The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats

Received for publication, March 21, 2017, and in revised form, May 11, 2017 Published, Papers in Press, May 19, 2017, DOI 10.1074/jbc.M117.787291

Camila Cristina Pasquali†1, Zeyal Islam†1, Douglas Adamoski†1, Igor Monteze Ferreira‡5, Ricardo Diogo Righeto‡2, Jefferson Bettini*, Rodrigo Villares Portugal†, Wyatt Wai-yin Yue‡, Ana Gonzalez‡*, Sandra Martha Gomes Dias‡3, and Andre Luis Berteli Ambrosio‡4

From the †Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, São Paulo 13083–970, Brazil, the Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, United Kingdom, the §Laboratório Nacional de Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, São Paulo 13083–970, Brazil, the ¶School of Electrical and Computer Engineering, University of Campinas, São Paulo 13083–852, Brazil, and the **Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025

Edited by Norma Allweil

On the basis of tissue-specific enzyme activity and inhibition by catalytic products, Hans Krebs first demonstrated the existence of multiple glutaminases in mammals. Currently, two human genes are known to encode at least four glutaminase isoforms. However, the phylogeny of these medically relevant enzymes remains unclear, prompting us to investigate their origin and evolution. Using prokaryotic and eukaryotic glutaminase sequences, we built a phylogenetic tree whose topology suggested that the multigene architecture was inherited from bacterial ancestors, probably simultaneously with the hosting of the proto-mitochondrion endosymbiont. We propose an evolutionary model wherein the appearance of the most active enzyme isoform, glutaminase C (GAC), which is expressed in many cancers, was a late retrotransposition event that occurred in fishes from the Chondrichthyes class. The ankyrin (ANK) repeats in the glutaminases were acquired early in their evolution. To obtain information on ANK folding, we solved two high-resolution structures of the ANK repeat-containing C termini of both kidney-type glutaminase (KGA) and GLS2 isoforms (glutaminase B and liver-type glutaminase). We found that the glutaminase ANK repeats form unique intramolecular contacts through two highly conserved motifs; curiously, this arrangement occludes a region usually involved in ANK-mediated protein-protein interactions. We also solved the crystal structure of full-length KGA and present a small-angle X-ray scattering model for full-length GLS2. These structures explain these proteins’ compromised ability to assemble into catalytically active super-tetrameric filaments, as previously shown for GAC. Collectively, these results provide information about glutaminases that may aid in the design of isoform-specific glutaminase inhibitors.

Glutamine, the most abundant amino acid in human plasma and muscles (1), is consumed by rapidly proliferating tumor cells to meet their increased energy and biosynthetic precursor demands (2, 3). The enzyme glutaminase converts glutamine into glutamate, which is further catabolized to produce ATP, nucleotides, amino acids, lipids, and glutathione. In this regard, glutaminase is a well-established target for the inhibition of cell transformation (4). Accordingly, several oncogenes and tumor suppressors have already been described to be involved in glutaminase-dependent glutamine breakdown (4). Notably, c-Myc stimulates glutamine uptake and degradation by GLS (specifically the product of the GLS gene) to sustain cellular viability and Krebs cycle anaplerosis (5). Similarly, nuclear factor-κB (NF-κB) enhances glutamine metabolism by decreasing the expression of miR-23a, which targets the GLS mRNA (6); Rho GTPases, ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C)-Cdh1, and c-Jun also regulate glutamine metabolism by acting on GLS and consequently supporting cancer growth (7–9). However, a second gene encodes another glutaminase enzyme, named GLS2. In contrast to GLS, GLS2 is regulated by the tumor suppressor p53 and has tumor suppressor features toward gliomas, glioblastomas, lung cancer, and hepatocarcinomas (10–14).

The existence of multiple glutaminases in mammals was first reported by Hans Krebs (15), on the basis of the detection of tissue-specific kinetic parameters and their susceptibility to inhibition by their catalytic products, when he probed the conversion of glutamine into glutamate in tissue extracts. In addition to being expressed by two different genes, at least four glutaminase isoforms have been described in mammals. The kidney-type isoforms, kidney-type glutaminase (KGA)5 and

5 The abbreviations used are: KGA, kidney-type glutaminase; LGA, liver-type glutaminase; GAB, glutaminase B; ANK, ankyrin; STAS, sulfate transporter and anti-α-factor antagonist; EH, Eps15 homology; SAD, single-wavelength anomalous dispersion; r.m.s., root mean square; BPTES, (bis-2-[5-phenylac-
The recombinant construct that spans a common region through the use of alternative transcription initiation sites (19). GLS, the glutaminase C (GAC), are generated by alternative splicing of diversity that poses challenges that still must be overcome to glutaminase isoforms as well as explain the mechanistic contribute to the general knowledge on the many human and cryo-electron microscopy. Collectively, these results proposed on the basis of small-angle X-ray scattering data shown for GAC. A similar model for human GLS2 is also lytically active supra-tetrameric filaments, as previously terminus-containing crystallographic structure of human bacteria to higher eukaryotes. Finally, we present the first C tallographic structures of the C-terminal ANK repeats from position events. In addition, we solved high-resolution crys- tunicates gave rise to the currently known human isozymes evolutionary model wherein a endosymbiont. From this phylogenetic tree, we also propose an main architecture of the mammalian glutaminases GLS and and building a phylogenetic tree, we propose that the multidomain architecture of the mammalian glutaminases GLS and GLS2 is a feature inherited from bacterial ancestors, probably simultaneously with the hosting of the proto-mitochondrion endosymbiont. From this phylogenetic tree, we also propose an evolutionary model wherein a GLS-like predecessor gene in tunicates gave rise to the currently known human isoforms through exon remodeling, gene duplication, and retrotransposition events. In addition, we solved high-resolution crytallographic structures of the C-terminal ANK repeats from both human KGA and GAB/LGA. These structures display a unique type of dimerization for ANK domains, governed by two short amino acid motifs that are highly conserved from bacteria to higher eukaryotes. Finally, we present the first C terminus-containing crystallographic structure of human KGA, which explains its inability to assemble into the cata-lytically active supra-tetrameric filaments, as previously shown for GAC. A similar model for human GLS2 is also proposed on the basis of small-angle X-ray scattering data and cryo-electron microscopy. Collectively, these results contribute to the general knowledge on the many human glutaminase isoforms as well as explain the mechanistic diversity that poses challenges that still must be overcome to increase the chances of successfully inhibiting the enzyme in a clinical setting.

Results

The multidomain origin of glutaminases

An evolutionary analysis of 2,796 bacterial glutaminase protein sequences and 789 eukaryotic sequences selected from a BLAST search was performed using the glutaminase domain from human KGA (Ile221-Arg544) as the query sequence (E value ≤ 0.0001). Particularly, no glutaminase homologs were identified in archaea. The bacterial sequence sizes presented three normal distributions. The first and major distribution (~85% of the sequences) was centered at 313 ± 10 amino acids (Fig. 1A), thus suggesting that the proteins predominantly contained only the catalytic domain (24). The second distribution was smaller (10% of the sequences) and centered at 429 ± 16 amino acids. Detailed analysis suggested that this group possesses a sulfate transporter and anti-σ factor antagonist (STAS) domain located C-terminal to the glutaminase domain. Notably, three cyanobacteria sequences from this group possess a glutaminase domain flanked by an Eps15 homology-like domain (EH-like) at the N terminus and two clusters of ANK repeats at the C terminus (Fig. 1A). The Eps15 homology domain is a pair of EF-hand motifs that recognize and bind to proteins containing Asn-Pro-Phe (NPF) sequences (25). Aside from these three-domain cyanobacteria sequences, 17 proteo- and actinobacteria sequences situated between the second and third normal distributions (Fig. 1A, blue oval) also exhibit a glutaminase domain flanked N-terminal by an EH-like domain and C-ter- minal by only one cluster of ANK repeats.

Eukaryotic glutaminases, which primarily belong to the kingdoms Animalia and Fungi, are consistently longer and present a single Gaussian size distribution averaging 631 amino acids in length (mode at 615), with a skewed standard deviation of 45 amino acids toward shorter sequences and 26 amino acids toward longer sequences (Fig. 1B). The left-most end of the distribution contained the fungal glutaminases, which are 450 to 500 amino acids long and present a unique architecture, wherein the glutaminase domain is followed by a long C terminus (156 ± 7 amino acids) with an unknown fold (Fig. 1B, pink bar). The remaining eukaryotic glutaminases that are longer than 500 amino acids in length mostly belong to the Animalia kingdom and have extended N-terminal sequences of varying lengths (Fig. 1B). The N-terminal region immediately adjacent to the glutaminase domain is more than 40% similar to the EF-hand-like four-helix bundle, which is structurally similar to the EH-like domain found in bacteria. Indeed, the X-ray structures of human GLS containing this domain have an EH-like structure (20).

