

Insights into the molecular features of the von Hippel-Lindau like protein

Federica Quaglia^{1#}, Giovanni Minervini^{1#}, Francesco Tabaro^{1,2}, Silvio C.E. Tosatto^{1,3*}

¹*Department of Biomedical Sciences, University of Padova, Viale G. Colombo 3, 35121, Padova, Italy.*

²*Institute of Biosciences and Medical Technology, Tampere, Finland*

³*CNR Institute of Neuroscience, Padova, Italy*

[#]*contributed equally*

^{*}*Corresponding Author: silvio.tosatto@unipd.it*

Keywords

Von Hippel-Lindau disease, hereditary neoplastic syndrome, cancer, bioinformatics

Abstract

We present an *in silico* characterization of the von Hippel-Lindau like protein (VLP), the only known human paralogue of the von Hippel-Lindau tumor suppressor protein (pVHL). Phylogenetic investigation showed VLP to be mostly conserved in upper mammals and specifically expressed in brain and testis. Structural analysis and molecular dynamics simulation show VLP to be very similar to pVHL three-dimensional organization and binding dynamics. In particular, conservation of elements at the protein interfaces suggests VLP to be a functional pVHL homolog potentially possessing multiple functions beyond HIF-1 α -dependent binding activity. Our findings show that VLP may share at least seven interactors with pVHL, suggesting novel functional roles for this understudied human protein. These may occur at precise hypoxia levels where functional overlap with pVHL may permit a finer modulation of pVHL functions.

Introduction

Cellular specialization arises from finely tuned interactions between different molecular components, such as proteins, DNA, RNA, transcription factors, degradative machinery and external factors [1]. This complex organization allows coordination of several metabolic networks yielding cellular metabolism. Oxygen sensing is an example of a biological network aimed at promoting metabolic adaptation to environmental variations, as is a well-characterized molecular pathway triggering complex responses to hypoxia [2,3]. Oxygen variation is perceived by the cell through interactions of highly specialized proteins: the von Hippel-Lindau tumor suppressor protein (pVHL), Hypoxia Inducible Factor 1 α (HIF-1 α) and prolyl-hydroxylase protein family (PHDs), which form the so-called core hypoxia response network [4]. At physiological oxygen concentrations, PHDs catalyze

the hydroxylation of two specific HIF-1 α prolines (P402 and P564) [5,6]. Hydroxylated HIF-1 α is then targeted for proteasomal degradation by pVHL, an E3 ubiquitin ligase complex substrate recognition element [7]. Hypoxia inhibits PHD activity, impairing HIF-1 α and pVHL recognition, promoting HIF-1 α stabilization [8]. Once stabilized, HIF-1 α is translocated to the nucleus, where it activates Hypoxia response elements (HRE) and promotes hypoxia-regulated gene expression [9]. Deregulation of this network is known to predispose to cancer onset, as in von Hippel-Lindau syndrome [10]. Biological networks, beyond their complexity, are also characterized by robustness [11], a fundamental property which allows the cell to react to internal and external perturbations (e.g. hypoxia, nutrient variation, DNA mutations, etc.) [11]. A bow-tie architecture is frequently used to describe biological networks [11,12]. This representation implies few highly connected hub proteins forming a conserved core, while a number of proteins and various stimuli form the network input/output. Elements in the conserved core are associated with the main tumor suppressors and oncogenes, such as p53 [13,14], pVHL [15] and c-Myc [16] just to name a few. Another property directly derived from a bow-tie network organization is the appearance of specialized protein hub isoforms (or alternative genes) presenting slightly different functions. The p53-like protein family is an example of this functional overlap within members of the same family [17]. p53, p63 and p73 are known to form a transcription factor family [18], acting as main regulators of cell cycle progression. Although partially redundant, each member has its own unique function, varying between coordination of cell differentiation, stress response, functional apoptosis and aging [17]. From a biological network perspective, functional overlap can be seen as a natural mechanism to improve network robustness, i.e. together replicating crucial nodes while allowing evolution of new functions. In 2004, the von Hippel-Lindau like protein (VLP), a human paralog of the VHL gene, was reported in the literature [19]. VLP contains 139 amino acids, corresponding to the pVHL β -domain, while the pVHL E3-component recognition domain, required for pVHL degradative function, is missing. The first characterization attempts revealed that VLP binds HIF-1 α and suggested it may act as a dominant-negative VHL, protecting HIF-1 α from degradation [19]. Three different pVHL isoforms [20] are already known in the literature. Although pVHL was described to be directly involved in several different pathways interacting with a number of different interactors [15], the exact role of each different isoform and paralogous VLP is far from being clearly understood. Here, we present an *in silico* characterization of human VLP. Phylogenetic analysis was paired with *in silico* structural investigations. Our analysis showed that VLP is expressed in specific human tissues and shares interactors with pVHL. Collectively, our findings suggest novel functional roles for this human pVHL paralog.

Materials and Methods

Expression data

Expression profile data was retrieved by combining different database searches. Bgee [21] was used to extract and compare gene expression patterns between homologous genes from different species and CleanEx [22] to extract specific entries according to their biological annotation. Ensembl [23], GeneCards [24], NextProt [25], GTEx Portal [26] and Gene Expression Atlas [27] were used to retrieve human gene expression data at the tissue level as well as to extract data for gene regulation and genetic variation.

Sequence feature and interactor analysis

Orthologous VLP sequences from different species were retrieved in OMA Browser [28], OrthoDB [29], EggNOG [30], PhylomeDB [31] and enriched by manual Blast [32] search against the NR database. The extracted dataset was then manually curated to remove pVHL-specific homologous sequences and aligned with Jalview [33] to derive a conservation score for each residue. The final conservation score and chemical properties were considered as a cutoff for protein-protein interaction surface prediction. A VLP interaction network was derived merging data from the BIOGRID [34], STRING [35] and IntAct [36]. Putative VLP molecular functions were predicted with INGA [37], while dSysMap [38], COSMIC [39] and PolyPhen-2 [40] were used to address VLP role in human disease onset.

Phylogenetic analysis

A bayesian analysis was performed with MrBayes 3.2.6 [41] using the Jones-Taylor-Thornton model of protein evolution plus invariant sites, JTT + I, identified with MEGA 7 [42], with 4 chains (3 heated chains and 1 cold chain) and temperature 0.1 and 1,000,000 generations. The tree layout and bootstrap values at the nodes of the tree were visually inspected with FigTree (<http://tree.bio.ed.ac.uk/>).

Homology modeling

The crystal structure of pVHL bound to the CODD fragment of HIF-1 α (PDB code: 1LM8) was selected as a template to model the VLP structure. The human VLP sequence (accession code: Q6RSH7) was retrieved from UniProt [43] and aligned with T-Coffee [44] against pVHL using standard parameters. This was used for model building in Modeller [45]. The VLP-HIF-1 α dimer was modeled through structural superimposition in Chimera [46]. GROMACS 4.6.5 [47] was used to relax the model structures for 1.000 steps of steepest descent minimization. Model quality was assessed

with Qmean [48] and TAP [49]. A network of interacting residues was generated using RING 2.0 [50] and used to inspect the main structural feature of VLP.

Molecular dynamics simulation

Molecular dynamics (MD) simulations were carried out with Gromacs 4.6.5, using the CHARMM-22 force field [51] on a standard x86 Linux workstation. Hydrogen atoms were added to the system using the Gromacs pdb2gmh routine. The simulation protocol consists in a minimization step, 100 ps of NVT (constant number of molecules, volume and temperature) simulation, 100 ps of NPT (constant number of molecules, pressure and temperature) simulation and 50 ns of classical MD. Temperature for production simulations was kept at 300 K. A solvent box was generated imposing a cut-off distance of 10 Å from the farthest point of the protein boundaries in each dimension. Parameters for simulating the hydroxyproline residue were derived from [52].

