akers AL, Johnson E, Steinberg GK, Zabramski JM, Marchuk DA. Biallelic somatic and germline mutations in cerebral cavernous malformations (CCMs): evidence for a two-hit mechanism of CCM pathogenesis. Hum Mol Genet 2009;18:919–930.
31. Pagenstecher A, Stahl S, Sure U, Felbor U. A two-hit mechanism causes cerebral cavernous malformations: complete inactivation of CCM1, CCM2 or CCM3 in affected endothelial cells. Hum Mol Genet 2009;18:911–918.
32. Harel L, Costa B, Tcherpakov M, Zapatka M, Oberthuer A, Hansford LM, Vojvodic M, Levy Z, Chen ZY, Lee FS, Avigad S, Yaniv I, Shi L, Eils R, Fischer M, Brors B, Kaplan DR, Fainzilber M. CCM2 mediates death signaling by the TrkA receptor tyrosine kinase. Neuron 2009;63:585–591.
33. Kumar GS, Xie T, Zhang Y, Radhakrishnan I. Solution structure of the mSin3A PAH2-Pf1 SID1 complex: a Mad1/Mxd1-like interaction disrupted by MRG15 in the Rpd3S/Sin3S complex. J Mol Biol 2011; 408:987–1000.
34. Xie T, He Y, Korkeamaki H, Zhang Y, Imhoff R, Lohi O, Radhakrishnan I. Structure of the 30-kDa Sin3-associated protein (SAP30) in complex with the mammalian Sin3A corepressor and its role in nucleic acid binding. J Biol Chem 2011;286:27814–27824.