Putative ANK repeats are predicted for about 550 of the lon-ger eukaryotic glutaminases and are located C-terminal to the glutaminase domain with no exception. Another 76 sequences,
all of mammalian origin, have a shorter C terminus that closely resembles human GAC in sequence (Fig. 1B). Finally, a few glutaminase sequences (11 in total) belonging to the major haplophyte and heterokont lines of eukaryotes were identified. These sequences are much longer, ranging from 800 to 1250 amino acids in length, and the predicted architectures of some of them consist of two complete tandemly associated ANK-containing glutaminases (data not shown).

To detect evolutionary relationships based on the architectural organization, we next reconstructed an unrooted cladogram using the analyzed sequences (Fig. 1C). The derived tree structure has a taxon organization similar to that found in rRNA-based phylogenetic trees (26). Many unresolved nodes (polytomy) were observed, particularly for the bacterial glutaminases. We believe that such observation stems from the fact that many bacterial species possibly have more than one glutaminase, often as a result of the combination of a single protein with a multidomain isozyme. The obtained ancestral sequences were aligned with the human GLS (KGA and GAC) and GLS2 (GAB and LGA) and the sequence similarity displayed as a heatmap of pairwise distances constructed in SDT using MUSCLE alignment. The obtained result suggests that GLS is the most primitive gene, which has been further duplicated to generate GLS2. cNMP, cyclic nucleotide-binding domain; SRP/BC, START/RHO-like/CA/PITP/Bet v1/CoxG/CaIc.

Finally, based on a theoretical reconstruction of the ancestral glutaminases from each of the main eukaryotic branches (chordata, arthropods, nematodes, and fungi), the most primitive glutaminase acquired from a bacterial ancestor was most probably the kidney-type glutaminase (KGA-like), as shown in Fig. 1D. The general organization EH-like + glutaminase + ANK is preserved in all four reconstructed ancestral sequences. Comparison of the sequence identity of human genes with the reconstructed common ancestor of GLS and GLS2 from Gnathostomata suggested that GLS was restrained during evolution to maintain the same function, whereas GLS2 exhibited more changes, probably acquiring new functions and/or regulatory mechanisms.
The origin of glutaminase isoforms

Having established a plausible origin for the multidomain architecture of the glutaminases, we next asked how the different variants of the enzyme, currently observed in humans, might have evolved. We first performed a “top-down” search for glutaminase genes in phylum Chordata using the complete genomic sequence of the human GLS and GLS2 genes as templates. Second, we identified regions homologous to exon 14 of GLS (after which the GLS splicing event occurs) and exon 15 (the “GAC exon”).

In the lower chordates, we verified that the genome of the tunicate sea squirt (27) (Ciona intestinalis) has a single GLS-like gene containing 12 exons. A region similar to human exon 14 (70% identity, red bar in the tunicate branch, Fig. 2A) is inserted in exon 11. The Cephalochordate amphioxus (28) (Branchiostoma floridiae) also has a single glutaminase gene, which, interestingly, has only two exons. In this gene, the region homologous to the human exon 14 is appropriately located at the 3’ boundary of exon 1; however, no homology to the GAC exon was identified in the downstream sequence. A gene duplication event was observed further upstream in lampreys (29) (Pekinus marinus, Hyperottia), which generated distinct sequences related to the human GLS and GLS2 genes.

Next, the Chondrichthyes Callorhinichus mili (elephant shark (30)) also has two glutaminases that are phylogenetically related to the GLS and GLS2 genes without annotated splicing regulation. However, in the GLS gene, a region homologous to the human GAC exon is for the first time observed and located downstream of an exon 14-equivalent sequence, thereby suggesting an origin for the splicing variant GAC. Zebrafish (Danio rerio (31)) and other fish species possess 5 glutaminase genes, thus suggesting subsequent autopomorphic duplication events for both GLS and GLS2. Interestingly, the GAC isoform was maintained in both copies of the GLS gene, thus supporting the hypothesis that the GAC exon appeared before this duplication; a new GLS-like (GLSL) sequence was identified in species within this branch, such as zebrafish, elephant shark, and others.

Of note, the amphibian Xenopus tropicalis (Western clawed frog (32)), with only one copy of both GLS and GLS2, points to a compaction of the GLS introns and exons against the spread of the GLS across a larger genomic region. Finally, the human glutaminase gene structures retained the same architecture observed in the amphibian genes, although the sequence lengths were smaller.

The identification of several transposable elements such as Alu and L2 (Fig. 2B), within the human GLS intron 15, strongly suggests that the insertion of the GAC exon was due to an early retrotransposition event; however, no consensus matching elements were found in the rapid evolving intron 14. Nevertheless, a thorough tblastn analysis failed to identify a region homologous to the GAC exon (exon 15) within the human genome, preventing us from proposing an original location for this sequence.

Crystal structures of the human glutaminase ankyrin repeats

We next determined the crystal structures of the ANK-containing C-terminal regions of both human KGA (KGA.ANK: Val551-Leu669) and GLS2 (GLS2.ANK: Lys485-Val602). Both proteins were subjected to limited proteolysis with trypsin (for KGA.ANK) or chymotrypsin (for GLS2.ANK) during purification to remove the likely intrinsically flexible regions and obtain well diffracting crystals. X-ray diffraction datasets were obtained for KGA.ANK crystals belonging to two different space groups, namely the tetragonal P4_2_2_1 and a monoclinic P2_1. Sulfur single-wavelength anomalous dispersion (SAD) phasing was used to produce a higher resolution molecular model (at 1.41 Å maximum) from the tetragonal crystals. The obtained structure was then used as a search model for molecular replacement of the monoclinic crystals, which diffracted up to 1.74-Å maximum resolution. Finally, the structure, which was collected at 2.55-Å maximum resolution, of the GLS2.ANK hexagonal crystals (space group P6_3) was solved by molecular replacement using the KGA.ANK structure. The statistics of the data collection, processing, phasing, and model refinement data for all the structures are shown in Table 1. Because of the high structural similarity observed for both KGA.ANK structures (backbone r.m.s. deviations of 0.4 Å), only the higher resolution model was used for the subsequent detailed analysis and comparison to the GLS2.ANK structure.

Overall, the C-terminal domains of KGA and GLS2 each contain three ANK repeats (labeled ANK1 to ANK3) and are very similar, both in sequence composition (80% identical) and overall structure (r.m.s. deviation of 0.5 Å over 89 equivalent Ca positions; Fig. 3A). The largest structural variation (a shift between 1.3 and 1.8 Å of the Cα atom positions) occurs in the β-hairpin that links ANK1 to ANK2 (β-hairpin 2), whereas the minimal deviations (~0.1 Å) are observed in the inner and outer α-helices of ANK2. The primary sequences of ANK1 and ANK3 repeats in both KGA and GLS2 diverge from the canonical ANK repeats at the TPLH tetrapeptide motif, which is located at the beginning of the inner α-helices and is responsible for the stabilization of ANK repeats through reciprocal hydrogen bonding between the threonine hydroxyl and histidine imidazole groups (33, 34) (Fig. 3B). The ANK1 inner helix of both KGA and GLS2 is a half-turn longer and lacks the β1 turn compared with the typical inner helix observed in ANK repeats (Fig. 3B), whereas in the ANK3 inner helix, the histidine is replaced with an aspartate, thus preventing the formation of the conserved hydroxyl-imidazole hydrogen bond. Moreover, in both structures, ANK2 contains an extra residue, a surface-exposed lysine (Lys511 at KGA and Lys543 at GLS2, indicated by a plus sign in Fig. 3B), at the C terminus of the outer helix. Finally, the GLS2 ANK3 outer helix is one turn longer than the canonical ANK repeats (Fig. 3B).