Results

The human VLP is a 139 amino acid protein with 67.6% identity and 78% similarity to the pVHL region 1–157. This corresponds to the N-terminus and β -domain of pVHL [53]. VLP proposed to co-regulate HIF-1 α stability [19], however, to date its molecular function is far from understood. The results from different *in silico* analyses aimed at characterizing VLP are discussed in the following.

VLP is found in cerebellum, testis and placenta

The expression patterns associated to VLP were investigated by a virtual expression analysis searching different databases. Multiple database searches in Bgee, CleanEX and Gene Atlas confirmed the previous observations reporting the human VLP in cerebellum and placenta [19]. The Gene Cards and GTEx Portal report VLP to be also expressed in testis (Figure 1). As these databases use data from different sources, this data cannot be considered a spurious observation. In particular, GTEx Portal collects novel analyses of human tissues from donors; it could be supposed that VLP expression could be caught in the near future within other tissues beyond brain cerebellum, placenta and testis. Of note, all of these tissues present a well known susceptibility to hypoxia [54–56]. Due to its molecular similarity with pVHL, a tissue-specific role for VLP could be hypothesized.

VLP is mainly conserved among primates.

Next, we asked whether VLP is conserved among other species. To answer this question, a comparative sequences analysis was done. Retrieving orthologous sequences is particularly challenging due to high sequence identity with the pVHL protein. Although an initial search using an automatic procedure generated a list of 47 putative orthologs, manual inspection showed that only a

limited number of sequences can be correctly classified as VLP. In particular, VLP is mainly present in primates (*Homo sapiens*, *Pan paniscus*, *Nomascus leucogenys*, *Aotus nancymae*, *Rhinopithecus roxellana*, *Mandrillus leucophaeus*), with *Vicugna pacos* and *Monodelphis domestica* (the grey short-tailed opossum, a marsupial) representing the only two other mammals whose sequences are not ambiguous and clearly orthologous to human VLP (Figure 2). A multiple-sequence alignment comparing a total of 7 VHL orthologs from distinct upper mammals shows the VLP globular region, corresponding to the β -domain of pVHL, to be highly conserved ($> 85\%$ similarity). A longer primate-specific region formed by three repetitions of the xEEEx motif was found at the N-terminus (Figure 2). This sequence organization is comparable to pVHL30, the longest human pVHL isoform [57] which presents a 54 residue tail formed by five GxEEEx repetitions. Although debated, this tail was proposed to mediate isoform-specific functions presumed to have emerged during primate evolution, i.e. a further protein-protein interaction interface [57].

VLP phylogenetic analysis

Despite great efforts in retrieving orthologous sequences of human VLP from various sources, multiple sequence alignment suggested that not all “VHL-like proteins” are truly orthologous to human VLP. Indeed, most of them share a sequence and domain composition resembling VHL rather than human VLP. Due to this ambiguity, we discarded these putative sequences from the following analysis. We next performed a Bayesian inference phylogenetic analysis to better investigate the evolutionary relationships between VLP, its paralog VHL, and their different functions. Our phylogeny suggests a complex evolutionary relationship behind VLP and its paralog VHL. A group of primate VHL sequences can be identified, including *Homo sapiens*, while the VLP and VHL proteins from *Vicugna pacos* seem to form a separated group, sharing higher sequence similarity with the VHL group (Figure 3). As the Alpaca (*Vicugna pacos*) is a camelid evolved to live at higher altitudes of the South American mountains, the observed differences may reflect an adaptation to constitutive low oxygen concentrations. The analysis also highlights a well defined group including all VLP sequences from Primates, suborder Haplorhini, with a clear distinction for each family: Hominidae, including *Homo sapiens*, *Gorilla gorilla*, *Pan paniscus* and *Pongo abelii*, and Hylobatidae represented by *Nomascus leucogenys*. On the other hand, we found no trivial localization for *Monodelphis domestica* as its putative VLP shares features with both VHL and VLP groups. These findings, paired with the high sequence similarity shared by the VHL β -domain and VLP sequences, suggest that one (or more) remote gene duplications during mammalian evolution could easily explain the differences observed between VLP sequences from distant groups. Further genetics characterization will help to clarify this point.

VLP presents a structured β -domain adorned by a long disordered tail

A VLP homology model, generated starting from pVHL structure (PDB code 1LM8 [58]), shows a clear pVHL β -domain fold, completely lacking the elongin binding α -domain with a long disordered N-terminus (Figure 4). Superimposition of the VLP model and pVHL crystal structure highlights full conservation (100% identity) for the β 2- and β 6- strands and a short loop in pVHL. The model quality calculated with QMEAN is 0.649, while TAP estimates a 0.7569 confidence value, suggesting that our VLP model is good. Lower reliability was estimated for the N-terminal tail. MobiDB [59] confirmed the disordered behavior of the region. We then decided to perform molecular dynamics simulations to verify VLP fold stability. A 50 ns MD simulation showed that the model is stable (average RMSD 1.8 Å and 2.5 Å for the globular domain and entire protein, respectively). Of note, transient short α -helix segments were noted during simulations in the disordered tail, suggesting that this region is dynamically active and prone to explore different conformations.

VLP interacts with HIF-1 α through interface B

Although interaction between VLP and HIF-1 α is not completely new, driving molecular details remain unknown. Due to its similarity with pVHL the interaction was proposed to rely on similar residues [19]. Our structural model shows amino acids substitutions at the putative HIF-1 α interaction interface are present which may interfere with VLP/HIF-1 α interaction. The corresponding VLP complex was modeled by superimposition with the pVHL/HIF-1 α complex crystal structure. Two independent MD simulations were performed to test complex stability with and without HIF-1 α hydroxylation (Figure 5). These showed VLP interacting with HIF-1 α through the β 3- and β 4 strands. The binding is maintained by electrostatic interactions between VLP residues and the HIF-1 α region comprising residues 561-565. The remaining residues appear to follow a different fate depending on presence of hydroxylated proline 564. Simulations showed that hydroxylation of proline 564 may play a regulative role in the interaction with VLP. The hydroxylated complex is more stable during the simulation, in particular a hydrogen bond between hydroxyproline 564 and histidine 97 of VLP lower the average distance between the two proteins, yielding further hydrophobic interactions with interface B residues (Figure 5).

VLP presents conserved structural motifs suggesting overlapping pVHL functions

In a previous research, we proposed three different pVHL protein-protein interaction interfaces (A, B and C, respectively) [53]. For each interface, conserved residues are known to engage in specific interactions with a subset of pVHL binding partners [60]. We asked whether a similar surface

organization is also conserved in VLP. Considering the high sequence identity between the two proteins, we mapped putative binding regions on the VLP interface. Residues forming interfaces B and C are mostly conserved in VLP. Considering the similarity between the two proteins, at least seven pVHL interactors can be inferred for VLP (Figure 6). Putative pathways in which VLP may play a role were assigned based on these findings. VLP may be involved in regulating HIF-1 α degradation, cell migration and cell cycle regulation interacting with VDU1/USP30 and VDU2/USP20 through interface B. Both are well known deubiquitinating enzymes involved in various cellular processes [61–63]. For interface C, interactions with EP300, eEF1A, Androgen receptor and TBP1 were predicted. This suggests VLP to be involved in various pathways, such as cell cycle regulation, nuclear export, growth and development of male reproductive organs and ATP-dependent degradation of ubiquitinated proteins.

VLP mutations

COSMIC contains 37 mutated VLP samples, including 26 missense, 9 synonymous, 1 nonsense substitution and 1 non-stop gain substitutions. No other complex mutations were found. Both confirmed somatic mutations and variants of unknown origin are described. A schematic view of the distribution of these mutations on the VLP sequence and their pathogenicity predictions are presented in Table 1. VLP mutations appear to uniformly affect the entire protein. Interestingly, tissues and organs affected are largely different from those affected in the VHL syndrome (Figure 7). Considering only missense mutations, 33.6% were found in adenocarcinoma of the large intestine, 15.4% in gastric adenocarcinoma and 11.5% in endometrioid carcinoma. Others organs affected include biliary tract, breast, small intestine, kidney, liver, lung, ovary, esophagus and skin.

Discussion

Oxygen sensing in mammal cell occurs at many levels, yielding both acute and chronic adaptation. The pVHL/HIF-1 α axis is deputed to this role through fine regulation of several hypoxia responsive elements. From a biological network point of view, pVHL is a conserved core element, or a master protein, driving important cell functions. Robustness is a property of biological systems. A relevant aspect of robustness is a clear distinction between proteins acting as signals (modulators) and members of the central control (conserved core) [11]. To guarantee system stability, defined as resistance to external aggressions such as mutations and metabolic imbalance, the central core elements frequently evolve specialized isoforms with partially overlapping functions. This concept is well known for the combined functions of p53 protein family [17]. A similar assumption of functional family can be hypothesized for pVHL. Currently, three isoforms and a paralog protein are described

in human for this important tumor suppressor [7,15,20]. The two major isoforms, i.e. pVHL19 and pVHL30, show different cellular localization and the exact nature of their functional segregation remains far from understood. Fewer molecular details are available for VLP. It is thought not to possess ubiquitin ligase activity, missing the pVHL VCB-forming α -domain [53]. This clearly excludes VLP from being a ubiquitin E3 ligase. Our analysis showed this protein is mainly found in primates, where it apparently evolved structural features posing it at middle between pVHL19 and pVHL30. Human VLP is found expressed in three specific tissues characterized by relevant susceptibility to hypoxic injury. The brain, in particular, is well known for its sensibility to variation of oxygen concentration. VLP was proposed as a dominant-negative-VHL exerting a HIF-1 α protective function [19]. It is known that neuron-specific inactivation of HIF-1 α correlates with increased brain injury during transient focal cerebral ischemia in mice models [64]. In the same study, was also observed that following an initial hypoxic stimulus, HIF-1 α levels remained elevated after eight days from a normal oxygenated blood perfusion. Our MD simulations suggest that VLP forms stronger interactions with hydroxylated than unmodified HIF-1 α . We believe this data in particular may suggest a precise role for VLP in brain tissue remodeling after a pathological hypoxia insult. In this context, a putative interaction with EP300 (E1A binding protein p300) is worth mentioning. In severe anoxic conditions, p53 outcompetes HIF-1 α for EP300 [65], promoting pVHL-independent HIF-1 α degradation and leading to a demodulation of hypoxia response. VLP may play a role regulating the competitive p53/ HIF-1 α interaction with EP300. Our analysis also sheds light on HIF-1 α -independent functions, in particular the putative interaction with eEF1A (Elongation factor 1- α 1). This protein is functionally connected with the testis-specific protein Y-encoded (TSPY1) [66], a male specific protein encoded on the Y chromosome expressed in testicular tissue and correlated with spermatogenesis and gonadoblastoma insurgence. As eEF1A mediates pVHL nuclear exporting [67], our investigations suggest the same mechanism to be present in VLP. Combining these findings with VLP expression in testis, we suggest this protein to have a novel regulative role modulating eEF1A/TSPY1 interaction.

Conclusion

Our analysis shows that VLP shares structural features with the pVHL oncosuppressor in terms of structure conservation and protein-protein interaction partners. It also suggests that VLP could be a member of a larger pVHL family sharing regulative functions. pVHL is a key component of proteasome mediated degradation system, while VLP clearly miss this function. Nevertheless, it may play a relevant role in HIF-1 α stabilization. In particular its propensity to form more stable complex with hydroxylated HIF-1 α suggests an active role in the modulation of hypoxia response. VLP

appears to be structurally very similar to pVHL. Conservation of functional elements at interfaces B and C, in particular, extends the VLP range of activity to HIF-1 α -independent functions. These may occur within a precise hypoxia gradient where a finer modulation of pVHL functions would be desirable as in the case of brain remodeling after ischemia injury. The degree to which these results *in silico* apply to living organism is unclear. Experimental studies are in progress to clarify VLP functions.

Acknowledgements

The authors are grateful to members of the BioComputingUP group for insightful discussions.

Conflict of interest and funding

This work was supported by Associazione Italiana per la Ricerca sul Cancro (AIRC) grant MFAG12740 and IG17753 to ST. FT is an AIRC research fellows. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare that have no significant competing financial, professional or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

References

1. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5: 101–113. doi:10.1038/nrg1272
2. Bárdos JI, Ashcroft M. Negative and positive regulation of HIF-1: a complex network. *Biochimica et biophysica acta.* 2005;1755: 107–20. doi:10.1016/j.bbcan.2005.05.001
3. Bruick RK. Oxygen sensing in the hypoxic response pathway: regulation of the hypoxia-inducible transcription factor. *Genes Dev.* 2003;17: 2614–2623. doi:10.1101/gad.1145503
4. Minervini G, Panizzoni E, Giollo M, Masiero A, Ferrari C, Tosatto SCE. Design and analysis of a Petri net model of the Von Hippel-Lindau (VHL) tumor suppressor interaction network. *PLoS ONE.* 2014;9: e96986. doi:10.1371/journal.pone.0096986
5. Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, et al. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science (New York, NY).* 2001;292: 468–72. doi:10.1126/science.1059796
6. Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, et al. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science.* 2001;292: 464–468. doi:10.1126/science.1059817
7. Hon W-C, Wilson MI, Harlos K, Claridge TDW, Schofield CJ, Pugh CW, et al. Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature.* 2002;417: 975–8. doi:10.1038/nature00767
8. McDonough MA, Li V, Flashman E, Chowdhury R, Mohr C, Liénard BMR, et al. Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci USA.* 2006;103: 9814–9819. doi:10.1073/pnas.0601283103
9. Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, Cockman ME, et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature.* 1999;399: 271–275. doi:10.1038/20459
10. Iliopoulos O, Kibel A, Gray S, Kaelin WG. Tumour suppression by the human von Hippel-Lindau gene product. *Nat Med.* 1995;1: 822–826. doi:10.1038/nm0895-822
11. Kitano H. Biological robustness. *Nat Rev Genet.* 2004;5: 826–837. doi:10.1038/nrg1471
12. Friedlander T, Mayo AE, Tlustý T, Alon U. Evolution of Bow-Tie Architectures in Biology. *PLOS Comput Biol.* 2015;11: e1004055. doi:10.1371/journal.pcbi.1004055
13. Bisio A, Ciribilli Y, Fronza G, Inga A, Monti P. TP53 Mutants in the Tower of Babel of Cancer Progression. *Human Mutation.* 2014;35: 689–701. doi:10.1002/humu.22514
14. Lamb P, Crawford L. Characterization of the human p53 gene. *Mol Cell Biol.* 1986;6: 1379–1385.
15. Frew IJ, Krek W. pVHL: a multipurpose adaptor protein. *Sci Signal.* 2008;1: pe30. doi:10.1126/scisignal.124pe30
16. Zeller KI, Zhao X, Lee CWH, Chiu KP, Yao F, Yustein JT, et al. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc Natl Acad Sci USA.* 2006;103: 17834–17839. doi:10.1073/pnas.0604129103