Glutaminase ankyrin repeats assemble into atypical dimers

A search for the biologically relevant assembly of the KGA and GLS2 ANK domains was performed using the PISA (protein interfaces, surfaces and assemblies) server (35). The program identified a thermodynamically stable interface involving a crystallographic homodimer, with a solvation free energy gain of ~5.4 kcal/mol upon formation (Fig. 3C). The identified interface occludes an area of 800 Å² and corresponds to a symmetric groove-to-groove interaction in both structures. Twelve hydrogen bonds and 12 salt bridges are formed between the short loops present in the β-hairpins 1 (linking ANK1 to ANK2) and
Figure 2. Evolution of glutaminase intron-exon structure. A, protein sequence corresponding to human GLS exon 14 was used as a query of translated blast (tblastn) searches against available genomes from *C. intestinalis* (Tunicate), *B. floridae* (Cephalochordate), *P. marinus* (Cyclostomata), *C. milii* (Chondrichthyes), *D. rerio* (Actinopterygii), and *X. tropicalis* (Amphibia). After an initial Reciprocal Best Hit region was found, the correct position of the exon was determined by pairwise LALIGN between sequences. Region downstream genomic sequence homologous to exon 14 were also evaluated by LALIGN against human GAC-exclusive exon (exon 15). As both *P. marinus* glutaminases are incomplete, available exons were found in contig using LALIGN approach. The human genes are represented at the top branch (Amniota). The thin red bars represent the homologous region to the GLS exon 14 and the light blue boxes indicate homologous regions to the exon 15 of human GLS. The introns, represented as dashed lines, are not drawn to scale and the omitted sections are indicated by two bars ( ). The asterisk (*) indicates that the deposited sequences are incomplete. B, a search for retrotransposition evidences in the genomic region comprehending exon 15 and introns 14 and 15 of human GLS using TranspoGene resulted in the identification of various transposable elements in the intron 15; however, no consensus matching elements were found in the intron 14.
2 (linking ANK2 to ANK3, Fig. 3B) in chain A with their respective equivalents in chain B (Fig. 3C, left panel, for β-hairpin 1 loop interactions and right panel for β-hairpin 2 loop interactions). The short loops of β-hairpins 1 and 2 contain the motifs Asp-Tyr-Asp (DYD, residues 583–585 in KGA and 516–518 in GLS2) and Asp-Arg-Trp (DRW, residues 617–619 in KGA and 550–552 in GLS2), respectively (Fig. 3C). Despite the overall relatively low sequence identity of the eukaryotic and bacterial glutaminase ANK repeats, both the DYD and DRW motifs are highly conserved (Fig. 3, D and E, respectively). Finally, additional polar interactions are observed between the β-hairpin 1 loop and the ANK3 inner helix and between the β-hairpin 2 loop and the ANK2 inner helix, complementing the identified symmetrical interface (data not shown).

Crystal structure of full-length human KGA

In addition to the determination of the crystal structures of the isolated C-terminal ANK dimers, we also solved the novel crystal structure of KGA containing the C terminus, which was bound to the inhibitor bis-2-[5-phenylacetamido-1,2,4-thiadiazol-2-yl]ethyl sulfide (BPTES) (36). The diffraction dataset was collected at a 3.6-Å maximum resolution (space group P42212) with lengthy cell parameters and a high solvent content of ~78% (Table 1). Phasing was achieved by molecular replacement using the coordinates of the previously solved N-terminal and glutaminase domains from GAC bound to BPTES (PDB code 4jkt) and the KGA.ANK dimer described here as search models. The P42212 asymmetric unit contains five KGA monomers; accordingly, one canonical tetramer (a dimer of dimers) is formed within the asymmetric unit by four monomers and, the fifth monomer is related to another tetramer by the crystal 2-fold screw axis. The final model obtained (Fig. 4A) was refined to R\(_{\text{factor}}\) and R\(_{\text{free}}\) of 27.3 and 31.8%, respectively.

The ANK dimers in the KGA structure (green surface in Fig. 4B) are spatially located between the N-terminal EF-hand-like domains (blue surfaces in Fig. 4B). Notably, both ANK dimers of a tetramer lie on the same side of a sectional plane defined by the two longest axes of the protein (Fig. 4B), breaking the previously established 2-fold dihedral symmetry of the glutaminase structures lacking the C terminus (20).

The low-resolution model did not allow us to provide a detailed description of the side chain interactions between the ANK dimers and the N-terminal domain. However, the close contact between the ANK1 outer helix of chain A with the EF-hand-like domain helix H1 of chain B is unmistakable (Fig. 4C). The interface contains large, exposed polar residues, such as glutamic acid and arginine, which are likely to be responsible for the contact. An equivalent arrangement was also observed for the ANK1 outer helix of chain C with the EF-hand-like domain helix H1 of chain D (data not shown). The majority of the C-terminal region of KGA adjacent to the ANK repeats (residues Thr\(^{447}\)–Leu\(^{569}\)) is likely heterogeneous in conforma-
tion and therefore could not be modeled. According to secondary structure predictions by the JPreD4 server (37), this region is mostly unstructured. This information, added to the fact that such region is present in a KEN box involved on KGA degradation (8), suggest that this segment is likely functionally independent from the ankyrin domain.

As shown in our previous study (23), the phosphate-dependent enzymatic activity of the GLS isoforms is directly related to their ability of self-assembling into supra-tetrameric helical filaments; GAC is the most active isoform and forms longer filaments. Moreover, we previously demonstrated that this assembly was generated by end-to-end associations via the N-terminal domains; the KGA assembly was always shorter, although it formed the same structure. Based on the KGA crystal structure presented here, we hypothesized that the ANK repeats located between the N-terminal domains prevent KGA from forming longer active filaments due to the fact that it destabilizes the polymerization. More specifically, the ANK dimers limit the association of the KGA tetramers into long polymers (23) because they limit the formation and availability of the filamentation interface (Fig. 4D). This interface is the groove region by which the single strand filaments of the GAC

Figure 3. Overall structural features of ANK domain of glutaminases. A, schematic representation of three ANK repeats (ANK1 to ANK3) present as a C-terminal domain of KGA and GLS2. Both ANK repeats are very similar in sequence as well as in the overall structure. B, alignment of the glutaminase ANK repeats and the consensus sequence (shown on top), proposed in Ref. 39. Highly conserved residues are capitalized and in red, semiconserved residues are colored cyan and not capitalized. Residues involved in the dimer interface are indicated by a bar. ANK2 in both KGA and GLS2 contains an extra surface-exposed lysine, which is indicated by the + signal. On the right, the superposition of ANK1, ANK2, and ANK3 is represented. C, dimer interface and associated interaction of ANK repeats. The side chains of the motifs DYD (left) and DRW (right) are represented in sticks.

D, representation of sequence conservation of glutaminases ANK from eukaryotes and E, bacteria. The size of the letters is proportional to the degree of conservation of residues. The motifs DYD and DRW are highlighted. The residues in red shows the 100% conservation throughout the alignment. Residues are numbered according to human KGA.

The multidomain architecture of human glutaminases
isoform grows indefinitely, via end to end interaction between pairs of N-terminal domains (23).

To further confirm this hypothesis, we generated a C terminally truncated GLS construct containing the common sequence between KGA and GAC (N terminus and glutaminase domains, called \( \gamma/H9004 \), but lacking the ANK repeats. Accordingly, the construct had a larger Stokes radius in comparison to KGA and GAC and was more heterogeneous in size (Fig. 4E). Moreover, as expected, the \( \gamma/H9004 \) construct was as active as GAC, possibly due to its increased ability to form extended filaments (Fig. 4F).

**Low resolution structure of GLS2**

Last, the availability of the crystal structure of individual domains of GAB/LGA, such as the C-terminal ANK dimers reported here and the glutaminase domain tetramer (PDB code 4bqm), allowed for a proposition of the full-length structure of GLS2. A, electron density and final model for the complete tetramer of KGA. B, planar representation, dividing the tetramer into front and back halves to show the position of two ANK dimers at the same side of the plane. C, upper view of the tetramer, indicating the close contact between the ANK1 outer helix and the helix H1 of the N-terminal EF-hand-like domain. D, the filament interface between two N-terminal domains (region delimited by gray), and by which the single strand polymer of GAC (PDB code 4jkt) grows (23), is occluded in KGA because of the presence of the ankyrin dimers. E, the Stokes radii calculated from the size exclusion chromatographic profiles of KGA, GAC, and \( \Delta C \) (KGA lacking the ANK repeats), indicating that ANK repeats prevent the self-assembly of glutaminases into the higher supra-tetrameric filament forms. F, the comparative enzymatic efficiency of KGA, GAC, and \( \Delta C \) toward glutamine in the presence of 20 mM phosphate. The lack of ANK repeats enhances the catalytic efficiency of KGA. Red lines indicate mean ± S.D.