17. Murray-Zmijewski F, Lane DP, Bourdon J-C. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ.* 2006;13: 962–972. doi:10.1038/sj.cdd.4401914
18. Strano S, Rossi M, Fontemaggi G, Munarriz E, Soddu S, Sacchi A, et al. From p63 to p53 across p73. *FEBS Lett.* 2001;490: 163–170.
19. Qi H, Gervais ML, Li W, DeCaprio JA, Challis JRG, Ohh M. Molecular Cloning and Characterization of the von Hippel-Lindau-Like Protein11National Cancer Institute of Canada with funds from the Terry Fox Foundation (Grant no. 13030). H.Q. is a recipient of the Canadian Institutes of Health Research (CIHR) fellowship. Note: M.O. is a Canada Research Chair in Molecular Oncology. *Mol Cancer Res.* 2004;2: 43–52.
20. Chesnel F, Hascoet P, Gagné JP, Couturier A, Jouan F, Poirier GG, et al. The von Hippel–Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. *Br J Cancer.* 2015; doi:10.1038/bjc.2015.189
21. Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M. Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species. In: Bairoch A, Cohen-Boulakia S, Froidevaux C, editors. *Data Integration in the Life Sciences*. Springer Berlin Heidelberg; 2008. pp. 124–131. doi:10.1007/978-3-540-69828-9_12
22. Praz V, Bucher P. CleanEx: new data extraction and merging tools based on MeSH term annotation. *Nucleic Acids Res.* 2009;37: D880–884. doi:10.1093/nar/gkn878
23. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. *Database (Oxford).* 2016;2016. doi:10.1093/database/bav096
24. Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, et al. In-silico human genomics with GeneCards. *Hum Genomics.* 2011;5: 709–717.
25. Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, et al. neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res.* 2012;40: D76–D83. doi:10.1093/nar/gkr1179
26. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45: 580–585. doi:10.1038/ng.2653
27. Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, et al. Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* 2010;38: D690–D698. doi:10.1093/nar/gkp936
28. Schneider A, Dessimoz C, Gonnet GH. OMA Browser--exploring orthologous relations across 352 complete genomes. *Bioinformatics.* 2007;23: 2180–2182. doi:10.1093/bioinformatics/btm295
29. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 2017;45: D744–D749. doi:10.1093/nar/gkw1119
30. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 2016;44: D286–293. doi:10.1093/nar/gkv1248
31. Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Marcet-Houben M, Gabaldón T. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 2014;42: D897–902. doi:10.1093/nar/gkt1177

32. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). Cold Spring Harb Protoc. 2007;2007: pdb.top17. doi:10.1101/pdb.top17
33. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25: 1189–1191. doi:10.1093/bioinformatics/btp033
34. Chatr-Aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015;43: D470-478. doi:10.1093/nar/gku1204
35. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41: D808-815. doi:10.1093/nar/gks1094
36. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, et al. IntAct: an open source molecular interaction database. Nucleic Acids Res. 2004;32: D452-455. doi:10.1093/nar/gkh052
37. Piovesan D, Giollo M, Leonardi E, Ferrari C, Tosatto SCE. INGA: protein function prediction combining interaction networks, domain assignments and sequence similarity. Nucleic Acids Res. 2015;43: W134–W140. doi:10.1093/nar/gkv523
38. Mosca R, Tenorio-Laranga J, Olivella R, Alcalde V, Céol A, Soler-López M, et al. dSysMap: exploring the edgetic role of disease mutations. Nat Meth. 2015;12: 167–168. doi:10.1038/nmeth.3289
39. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43: D805-811. doi:10.1093/nar/gku1075
40. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current protocols in human genetics / editorial board, Jonathan L Haines . [et al]. 2013;07: Unit7.20-Unit7.20. doi:10.1002/0471142905.hg0720s76
41. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012;61: 539–542. doi:10.1093/sysbio/sys029
42. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33: 1870–1874. doi:10.1093/molbev/msw054
43. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014;42: D191-198. doi:10.1093/nar/gkt1140
44. Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology. 2000;302: 205–217. doi:10.1006/jmbi.2000.4042
45. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci. 2007;Chapter 2: Unit 2.9. doi:10.1002/0471140864.ps0209s50
46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera a visualization system for exploratory research and analysis. J Comput Chem. 2004;25: 1605–1612. doi:10.1002/jcc.20084

47. Hess B, Kutzner C, Biophysics C, Spoel DVD, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. 2008; 435–447.
48. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Genetics*. 2008;71: 261–277. doi:10.1002/prot.21715
49. Tosatto SC, Battistutta R. TAP score: torsion angle propensity normalization applied to local protein structure evaluation. *BMC Bioinformatics*. 2007;8: 155. doi:10.1186/1471-2105-8-155
50. Piovesan D, Minervini G, Tosatto SCE. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res*. 2016; doi:10.1093/nar/gkw315
51. Patel S, Mackerell AD, Brooks CL. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *Journal of computational chemistry*. 2004;25: 1504–14. doi:10.1002/jcc.20077
52. Minervini G, Masiero A, Moro S, Tosatto SCE. In silico investigation of PHD-3 specific HIF1- α proline 567 hydroxylation: A new player in the VHL/HIF-1 α interaction pathway? *FEBS Letters*. 2013;587: 2996–3001. doi:10.1016/j.febslet.2013.07.019
53. Leonardi E, Murgia a, Tosatto SCE. Adding structural information to the von Hippel-Lindau (VHL) tumor suppressor interaction network. *FEBS letters*. 2009;583: 3704–10. doi:10.1016/j.febslet.2009.10.070
54. Busl KM, Greer DM. Hypoxic-ischemic brain injury: pathophysiology, neuropathology and mechanisms. *NeuroRehabilitation*. 2010;26: 5–13. doi:10.3233/NRE-2010-0531
55. Aplin JD. Hypoxia and human placental development. *J Clin Invest*. 2000;105: 559–560.
56. Reyes JG, Farias JG, Henr#xed, quez-Olavarrieta S, n, Madrid E, et al. The Hypoxic Testicle: Physiology and Pathophysiology, The Hypoxic Testicle: Physiology and Pathophysiology. *Oxidative Medicine and Cellular Longevity, Oxidative Medicine and Cellular Longevity*. 2012;2012, 2012: e929285. doi:10.1155/2012/929285, 10.1155/2012/929285
57. Minervini G, Mazzotta GM, Masiero A, Sartori E, Corrà S, Potenza E, et al. Isoform-specific interactions of the von Hippel-Lindau tumor suppressor protein. *Sci Rep*. 2015;5: 12605. doi:10.1038/srep12605
58. Min J-H, Yang H, Ivan M, Gertler F, Kaelin WG, Pavletich NP. Structure of an HIF-1 α -pVHL complex: hydroxyproline recognition in signaling. *Science*. 2002;296: 1886–1889. doi:10.1126/science.1073440
59. Potenza E, Di Domenico T, Walsh I, Tosatto SCE. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res*. 2015;43: D315-320. doi:10.1093/nar/gku982
60. Tabaro F, Minervini G, Sundus F, Quaglia F, Leonardi E, Piovesan D, et al. VHLdb: A database of von Hippel-Lindau protein interactors and mutations. *Scientific Reports*. 2016;6: 31128. doi:10.1038/srep31128
61. Curcio-Morelli C, Zavacki AM, Christofollete M, Gereben B, de Freitas BCG, Harney JW, et al. Deubiquitination of type 2 iodothyronine deiodinase by von Hippel-Lindau protein-interacting deubiquitinating enzymes regulates thyroid hormone activation. *J Clin Invest*. 2003;112: 189–196. doi:10.1172/JCI18348
62. Berthouze M, Venkataramanan V, Li Y, Shenoy SK. The deubiquitinases USP33 and USP20 coordinate beta2 adrenergic receptor recycling and resensitization. *EMBO J*. 2009;28: 1684–1696. doi:10.1038/emboj.2009.128

63. Shenoy SK, Modi AS, Shukla AK, Xiao K, Berthouze M, Ahn S, et al. Beta-arrestin-dependent signaling and trafficking of 7-transmembrane receptors is reciprocally regulated by the deubiquitinase USP33 and the E3 ligase Mdm2. *Proc Natl Acad Sci USA*. 2009;106: 6650–6655. doi:10.1073/pnas.0901083106
64. Baranova O, Miranda LF, Pichiule P, Dragatsis I, Johnson RS, Chavez JC. Neuron-specific inactivation of the hypoxia inducible factor 1 alpha increases brain injury in a mouse model of transient focal cerebral ischemia. *J Neurosci*. 2007;27: 6320–6332. doi:10.1523/JNEUROSCI.0449-07.2007
65. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem Rev*. 2014; doi:10.1021/cr400585q
66. Kido T, Lau Y-FC. The human Y-encoded testis-specific protein interacts functionally with eukaryotic translation elongation factor eEF1A, a putative oncoprotein. *Int J Cancer*. 2008;123: 1573–1585. doi:10.1002/ijc.23697
67. Khacho M, Mekhail K, Pilon-Larose K, Pause A, Côté J, Lee S. eEF1A Is a Novel Component of the Mammalian Nuclear Protein Export Machinery. *Mol Biol Cell*. 2008;19: 5296–5308. doi:10.1091/mbc.E08-06-0562
68. Masson N, Willam C, Maxwell PH, Pugh CW, Ratcliffe PJ. Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. *EMBO J*. 2001;20: 5197–5206. doi:10.1093/emboj/20.18.5197

Tables

Position	Wild type	Mutant	Conserved	PolyPhen 2
15	Ala	Val	yes	possibly damaging
17	Thr	Ala	yes	benign
26	Cys	Tyr	yes	possibly damaging
32	Ala	Val	yes	benign
39	Arg	Gly	no	benign
49	Asn	Lys	yes	probably damaging
51	Arg	Cys	no	benign
52	Glu	Gly	yes	probably damaging
55	Arg	Trp	no	probably damaging
60	Asn	Lys	yes	probably damaging
62	Ser	Asn	yes	probably damaging
66	Val	Met	yes	probably damaging
69	Val	Leu	yes	benign
70	Trp	-	yes	n/a
78	Leu	Met	no	probably damaging
80	Tyr	Asp	yes	probably damaging
86	Gly	Ser	yes	probably damaging
90	Arg	His	no	possibly damaging
91	Ile	Thr	yes	benign
94	Phe	Leu	yes	possibly damaging
98	Pro	His	yes	probably damaging
99	Trp	Arg	yes	probably damaging
105	Arg	Ser	yes	benign
114	Gln	Lys	yes	possibly damaging
120	Pro	Leu	yes	probably damaging
121	Ser	Tyr	yes	probably damaging
140	-	Arg	-	n/a

Table 1. VLP mutations and pathogenicity prediction. For each VLP mutation found in COSMIC, we report the amino acid position and type for both wild type and mutated residues. Conservation in sequence and PolyPhen predictions (benign, possibly damaging or probably damaging) are also shown. “n/a” is used for non-missense mutations.

Figures

Figure 1. VLP sequence conservation. Multiple sequence alignment of VLP showing relevant conservation for the globular domain corresponding to the β -domain of pVHL. Blue box for the α -domain of pVHL missing in VLP. Color bars represent putative protein-protein interaction interfaces predicted for VLP using pVHL description as reported in [53]. VLP cancer mutations found in COSMIC are shown with red dots.

Figure 2. VLP expression in human. Box plot of human VLP gene expression data from GTEx Portal. VLP is actively expressed in testis and brain (cerebellum and basal ganglia).

Figure 3. Phylogenetic analysis of VLP and pVHL. The VLP group is highlighted in orange, with sky-blue box used for pVHL. Bootstrap values are shown at nodes.

Figure 4. Overview of VLP structural features. Structural comparison of domain organization and electrostatic surface between VLP and pVHL. VLP homology model in ribbon representation with semitransparent surface colored by conservation (purple, high; turquoise, low). VLP corresponds to β -domain of pVHL (blue box).

Figure 5. Molecular dynamics simulation of VLP/HIF-1 α interaction and VLP mutation analysis. The starting VLP/HIF-1 α complex is shown on the left (blue box). The results of 40 ns of simulation for wild type HIF-1 α (top right, red box) revealed VLP forms stable interaction with the LxxLAP motif [68], while presence of hydroxylation (bottom right, green box) reinforces the interaction yielding enlarged interaction surface. The HIF-1 α ODD motif is shown in light blue and red is used for VLP interacting residues. Light grey is for region of VLP not involved in the binding.

Figure 6. VLP in the cellular context. Conservation analysis suggests 7 putative interactors shared with pVHL beyond HIF-1 α . Interactions rely on the putative interaction interfaces B and C highlighted in the structure as purple and green, respectively. Interactors associating multiple interface are presented as wine-colored. Putative functional pathways are presented above and associate to each interface. Spheres represent VLP mutation sites and are colored according to the affected VLP binding interfaces.

Figure 7. Organs affected by VLP missense mutations in human cancer. Fractions are shown in a single pie chart. Notice the prevalence of intestinal cancers.