**Discussion**

The human genome has two copies of glutaminase-coding genes, which are known to produce at least four isoforms through alternative splicing or transcription at alternative initiation sites (4). Genetic evolutionary models often predict an increase in the gene copy number prior to the specialization or gain of new functions, with a momentary intermediate state of specialization with an overlap of functions (38). In addition, this apparent redundancy in glutaminase isoforms can be explained by differential tissue expression patterns, as well as the cell proliferation state. In this regard, KGA is expressed in kidney tubule epithelial cells, in which ammonia production is key to renal acid-base regulation (16); LGA is expressed in periportal hepatocytes, where it participates in urea synthesis (17). In addition, although GAC meets the glutamine-dependent metabolic demands of several types of cancer (20), GAB has tumor-suppressing activity in hepatocarcinomas and gliomas (40).

Despite this functional diversity, the human glutaminase isoforms converge into a conserved multidomain structure. Based on a combination of computational predictions and experimental evidence mainly produced by X-ray crystallography studies, we found that all four isoforms contain an N-terminal EF-hand-like domain, followed by a glutaminase domain in the middle and three ANK repeats at the C terminus. The only exception is GAC, in which the C terminus is a short
unstructured region, because of alternative splicing of the GLS gene.

In this work, the multidomain architecture was shown to be shared by all chordates, including fishes, reptiles, birds, and amphibians, as well as by arthropods and nematodes. By surveying the database-deposited glutaminase protein sequences, we observed similar organizational characteristics in free-living proteo- and cyanobacteria, but not fungi. From our data, we propose that the multidomain glutaminase structure has a very ancient origin. In addition, we verified that multiple GLS genes and isoforms are present in a wide range of vertebrates. By further analyzing the genomes of representative species of the phylum Chordata, we were able to identify early events of exon repositioning among tunicates and cephalochordates, followed by gene duplication in Hyperoartia and exon retrotransposition in Chondrichthyes, concomitant with the change from a simple kidney structure pronephros to a more complex mesonephros with tubules (41). Together, all these features are likely the basis of the isozyme diversity observed in vertebrates. We also identified a glutaminase ancestor that was more similar to the human kidney-type glutaminase and gave rise to the liver-type glutaminase, whereas the appearance of GAC-like glutaminases, with shorter C termini, was a late event in glutaminase evolution. GAC is the most active isoform compared with KGA and LGA, a feature linked to its capacity to assemble into long filament-like superstructures (23). In the present study, we found that this feature is possible only because GAC lacks the original C-terminal bulky ANK repeats of parental KGA, because the AC-terminal construct assembles in longer filaments and is as active as GAC. Therefore, we propose that filament formation is a gain-of-function characteristic of vertebrate glutaminases, which has been positively selected to create a more active enzyme due to the substrate channeling effect.

Our group and others have previously published the structure of the N-terminal EF-hand-like and glutaminase domains of GLS (20, 42), but the structure of the C-terminal ANK repeats remained unsolved. Here, we provide novel crystal structures for the ANK domains of KGA and GLS2, which surprisingly form an atypical dimer. ANK repeats usually mediate the interaction of a protein with a different partner. However, KGA.ANK and GLS2.ANK mediated the formation of homo-oligomers, a rare feature. A survey of structures in the Protein Data Bank revealed 68 unique crystal structures of ANK-containing proteins, containing between 2 to 24 ANK repeats (supplemental Table S1). Of these crystal structures, only 5 displayed an ANK-to-ANK association (supplemental Fig. S1). However, none contained an interface that resembles the glutaminase ANK dimer (contact between the DYD and DRW motifs). Therefore, we propose that the ANK dimer association described here is structurally unique to glutaminases and was selected as a conserved feature throughout glutaminase evolution. The explanation for this positive selection remains elusive.

The short regions located immediately after the ANK repeats in KGA and GLS2 (KEN and ESMV motif, respectively) are involved in E3 ubiquitin ligase-dependent degradation (8) and direct interactions with PDZ domain-containing proteins (43), respectively. An analysis of protein complexes including ANK-containing proteins and their interaction partners (44–47)
showed that most of the contacts necessary for ANK-to-ANK interaction involves the ANK grove (the concave surface). Although most of the known cases involve the concave surface, some ANK repeats mediate protein-protein contacts through the convex face. One example is the human oncoprotein g-ankyrin, which was crystallized in complex with an antibody fragment. In this structure, the interaction occurred through the outer α-helices of the ANK4-ANK6 repeats (48). The vac-
cinia virus K1 protein consists entirely of ANK repeats that are involved in interactions mediated by the convex surface (49). In addition, the VP59-domain ANK repeat protein binds to Rab32 through the convex side of its ANK repeat (50).

In this regard, although the concave faces of the ANK repeats of KGA and GLS2 make contacts within the glutaminase, we predict that the ANK repeats may still mediate protein-protein interactions through the convex surface. Recently, GAB has been shown to bind and inhibit the small GTPase Rac1 by preventing its interaction with a guanine-nucleotide exchange factor, an interaction that involves the C terminus of GLS2 (residues 464 to 602) (40). Because the region contains the ANK repeats of GAB, we predict that the contact involves the ANK concave face. Considering the long list of partners that were shown to interact with glutaminases using mass spectrometry-based approaches (51), further investigations are required to confirm the mechanism by which the ANK repeats in glutaminases mediate protein-protein interactions.

Last, although at low resolution, the novel ANK-containing structure of human KGA confirms our previous hypothesis that the long-range propagation of filaments for this isoform was thwarted by the presence of the ankyrin repeats themselves (23). As observed in the crystallographic model, the ANK dimer occludes the formation of the filamentation groove, by which the single strand filament of the GAC isoform grows indefi-
nitely, via end to end interaction between pairs of N-terminal domains. Because GAC possesses a shorter unstructured C-
terminal sequence, this hypothesis is confirmed when the ANK domain is completely deleted in a mutant construct, which can form longer polymeric species, and therefore, resulting in a more active protein.

**Experimental procedures**

**Phylogenetic reconstruction**

726 Glutaminase protein sequences obtained from Gen-
Bank™ (52), by performing a BLAST search with human glu-
taminase domain sequence from GLS as the query, were aligned
using kalign (53). Maximum likelihood phylogenetic recon-
struction was performed with RaxML (54) with 260 bootstrap
pseudoreplicates and using automated model search. Rep-
lates were summarized with sumtree from DendroPy (55).

The generated phylogenetic tree and respective alignment were used to reconstruct ancestral states for specific nodes using maximum parsimony, as implemented in phangorn (56). Heat-
maps from pairwise distances were constructed in SDT (57)
using MUSCLE (58) alignment.

**Evolution of glutaminase exon structure**

Protein sequence corresponding to human GLS exon 14 was
used as input of translated blast (tblastn) searches against avail-
able genomes from *C. intestinalis* (27), *B. floridiae* (28), *P. mar-
nus* (29), *C. millii* (30), *D. rerio* (31), and *X. tropicalis* (32). After an initial Reciprocal Best Hit (59) region was found, the correct position of the exon was determined by pairwise LALIGN between sequences (60). Regions downstream from the genomic sequence homologous to exon 14 were also evaluated by LALIGN against human GAC-exclusive exon (exon 15). As both *P. marinus* glutaminases are incomplete, available exons were found in contig using the LALIGN approach. The GLS region comprising intron 14, exon 15, and intron 15 was used as input of TransposGene (61) retrotransposon searching.