**Protein
degradation**

VDU1

VDU2

B

B/C

C

**Cell growth
Division
Differentiation**

EP300

eEF1A

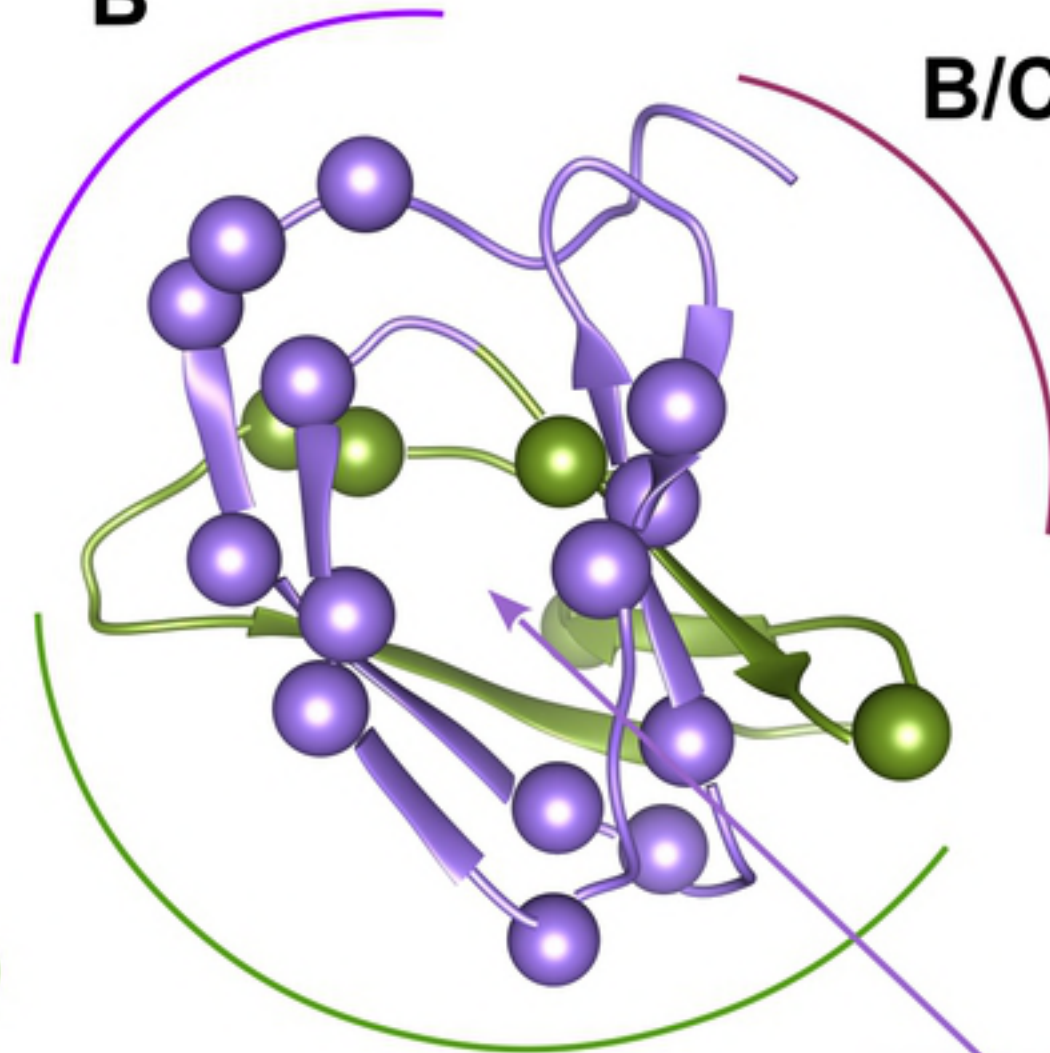
Androgen
Receptor

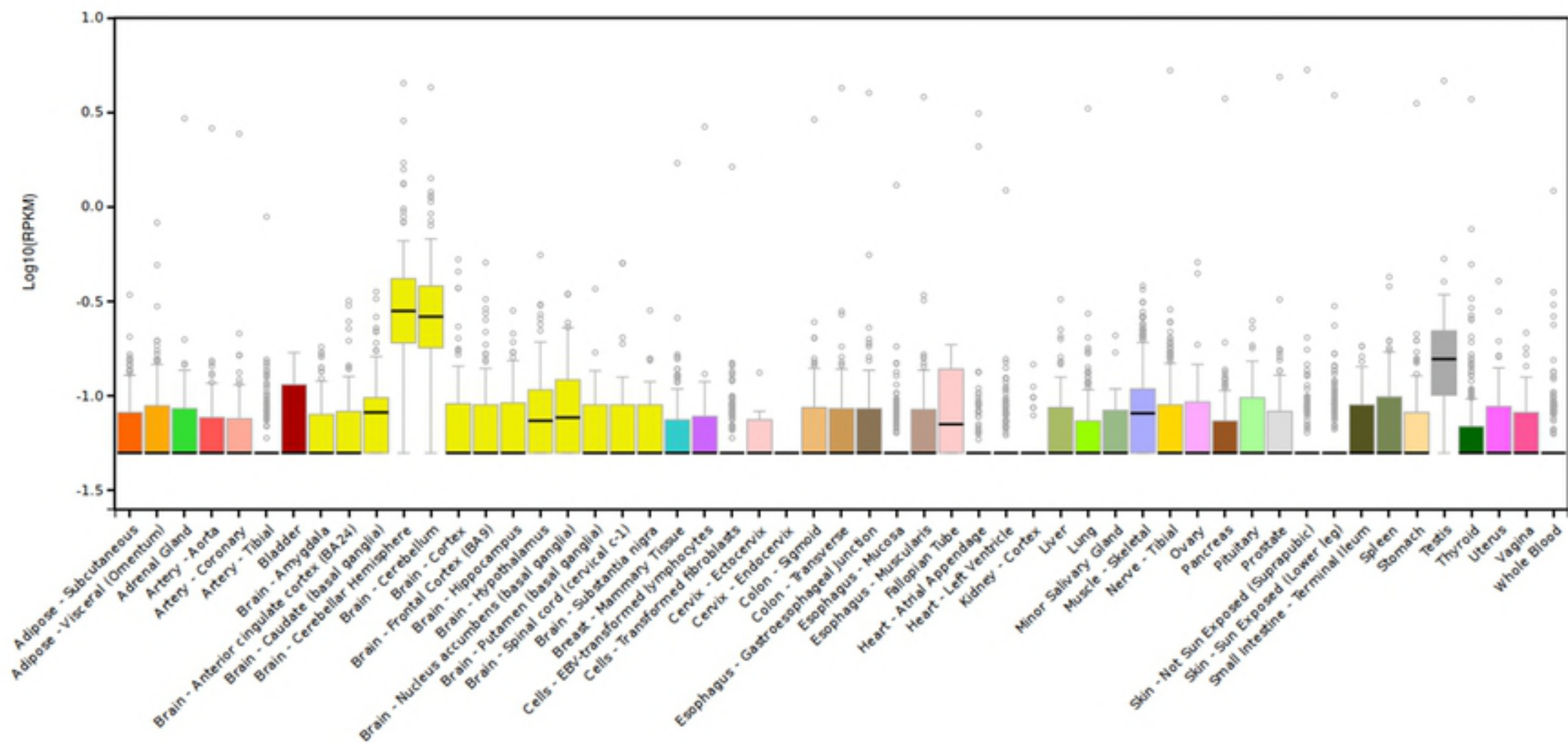
PRKCZ

TBP1

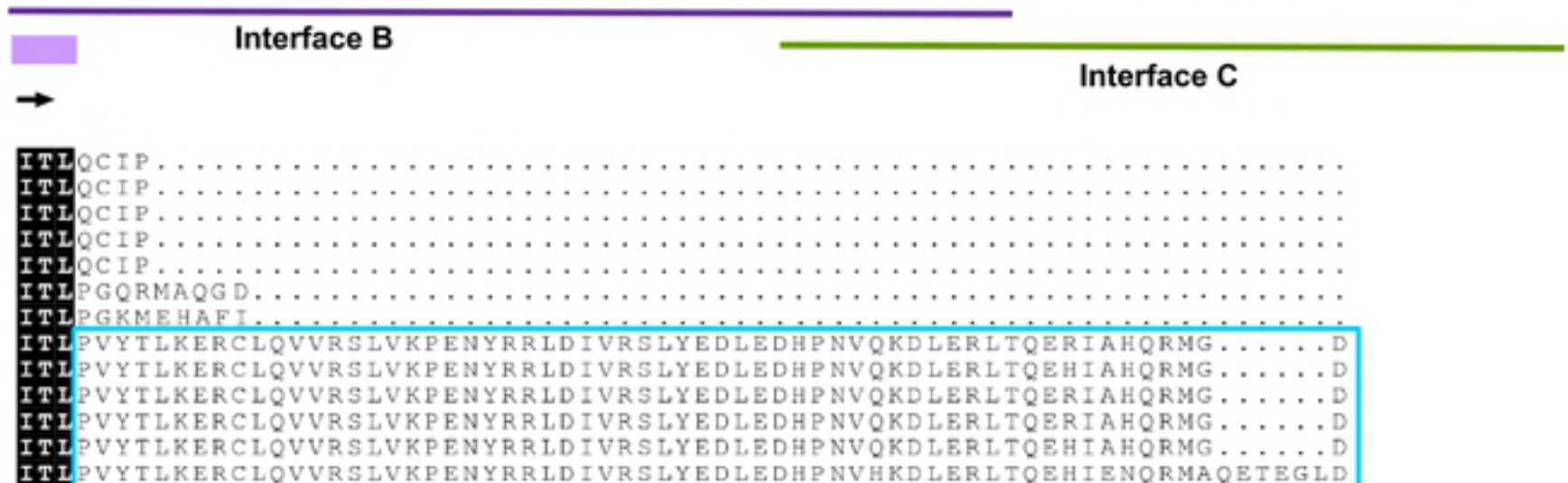
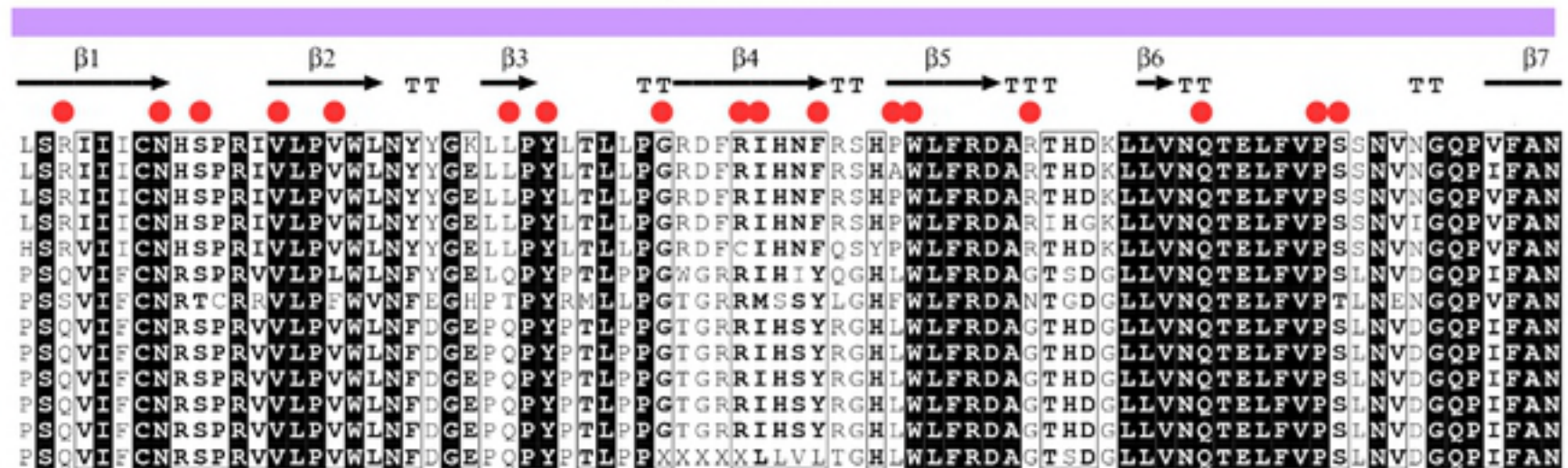
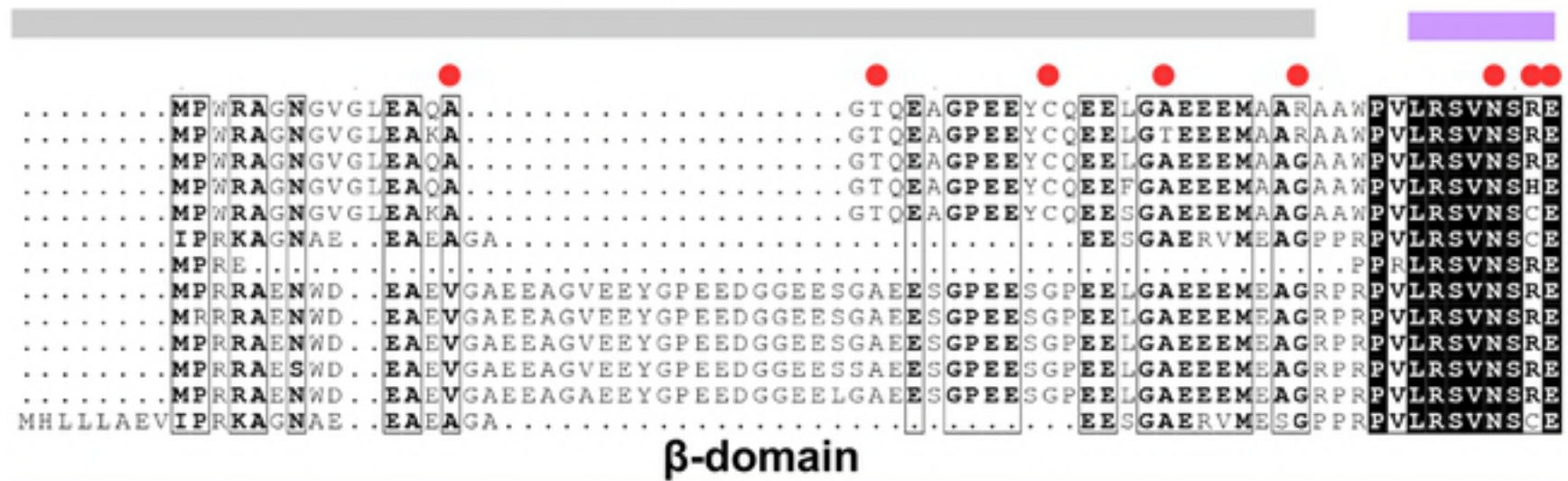
**Intra cellular signal transduction
Insulin receptor signaling
Protein degradation**