**Protein expression and purification**

The constructs of human KGA Val124-Leu669 and Val551-
Leu669 were amplified from the pcDNA3.1/hKGA-V5 clone, kindly provided by Dr. Richard Cerione (Cornell University,
Ithaca, NY), and subcloned into the pET28a plasmid (Novagen)
using the Ndel and Xhol restriction sites with N-terminal His6

tag. KGA Val124-Arg550 was generated from the pET28a KGA
Val124-Leu669 construct by site-directed mutagenesis of the
Val551 residue into a stop codon (TAA) using the QuikChange
II XL Site-directed Mutagenesis Kit (Agilent Technologies).
The construct of human GAB Lys485-Val602 was cloned into
pNIC28-Bsa4 plasmid. The construct of GAC Met128-Ser603
was amplified from a mouse fetal brain tissue cDNA library and
cloned into the pET28a plasmid. The four constructs were
transformed into *Escherichia coli* Rosetta-2 thermocompetent
cells (Merck). Overnight cultures, grown in LB medium in the
presence of 50 μg ml⁻¹ of kanamycin and 50 μg ml⁻² of clor-
amphenicol, were inoculated in a ratio of 1:100 in 3-liter cul-
tures supplemented with the same antibiotics and left shaking
at 200 rpm for 5 h at 37 °C. The cultures were then down-
tempered to 18 °C for 1 h before induction with 200 nM isopro-
pyl β-D-thiogalactopyranoside for 16 h at 18 °C. Cells were
collected by rapid centrifugation and resuspended in 500 mM
NaCl, 50 mM Tris HCl, pH 8.5 (or 50 mM HEPES, pH 7.5, for
GAB construct), 10% glycerol, and 2 mM PMSF (phenylmethylsulfonyl fluoride). Cell lysis was performed chemically, by incubation with hen egg white
lysozyme, DNase I, and deoxycholate (all three reagents from
Sigma) for about 1 h, incubated on ice. The soluble fractions
were separated from the debris by high speed centrifugation
and subsequently loaded, by gravity and in a cold room, on
immobilized metal affinity column, Co²⁺-charged TALON
(Clontech) for the GLS constructs and nickel-nitrilotriacetic
acid Superflow (Qiagen) for the GAB construct, previously equilibrated with the running buffer 10 mM NaCl and 50 mM
Tris-HCl, pH 8.5, HEPES, pH 7.5. The constructs were eluted
stepwise using running buffer to which up to 500 mM imidazole
(v/v) had been added. For the GLS constructs, the tag was
removed by overnight digestion with bovine thrombin (Sigma)
and the samples loaded into a HiTrap Q HP anion exchange
chromatography column (GE Healthcare). Elution was done by
performing a linear gradient with a buffer containing 1 M NaCl,
50 mM Tris-HCl, pH 8.5, and 2 mM β-mercaptoethanol.
The fractions containing the GLS constructs were loaded in a
HiLoad 16/600 Superdex 200 (for KGA Val124-Leu669, KGA
Val124-Arg550, and GAC Met128-Ser603 constructs) or 75 pg (GE
Healthcare) (for KGA Val551-Leu669). The final buffer conditions were 150 mM NaCl, 30 mM HEPES, pH 8, and 0.5 mM tris(2-carboxyethyl)phosphine. For the GAB construct, the eluate from the affinity chromatography was directly loaded in a HiLoad 16/600 75 pg. Protein concentration was determined by UV 280 nm using calculated extinction coefficients. The hydrodynamic parameters (Stokes radii, Rh) of human KGA, GAC, and the deletion mutant ΔC were determined by gel filtration chromatography using a prepacked Superdex 200 HR 10/30 column (GE Healthcare) in 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine). For each purified protein, ~1 mg ml⁻¹ concentration in a 500-µl volume was injected into the column. To induce the formation of the higher-order oligomeric species, 20 mM K₂HPO₄ as a final concentration was added in the protein solution before loading it onto the column. The gel filtration buffer was also supplemented with 0.5 ml min⁻¹ flow rate was maintained at 0.5 ml min⁻¹. To establish the hydrodynamic radius (Rh) and elution volume relationship of the protein, proteins of known Rh were run on the same column. The proteins used as standards were ferritin (440 kDa, Rh = 60.8 Å), aldolase (158 kDa, Rh = 48.1 Å), conalbumin (75 kDa, Rh = 36.4 Å), ovalbumin (44 kDa, Rh = 30.5 Å), and ribonuclease A (13.7 kDa, Rh = 15.9 Å).

Crystallization

Following size-exclusion purification, KGA Val124-Leu669 was concentrated using an Amicon 30-kDa cutoff concentrator (Millipore) to a final concentration of ~7.5 mg ml⁻¹. Crystallization experiments were performed at 277 K using the conventional sitting drop vapor diffusion technique. Drops were made by mixing two parts of protein previously incubated with 1.2 mM BPTES to one part of the well solution, containing 1.8 M sodium formate, 0.5 M NaCl, and 0.1 M BIS-TRIS propane, pH 6.8. Before data collection at cryogenic temperature (100 K), harvested crystals were cryoprotected with 10% ethylene glycol added to the mother liquor.

For large-scale limited proteolysis, KGA Val551-Leu669 after a size-exclusion purification step was incubated with 1:100 trypsin (Sigma) at 23 °C for 20 min. The proteolysis was stopped by the addition of 1.5 mM PMSF. For limited proteolysis of GAB Lys485-Val602, purified protein was incubated overnight with 1:1000 α-chymotrypsin at 23 °C. The digested fragments were immediately purified by size-exclusion chromatography. For crystallization trials, digested KGA Val551-Leu669 and GAB Lys485-Val602 were concentrated, respectively, to 25 and 50 mg ml⁻¹ using Amicon 10-kDa cutoff concentrators.

Both constructs were crystallized by sitting drop vapor diffusion, by mixing equal parts of protein solution and mother liquor. KGA Val551-Leu669 crystals were grown in (a) 3.2 M NaCl, 0.1 M Tris, pH 8.5 and (b) 3.8 M NH₄Cl, 0.1 M BIS-TRIS propane, pH 7. GAB Lys485-Val602 crystals were grown in 1.1 M trisodium citrate, 0.1 M imidazole, pH 8, 20 mM L-glutamine.

X-ray crystallography

Diffraction data were collected at beamlines 103 at the Diamond Light Source (UK) and 12-2 at the Stanford Synchrotron Radiation Lightsource, respectively, for KGA Val124-Leu669 and for KGA/GAB ANK crystals. Datasets were integrated using Mosflm (62) (for native datasets) and XDS (63) (for sulfur-SAD dataset) and scaled with Aimless (64). The first set of phases of KGA Val124-Leu669 was obtained by the molecular replacement technique as implemented in the program Phaser (65), using the coordinates of the mouse GAC isoform (PDB code 3SS3). KGA ANK was solved by sulfur-SAD using SHELX (66) and the model was refined using higher resolution native datasets. The model obtained for KGA was employed as a search model for solving the GAB ANK structure by molecular replacement (Table 1). Positional and B-factor refinement cycles, as well as solvent modeling, were performed with Refmac (67) and Phenix (68), followed by visual inspection using COOT (69).

Glutaminase activity assay

To obtain the kinetic parameters for KGA Val124-Leu669, KGA Val124-Arg550, and GAC Met128-Ser603, a mixture containing 10 nm glutaminase, 50 mM Tris acetate, pH 8.6, 3 units of bovine L-glutamate dehydrogenase (Sigma), 2 mM NAD (Sigma) was pipetted into 96-well plates previously filled with 6 or 12 serial dilutions of L-glutamine, to achieve a range of concentrations from 60 to 0.15 mM. K₂HPO₄ (2 M stock, pH 9.4) was added to the mixture at a final concentration of 20 mM. The formation of NADH was tracked by absorbance readings at 340 nm, for up to five consecutive minutes, at room temperature. Measurements were done in triplicate. The initial velocities, in picomoles of NADH produced per second, were calculated using an extinction coefficient for NADH of 6,220 M⁻¹ cm⁻¹ at 340 nm and 0.5 cm of path length. The total volume per reaction was 200 µl. Plate-reader used was an EnSpire (PerkinElmer Life Sciences). Measurements were done in triplicates and analyzed using GraphPad Prism 5.00 (GraphPad Software, San Diego, CA).

Small-angle X-ray scattering

Scattering data were collected at λ = 1.488 Å, for sample-detector distances of 1.1 m covering the momentum transfer ranges 0.015 < s < 0.442 Å⁻¹ (s = 4π sinθ/λ, where 2θ is the scattering angle). The data were normalized to the intensity of the incident beam and corrected for the detector response using an in-house program. Two frames of 250 s were collected and compared for radiation damage using the program PRIMUS (70). The same program was used to average the frames and subtract the buffer. The different protein concentrations were evaluated for aggregation by following increases in the measured Rg (radius of gyration) as calculated by auto Rg. The Rg was confirmed by using the indirect Fourier transform program GNOM (71), which was also used to calculate the distribution function P(r) and Dmax. The data were analyzed and processed, including ab initio construction and model averaging, using the programs contained in the ATSAS package (72).

Cryo-electron microscopy

For visualization of cryogrids, purified GAB samples were frozen onto a Gatan 626 sample holder, prepared with FEI Vitrobot Mark II (force of ~5 for 2s). Images were acquired using a JEM 2100 (200 kV) electron microscope with a LaB6 filament Gatan 4k × 4k slow scan CCD camera (US4000). Exposure time
was 1 s/frame, with a dose of 20 e Å⁻² s⁻¹. The micrographs were processed using IMAGIC (73) and EMAN 2.1 (74). Iterative stable alignment and clustering (ISAC) (75) was used to generate reference-free class-averages from both the IMAGIC stack (25,000 particles) and the EMAN2.1 stack (33,608 particles). Using IMAGIC and UCSF Chimera (76), 10,000 projection images in random orientations of the GAC crystallographic model (PDB code 3SS3) were generated, in both dimer and tetramer configurations. The bounding rectangle dimensions of 200 particles randomly extracted from the IMAGIC, EMAN2.1, ISAC/IMAGIC, and ISAC/EMAN2.1 datasets (50 from each) were classified according to longer (L) and shorter (S) dimensions. Images were low-pass filtered to reduce noise. Only the particles contained in the 50 “best” class averages generated by each program were considered (total: 200 particles). Measurements were performed in Digital Micrograph based on the integration profile across each perpendicular direction, as shown in Fig. 5C. Measurements were taken as the distance, in pixels, between the valleys confining the particle signal. For comparison, we also measured 200 randomly selected projections of the GAC atomic model in dimer and also in tetramer configuration.

Acknowledgments—We thank LNBio for accessibility to core facilities as well as for financial support. We are very grateful to Dr. Alessandra Girasole for expert technical support. We thank the Biotechnology group at the SGC Oxford for cloning and expression testing of glutaminase constructs. We thank Dr. Artur T. Cordeiro and Américo T. Ranzani for help with crystallographic data collection. We thank Dr. Wah Chi Wu for access to the cryo-electron microscopy facility at Baylor College of Medicine. We acknowledge the staff of all the beamlines used in this work. The KGA.ANK and GLS2.ANK data were collected during the RapiData 2015 Course for Data Collection and Structure Solving. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and National Institutes of Health, National Institute of General Medical Sciences Grant P41GM103393.

References
1. Bergström, J., Fürst, P., Norée, L. O., and Vinnars, E. (1974) Intracellular free amino acid concentration in human muscle tissue. J. Appl. Physiol. 36, 693–697
2. DeBerardinis, R. J., and Cheng, T. (2010) Q’s next: the diverse functions of glutamine in metabolism, cell biology and cancer. Oncogene 29, 313–324
3. Levine, A. J., and Puzio-Kuter, A. M. (2010) The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. Science 330, 1340–1344
4. Altman, B. J., Stine, Z. E., and Dang, C. V. (2016) From Krebs to clinic: glutamine metabolism to cancer therapy. Nat. Rev. Cancer 16, 619–634
5. Wise, D. R., DeBerardinis, R. J., Mancuso, A., Sayed, N., Zhang, X.-Y., Pfeiffer, H. K., Nissim, I., Daikhin, E., Yudkoff, M., McMahon, S. B., and Thompson, C. B. (2008) Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc. Natl. Acad. Sci. U.S.A. 105, 18782–18787
6. Bathore, M. G., Saumet, A., Rossi, J.-F., de Bettignies, C., Tempé, D., Lecleciel, C.-H., and Villalba, M. (2012) The NF-kB member p65 controls glutamine metabolism through miR-23a. Int. J. Biochem. Cell Biol. 44, 1448–1456
7. Wang J. B., Erickson, J. W., Fuji, R., Ramachandran, S., Gao, P., Dinavahi, R., Wilson, K. F., Ambrosio, A. L., Dias, S. M., Dang, C. V., and Cerione, R. A. (2010) Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. Cancer Cell. 18, 207–219
8. Colombo, S. L., Palacios-Callender, M., Frakich, N., De Leon, J., Schmitt, C. A., Boom, L., Davis, N., and Moncada, S. (2010) Anaphase-promoting complex/cyclosome-Cdh1 coordinates glycolysis and glutaminolysis with transition to S phase in human T lymphocytes. Proc. Natl. Acad. Sci. U.S.A. 107, 18868–18873
9. Lukey, M. J., Greene, K. S., Erickson, J. W., Wilson, K. F., and Cerione, R. A. (2016) The oncogenic transcription factor c-Jun regulates glutamine expression and sensitizes cells to glutaminase-targeted therapy. Nat. Commun. 7, 11321
10. Suzuki, S., Tanaka, T., Poyurovsky, M. V., Nagano, H., Mayama, T., Ohkubo, S., Lokshin, M., Hosokawa, H., Nakayama, T., Suzuki, Y., Sugano, S., Sato, E., Nago, T., Yokote, K., Tatsuno, I., and Prives, C. (2010) Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. Proc. Natl. Acad. Sci. U.S.A. 107, 7461–7466
11. Hu, W., Zhang, C., Wu, R., Sun, Y., Levine, A., and Feng, Z. (2010) Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. Proc. Natl. Acad. Sci. U.S.A. 107, 7455–7460
12. Szeliga, M., and Albrecht, J. (2015) Opposing roles of glutaminase isoforms in determining glioblastoma cell phenotype. Neurochem. Int. 88, 6–9
13. Szeliga, M., Bogacińska-Karaś, M., Różycka, A., Hilgier, W., Marquez, J., and Albrecht, J. (2014) Silencing of GLS2 and overexpression of GLS2 genes cooperate in decreasing the proliferation and viability of glioblastoma cells. Tumour Biol. 35, 1855–1862
14. Liu, J., Zhang, C., Lin, M., Zhu, W., Liang, Y., Hong, X., Zhao, Y., Young, K. H., Wu, W., and Feng, Z. (2014) Glutaminase 2 negatively regulates the PI3K/AKT signaling and shows tumor suppression activity in human hepato-cellular carcinoma. Oncotarget 5, 2635–2647
15. Krebs, H. (1935) Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues. Biochem. J. 29, 1951–1969
16. Curthoys, N. P., Kuhlenschmidt, T., Weiss, R. F., and Godfrey, S. S. (1976) Phosphate-dependent activity glutaminase from rat kidney and identity with glutaminase in response to acidosis from other tissues. Arch. Biochem. Biophys. 172, 162–167
17. Curthoys, N. P., and Watford, M. (1995) Regulation of glutamine activity and glutamine metabolism. Annu. Rev. Nutr. 15, 133–159
18. Elgadi, K. M., Meguid, R. A., Qian, M., Souba, W. W., and Abcouwer, S. F. (1999) Cloning and analysis of unique human glutaminase isoforms generated by tissue-specific alternative splicing. Physiol. Genomics 1, 51–62
19. Martin-Rufián, M., Tosina, M., Campos-Sandoval, J. A., Manzanares, E., Lobo, C., Segura, J. A., Alonso, F. J., Matés, J. M., and Márquez, J. (2012) The multidomain architecture of human glutaminases. Tumour Biol. 33, 207–219
20. Cassago, A., Ferreira, A. P., Ferreira, I. M., Fornezari, C., Gomes, E. R., Greene, K. S., Pereira, H. M., Garratt, R. C., Dias, S. M., and Ambrosio, A. L. (2012) Mitochondrial localization and structure-based phosphate activation mechanism of glutaminase C with implications for cancer metabolism. Proc. Natl. Acad. Sci. U.S.A. 109, 1092–1097
21. Campos-Sandoval, J. A., López de la Oliva, A. R., Lobo, C., Segura, J. A., Matés, J. M., Alonso, F. J., and Márquez, J. (2007) Expression of functional human glutaminase in baculovirus system: affinity purification, kinetic and molecular characterization. Int. J. Biochem. Cell Biol. 39, 765–773.
The multidomain architecture of human glutaminases

22. Olalla, L., Gutiérrez, A., Campos, J. A., Khan, Z. U., Alonso, F. J., Segura, J. A., Márquez, J., and Aledo, J. C. (2002) Nuclear localization of t-type glutaminase in mammalian brain. J. Biol. Chem. 277, 38939–38944

23. Ferreira, A. P., Cassago, A., Gonçalves Kde, A., Dias, M. M., Adamoski, D., Ascenção, C. F., Honorato, R. V., de Oliveira, J. F., Ferreira, I. M., Fornezari, C., Bettini, I., Oliveira, P. S., Paes Leme, A. F., Portugal, R. V., Ambrosio, A. L., and Dias, S. M. (2013) Active glutaminase C self-assembles into a supratetrameric oligomer that can be disrupted by an allosteric inhibitor. J. Biol. Chem. 288, 28009–28020

24. Brown, G., Singer, A., Proudfoot, M., Skarina, T., Kim, Y., Chang, C., Dehal, P., Satou, Y., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, J. J., Kuraku, S., McLaren, K., Matthews, L., McLaren, S., Hastings, K. E., Ho, I., Hotta, K., Huang, W., et al. (2008) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 319, 2157–2163