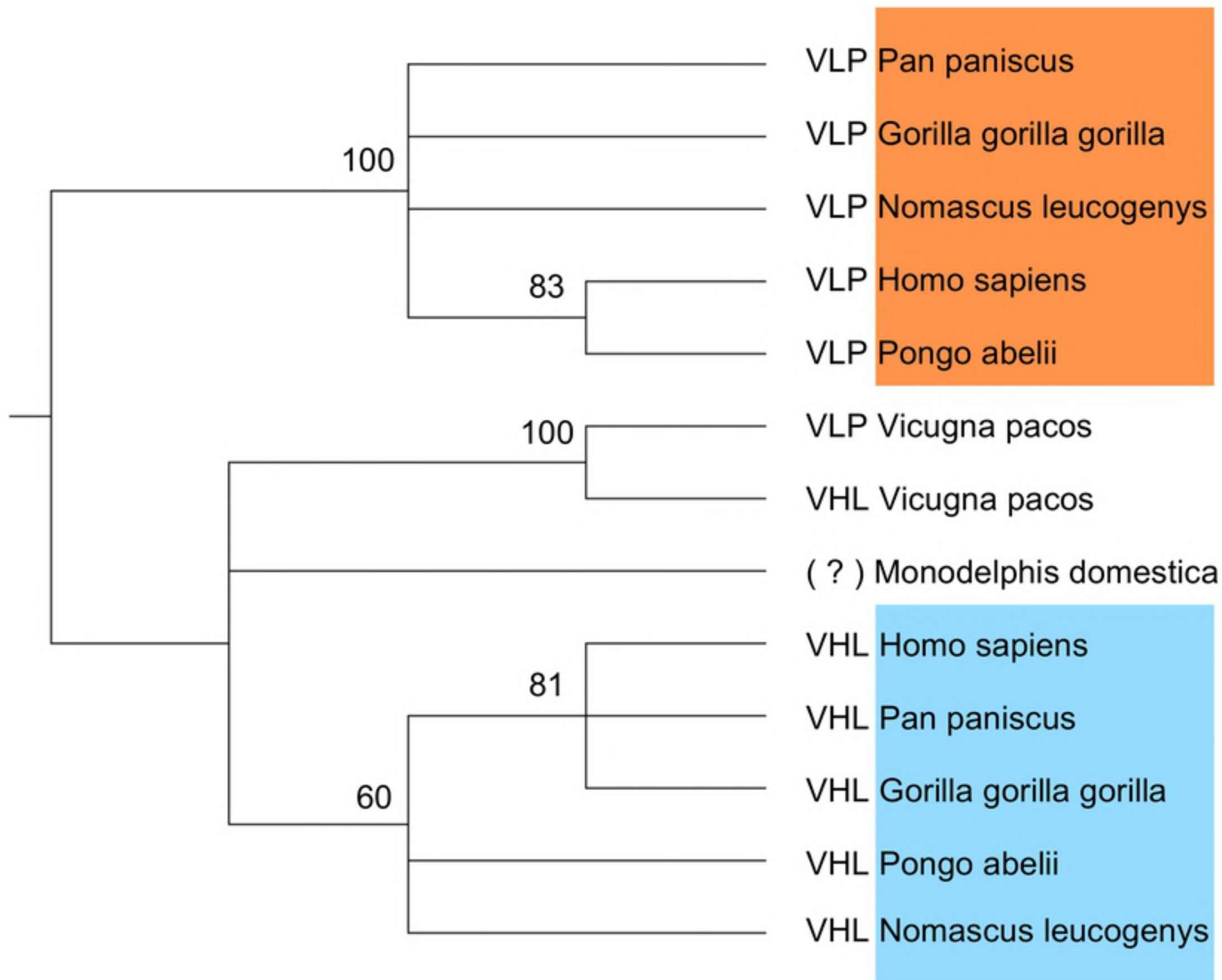
HIF-1 α
ODD motif
PDB:1LM8



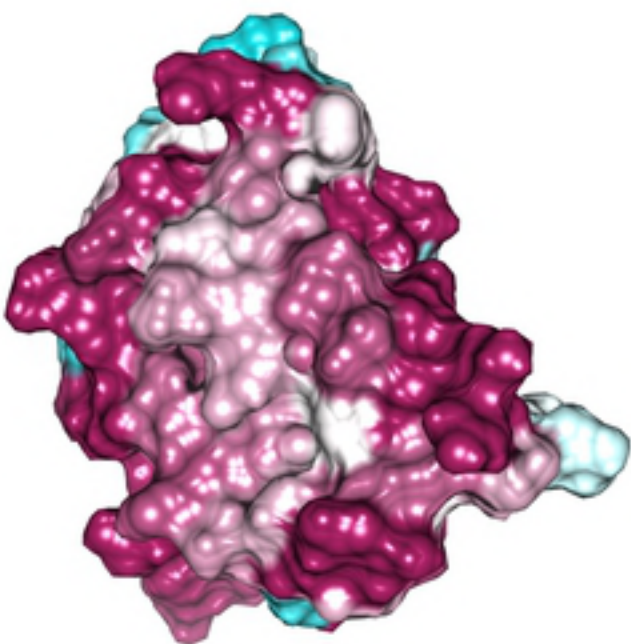


N Terminus (xEEEx)x2





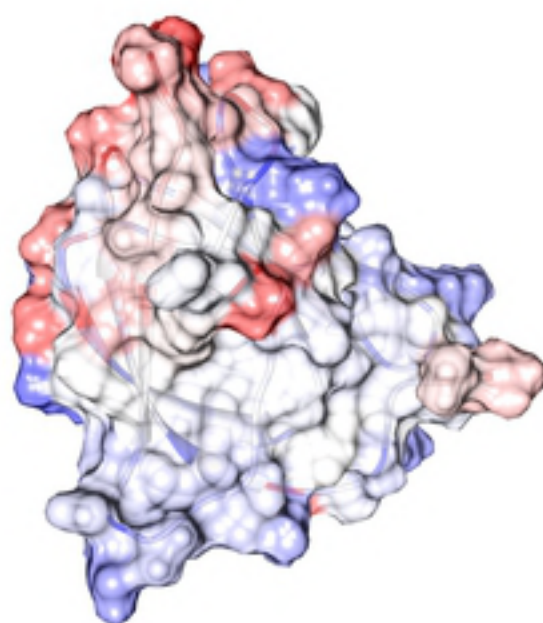
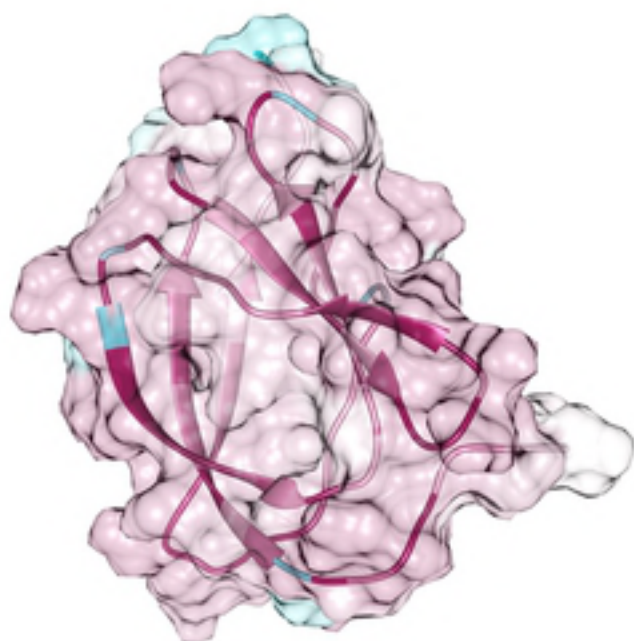
Conservation map



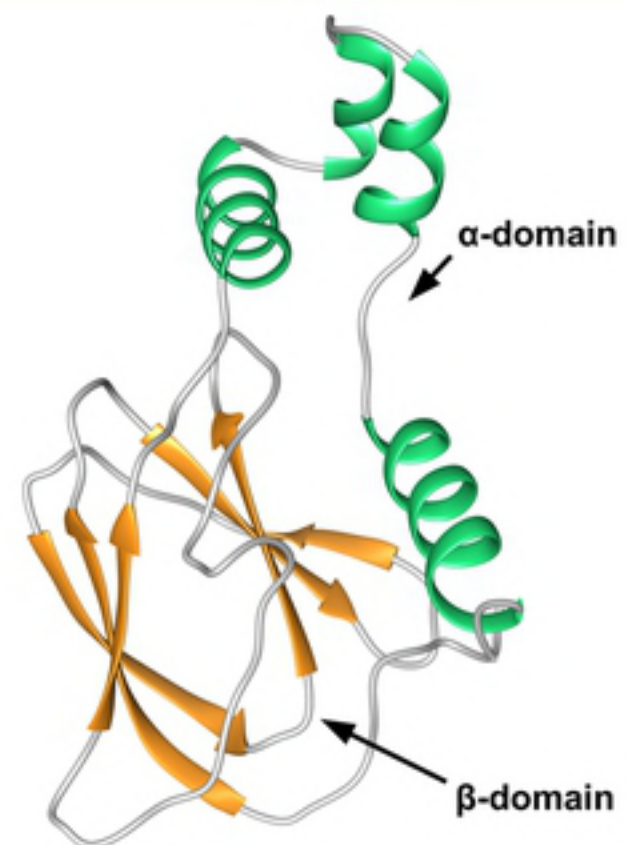
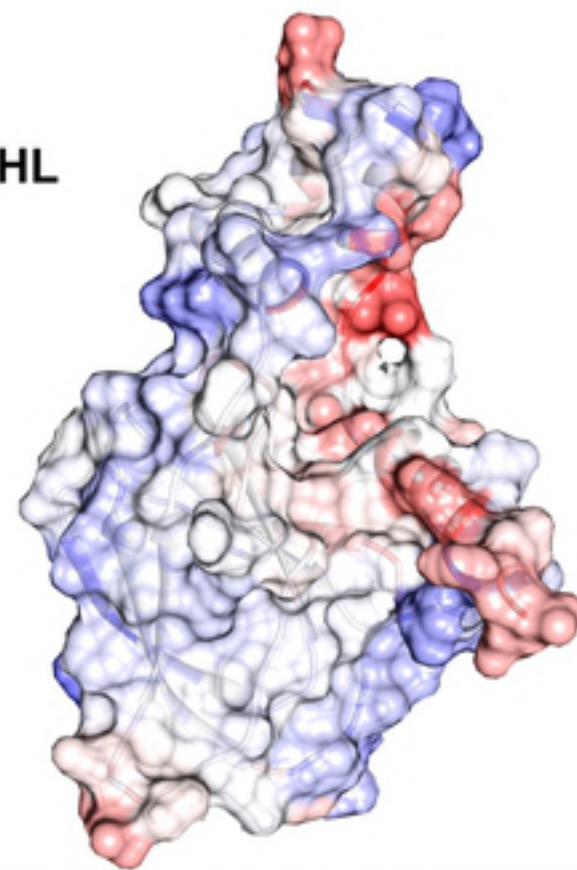
Electrostatic surface



VLP



pVHL



t = 0



t = 50 ns