25. de Beer, T., Carter, R. E., Lobel-Rice, K. E., Sorkin, A., and Overduin, M. (1998) Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. Science 281, 1357–1360

26. Pace, N. R. (2009) Mapping the tree of life: progress and prospects. *Molecular Systematics*. 1584

27. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Irimia, M., Korzh, V., et al. (2015) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 510, 174–179

28. Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, J. J., Grigoriev, I. V., et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071

29. Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K. L., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., Hoon, S., Gandy, V., Roy, S. W., Ichimura, K., Kozh, V., et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505, 415–421

30. Holzer, S., Ban, N., and Klinge, S. (2013) Crystal structure of the yeast ribosomal protein rpS3 in complex with its chaperone Yar1. *J. Biol. Chem.* 288, 4154–4160

31. Hublin, E. L., Ting, L., Buckner, R. J., Gebreab, F., Gygij, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarini, V., Vaites, L., Orude, A., Rad, R., et al. (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* 162, 425–440

32. Clark, K., Karchs-Mizrahi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2016) GenBank. *Nucleic Acids Res.* 44, D67–D72

33. Lassmann, T., and Sonnhammer, E. L. (2005) Kalign: an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6, 298

34. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313

35. Sukumaran, J., and Holder, M. T. (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571

36. Schliep, K. P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593

37. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797

38. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* 278, 631–637

39. Mosavi, L. K., Minour, D. L. Jr., and Peng, Z.-Y. (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16029–16034

40. Zhang, C., Liu, J., Zhao, Y., Yue, X., Zhu, Y., Wang, X., Wu, H., Blanco, F., Li, S., Haffty, B. G., Hu, W., and Feng, Z. (2016) Glutaminase 2 is a novel negative regulator of small GTPase Rac1 and mediates p53 function in suppressing metastasis. *Elife* 5, e10727

41. Scott, A. P. (1987) Reproductive endocrinology of fish. In *Fundamentals of Comparative Vertebrate Endocrinology* (Chester-Jones, I., Ingleton, P. M., Phillips, and J. G., eds), pp. 223–256, Springer US, New York, NY

42. Delabarre, B., Gross, S., Fang, C., Gao, Y., Jha, A., Jiang, F., Song, J. I., Wei, W., and Hurov, J. B. (2011) Full-length human glutaminase C with an allosteric inhibitor. *Biochemistry* 50, 10764–10770

43. Olalla, L., Gutiérrez, A., Jiménez, A. J., López-Téllez, J. F., Khan, Z. U., Pérez, J., Alonso, F. J., de la Rosa, V., Campos-Sandoval, J. A., Segura, J. A., Aledo, J. C., and Márquez, J. (2008) Expression of the scaffolding PDZ protein glutaminase-interacting protein in mammalian brain. *J. Neurosci. Res.* 86, 281–292

44. Holzer, S., Ban, N., and Klinge, S. (2013) Crystal structure of the yeast ribosomal protein rpS3 in complex with its chaperone Yari. *J. Biol. Chem.* 425, 4154–4160

45. Robbin, G., Sato, T., Desplançq, D., Rochel, N., Weiss, E., and Martineau, P. (2014) Restricted diversity of antigen binding residues of antibodies revealed by computational alanine scanning of 227 antibody–antigen complexes. *J. Mol. Biol.* 426, 3729–3743

46. Li, Y., Meng, X., Xiang, Y., and Deng, J. (2010) Structure function studies of vaccinia virus host range protein k1 reveal a novel functional surface for ankyrin repeat proteins. *J. Virol.* 84, 3331–3338

47. Hesketh, G. G., Pérez-Dorado, I., Jackson, L. P., Wartosch, L., Schäfer, I. B., Gray, S. R., McCoy, A. J., Zeldin, O. B., Garman, E. F., Harbour, M. E., Evans, P. R., Seaman, M. N., Lizio, J. P., and Owen, D. J. (2014) VARP is recruited on to endosomes by direct interaction with retromer, where together they function in export to the cell surface. *Dev. Cell.* 29, 591–606

48. Pop, P. M., Philips, and J. G., eds), pp. 223–256, Springer US, New York, NY

49. Sukumaran, J., and Holder, M. T. (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571

50. Schliep, K. P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593

51. Muhire, B. M., Varsani, A., and Martin, D. P. (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* 9, e108277

52. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797

53. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* 278, 631–637

54. Huang, X., and Miller, W. (1991) A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12, 337–357

55. J. Biol. Chem. (2017) 292(27) 11572–11585
61. Levy, A., Sela, N., and Ast, G. (2008) TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. Nucleic Acids Res. 36, D47–D52

62. Leslie, A. G. W., and Powell H. R. (2007) Processing diffraction data with mosflm. In Evolving methods for macromolecular crystallography (Read, R. J., and Sussman, J. L., eds), pp. 41–51, Springer, Dordrecht, Netherlands

63. Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132

64. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. 69, 1204–1214

65. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674

66. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. 66, 479–485

67. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. D Biol. Crystallogr. 60, 2184–2195

68. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. I., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221

69. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501

70. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282

71. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 25, 495–503

72. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1: towards automated and web-supported small-angle scattering data analysis. J. Appl. Crystallogr. 40, s223–s228

73. van Heel, M., Portugal, R., Rohou, A., Linnemayr, C., Bebeacua, C., Schmidt, R., Grant, T., and Schatz, M. (2012) Four-dimensional cryo-electron microscopy at quasi-atomic resolution: IMAGIC4D. Int. Tables. Acta Crystallogr. Vol. F, chi. 19.9, 624–628

74. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, L., and Ludtke, S. J. (2007) EMAN2: An extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 36–46

75. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J., and Penczek, P. A. (2012) Iterative stable alignment and clustering of 2D transmission electron microscopy images. Structure 20, 237–247

76. Pettiens, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612
The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats
Camila Cristina Pasquali, Zeyaul Islam, Douglas Adamoski, Igor Monteze Ferreira, Ricardo Diogo Righeto, Jefferson Bettini, Rodrigo Vilares Portugal, Wyatt Wai-yin Yue, Ana Gonzalez, Sandra Martha Gomes Dias and Andre Luis Berteli Ambrosio

J. Biol. Chem. 2017, 292:11572-11585.
doi: 10.1074/jbc.M117.787291 originally published online May 19, 2017

Access the most updated version of this article at doi: 10.1074/jbc.M117.787291

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2017/05/25/M117.787291.DC1

This article cites 74 references, 19 of which can be accessed free at http://www.jbc.org/content/292/27/11572.full.html#ref-list-1
SUPPLEMENTAL MATERIAL

The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats

Camila Cristina Pascoal¹*, Zeyaul Islam¹*, Douglas Adamoski¹*, Igor Monteze Ferreira¹², Ricardo Diogo Righeto³⁴‡, Jefferson Bettini³, Rodrigo Villares Portugal³, Wyatt W Yue², Ana Gonzalez⁵, Sandra Martha Gomes Dias¹†, and Andre Luis Berteli Ambrosio¹†

From the ¹Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, 13083-100, Brazil; the ²Structural Genomics Consortium, University of Oxford, Oxford, OX3 7DQ, United Kingdom; the ³Laboratório Nacional de Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, 13083-100, Brazil, the ⁴School of Electrical and Computer Engineering, University of Campinas, SP, 13083-852, Brazil, and the ⁵Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.

*CCP, ZI and DA contributed equally to this work.
‡current address: Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Basel, CH-4058, Switzerland
### Supplemental Table S1. Survey in the Protein Data Bank (PDB) of deposited crystal structures containing ankyrin (ANK) repeats.

| ANK-containing proteins on PDB                                      | Organism                  | PDB ID | Biological Assembly | Number of ANK repetitions in the structure |
|-------------------------------------------------------------------|---------------------------|--------|---------------------|------------------------------------------|
| 2-5A-dependent ribonuclease                                       | Homo sapiens              | 4oav   | homo-2-mer          | 9                                        |
| 26S proteasome non-ATPase regulatory subunit 10                   | Homo sapiens              | 1qym   | monomer             | 7                                        |
| ABAYE2397                                                         | Acinetobacter baumannii   | 5d66   | monomer             | 6                                        |
| Ankyrin repeat and SOCS box protein 11                           | Homo sapiens              | 4uuc   | monomer             | 6                                        |
| Ankyrin repeat and SOCS box protein 9                            | Homo sapiens              | 3d9h   | monomer             | 7                                        |
| Ankyrin repeat domain-containing protein 27                       | Homo sapiens              | 4cym   | hetero-4-mer        | 5                                        |
| Ankyrin repeat domain-containing protein 2A                       | Arabidopsis thaliana      | 4tum   | monomer             | 4                                        |
| Ankyrin repeat family A protein 2                                 | Homo sapiens              | 3so8   | monomer             | 5                                        |
| Ankyrin repeat family protein                                     | Legionella pneumophila    | 2aja   | monomer             | 10                                       |
| Ankyrin repeat-containing protein                                 | Legionella pneumophila    | 4zhb   | monomer             | 3                                        |
| Ankyrin repeat-containing protein YAR1                            | Saccharomyces cerevisiae  | 4bsz   | hetero-8-mer        | 4                                        |
| Ankyrin-1                                                         | Homo sapiens              | 1n11   | monomer             | 12                                       |
| Ankyrin-2, Ankyrin-1                                              | Homo sapiens / Mus musculus (quimera) | 4rlv   | monomer             | 24                                       |
| Anti-ige inhibitor e2_79                                          | Escherichia coli          | 4grg   | hetero-4-mer        | 4                                        |
| Apoptosis-stimulating of p53 protein 2                            | Homo Sapiens              | 1ycs   | hetero-2-mer        | 4                                        |
| Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 1 | Homo sapiens              | 3jue   | monomer             | 4                                        |
| Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2 | Mus musculus              | 1dcq   | monomer             | 3                                        |
| Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 3 | Homo Sapiens              | 2b0o   | homo-2-mer          | 4                                        |
| B-cell lymphoma 3 protein                                         | Homo sapiens              | 1k1a   | monomer             | 7                                        |
| Bd3460                                                            | Bdellovibrio bacteriovorus | 5cea  | monomer             | 6                                        |
| BRCA1-associated RING domain protein 1                            | Homo sapiens              | 3c5r   | monomer             | 4                                        |
| Cyclin-dependent kinase 4 inhibitor D                             | Homo sapiens              | 1bd8   | monomer             | 5                                        |
| Cyclin-dependent kinase 6 inhibitor                               | Homo sapiens              | 1mx2   | monomer             | 5                                        |
| Cyclin-dependent kinase inhibitor 2A                              | Homo Sapiens              | 1bi7   | hetero-4-mer        | 4                                        |
The multidomain architecture of human glutaminases

| Protein Name                                                                 | Organism                        | PDB Code | Type          | Domain Count |
|------------------------------------------------------------------------------|---------------------------------|----------|---------------|--------------|
| DNA-binding protein RFXANK                                                   | *Homo sapiens*                  | 3uxg     | monomer       | 5            |
| EF_0377                                                                      | *Enterococcus faecalis*         | 3hra     | monomer       | 6            |
| Eukaryotic huntingtin interacting protein B                                  | *Legionella pneumophila*        | 5czy     | monomer       | 5            |
| GA-binding protein subunit beta-1                                            | *Mus musculus*                  | 1awc     | hetero-2-mer  | 5            |
| Histone-lysine N-methyltransferase EHMT1                                     | *Homo sapiens*                  | 3b7b     | monomer       | 7            |
| Integrin-linked protein kinase                                                | *Homo sapiens*                  | 36aq     | hetero-2-mer  | 5            |
| K1L                                                                          | *Vaccinia virus*                | 3kea     | monomer       | 9            |
| KN motif and ankyrin repeat domain-containing protein 2                      | *Homo sapiens*                  | 5d68     | monomer       | 4            |
| Krev interaction trapped protein 1                                           | *Homo sapiens*                  | 4hbd     | monomer       | 6            |
| LMJF_29_1100                                                                 | *Leishmania major*              | 3ljn     | monomer       | 8            |
| Methionyl-tRNA formyltransferase                                             | *Providencia alcalifaciensis*   | 4xzcz    | homo-2-mer    | 4            |
| Myotrophin                                                                   | *Homo sapiens*                  | 3aaa     | hetero-2-mer  | 3            |
| Neurogenic locus notch homolog protein 1                                     | *Mus musculus*                  | 1lymp    | monomer       | 4            |
| Neurogenic locus notch homolog protein 1                                     | *Homo Sapiens*                  | 1yyh     | monomer       | 6            |
| Neurogenic locus Notch protein                                               | *Drosophila melanogaster*       | 1ot8     | monomer       | 6            |
| NF-kappa-B inhibitor alpha                                                    | *Homo Sapiens*                  | 1ikn     | hetero-3-mer  | 6            |
| NF-kappa-B inhibitor beta                                                    | *Mus musculus*                  | 1k3z     | hetero-3-mer  | 6            |
| Nuclear factor NF-kappa-B p100 subunit                                       | *Homo sapiens*                  | 4ot9     | homo-2-mer    | 7            |
| Osteoclast-stimulating factor 1                                              | *Homo sapiens*                  | 3eh9     | monomer       | 4            |
| Palmitoyltransferase ZDHHC17                                                 | *Homo sapiens*                  | 3eu9     | homo-6-mer    | 7            |
| Phosphocholine transferase AnkX                                              | *Legionella pneumophila*        | 4ber     | monomer       | 4            |
| Probable 26S proteasome regulatory subunit p28                                | *Saccharomyces cerevisiae*      | 1ixv     | monomer       | 7            |
| Protein lin-12                                                                | *Caenorhabditis elegans*        | 2fo1     | hetero-3-mer  | 8            |
| Protein phosphatase 1 regulatory subunit 12A                                  | *Gallus gallus*                 | 1s70     | hetero-2-mer  | 7            |
| Putative ankyrin repeat protein TV1425                                        | *Thermoplasma volcanium*        | 2rfm     | monomer       | 6            |
| Regulatory protein SWI6                                                       | *Saccharomyces cerevisiae*      | 1sw6     | homo-2-mer    | 5            |
| RelA-associated inhibitor                                                    | *Homo Sapiens*                  | 2vge     | monomer       | 4            |
| Ribonuclease L                                                               | *Sus scrofa*                    | 4o1o     | homo-2-mer    | 9            |
| Signal recognition particle 43 kDa protein, chloroplastic                    | *Arabidopsis thaliana*          | 3deo     | monomer       | 4            |
| Protein Name                                                                 | Species            | PDB Code | Assembly   | Count |
|------------------------------------------------------------------------------|--------------------|----------|------------|-------|
| Tankyrase-1                                                                  | Mus musculus      | 3utm     | hetero-3-mer | 9     |
| Tankyrase-2                                                                  | Homo sapiens      | 3twr     | monomer    | 5     |
| Tankyrase-2                                                                  | Homo sapiens      | 3twq     | homo-2-mer | 5     |
| Transient receptor potential cation channel subfamily A member 1, Maltose-binding periplasmic protein chimera | Homo sapiens/Escherichia coli (chimera) | 3j9p     | homo-4-mer | 5     |
| Transient receptor potential cation channel subfamily V member 1             | Rattus norvegicus | 3j5p     | homo-4-mer | 6     |
| Transient receptor potential cation channel subfamily V member 2             | Homo Sapiens      | 2f37     | monomer    | 6     |
| Transient receptor potential cation channel subfamily V member 2             | Rattus norvegicus | 2eta     | monomer    | 6     |
| Transient receptor potential cation channel subfamily V member 3             | Mus musculus      | 4n5q     | monomer    | 6     |
| Transient receptor potential cation channel subfamily V member 4             | Homo Sapiens      | 4dx1     | monomer    | 6     |
| Transient receptor potential cation channel subfamily V member 6             | Mus musculus      | 2rfa     | monomer    | 6     |
| TRPV2                                                                        | Oryctolagus cuniculus | 5an8   | homo-4-mer | 5     |
| Ubiquitin thioesterase ZRANB1                                                | Homo Sapiens      | 3zrh     | monomer    | 2     |
| uncharacterized protein                                                       | Leptotrichia buccalis | 3t8k   | homo-2-mer | 5     |
| Vanilloid receptor-related osmotically activated channel protein             | Gallus gallus     | 3jxi     | monomer    | 6     |
Supplemental Figure S1. Crystallographic structures containing ankyrin-to-ankyrin associations involved in homo-oligomeric assembly. (A) Ankyrin repeat cluster (ARC4) of human Tankyrase 2 (PDB ID 3TWQ); (B) crystal structure of a functionally unknown protein from Leptotrichia buccalis (PDB ID 3T8K); (C) ANK domain of Huntingtin interacting protein 14 (PDB 3EU9); (D) fragment containing ANK repeats of the yeast cell cycle regulator Swi6 (PDB ID 1SW6); (E) C-terminal domain of p100/NF-kB2 (PDB ID 4OT9).