Molecular & Biochemical Parasitology 175 (2011) 192–195

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

Short communication

Regions of intrinsic disorder help identify a novel nuclear localization signal in *Toxoplasma gondii* histone acetyltransferase TgGCN5-B

Stacy E. Dixon[a], Micah M. Bhatti[a,1], Vladimir N. Uversky[b,c], A. Keith Dunker[b], William J. Sullivan Jr.[a,*]

[a] Department of Pharmacology and Toxicology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[b] Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[c] Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

ARTICLE INFO

Article history:
Received 22 July 2010
Received in revised form 18 October 2010
Accepted 22 October 2010
Available online 3 November 2010

Keywords:
Apicomplexa
Parasite
Cellular trafficking
GCN5
Chromatin
Epigenetics

ABSTRACT

We have previously shown that protozoan parasites, such as *Toxoplasma gondii*, contain a high prevalence of intrinsically disordered regions in their predicted proteins. Here, we determine that both TgGCN5-family histone acetyltransferases (HATs) contain unusually high levels of intrinsic disorder. A previously identified basic-rich nuclear localization signal (NLS) in the N-terminus of TgGCN5-A is located within such a region of predicted disorder, but this NLS is not conserved in TgGCN5-B. We therefore analyzed the intrinsically disordered regions of TgGCN5-B for basic-rich sequences that could be indicative of a functional NLS, and this led to the identification of a novel NLS for TgGCN5-B, RPAENKKRGR. The functionality of the GCN5-B NLS was validated experimentally and has predictive value. These studies demonstrate that basic-rich sequences within regions predicted to be intrinsically disordered constitute criteria for a candidate NLS.

The obligate intracellular protozoan *Toxoplasma gondii* (Apicomplexa) is a serious opportunist pathogen. Completion of genome sequencing revealed that ~58% of predicted *Toxoplasma* genes encode hypothetical proteins of unknown function (toxodb.org). The discovery of new protein motifs is essential for improving predictions about the location and function of unknown proteins. We have previously determined that the genomes of early branching eukaryotic protozoa contain a large proportion of predicted proteins with significant amounts of intrinsic disorder [1]. Disordered regions are characterized by moderate to low amino acid sequence complexity with very few bulky, hydrophobic amino acids and with an enrichment of polar and charged amino acids and the structure-breaking proline, and can be predicted using computational methods [2]. Determining the degree of disorder in a protein can assist in predicting the biological relevance of a given domain, as many regions of disorder map to areas of protein–protein interaction or post-translational modification [3–6].

We have previously described the presence of lengthy (600–800 amino acids), unconserved N-terminal extensions on the two GCN5-family member histone acetyltransferases (HATs), TgGCN5-A and -B [7,8]. These extensions are not present on the GCN5 homologues in other lower eukaryotes, save the fellow apicomplexan parasite *Plasmodium falciparum* [9], and they have no currently known protein motifs that would help indicate their function. Previously, we determined that the N-terminus of TgGCN5-A plays a role in localizing the HAT to the parasite nucleus by virtue of a unique nuclear localization signal (NLS) [10]. It was also noted that if TgGCN5-B were deprived of its N-terminal extension, the truncated protein was mainly in the cytoplasm [8]. However, the NLS for TgGCN5-A is not conserved in TgGCN5-B, suggesting that TgGCN5-B uses a different NLS to gain access to the *Toxoplasma* nucleus.

Primary amino acid sequences for TgGCN5-A and -B were analyzed using *PONDR*® *VLXT*, *VL3*, *VSL2*, and *PONDR-FIT* algorithms to identify regions of intrinsic disorder (Suppl. methods). In both cases the bromodomain, which recognizes acetylated lysine residues [11], and HAT catalytic domain are predicted to be highly structured (Fig. 1). In contrast, the various *PONDR*®s concur that the remainder of each TgGCN5 is likely to be remarkably disordered.

* Corresponding author at: Department of Pharmacology and Toxicology, Indiana University School of Medicine, 635 Barnhill Drive, MS A-525, Indianapolis, IN 46202, USA. Tel.: +1 317 274 1573; fax: +1 317 274 7714.
E-mail address: wjsulliv@iupui.edu (W.J. Sullivan Jr.).
[1] Present address: Pediatric Infectious Diseases, University of Chicago, Chicago, IL 60637, USA.

0166-6851/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2010.10.009

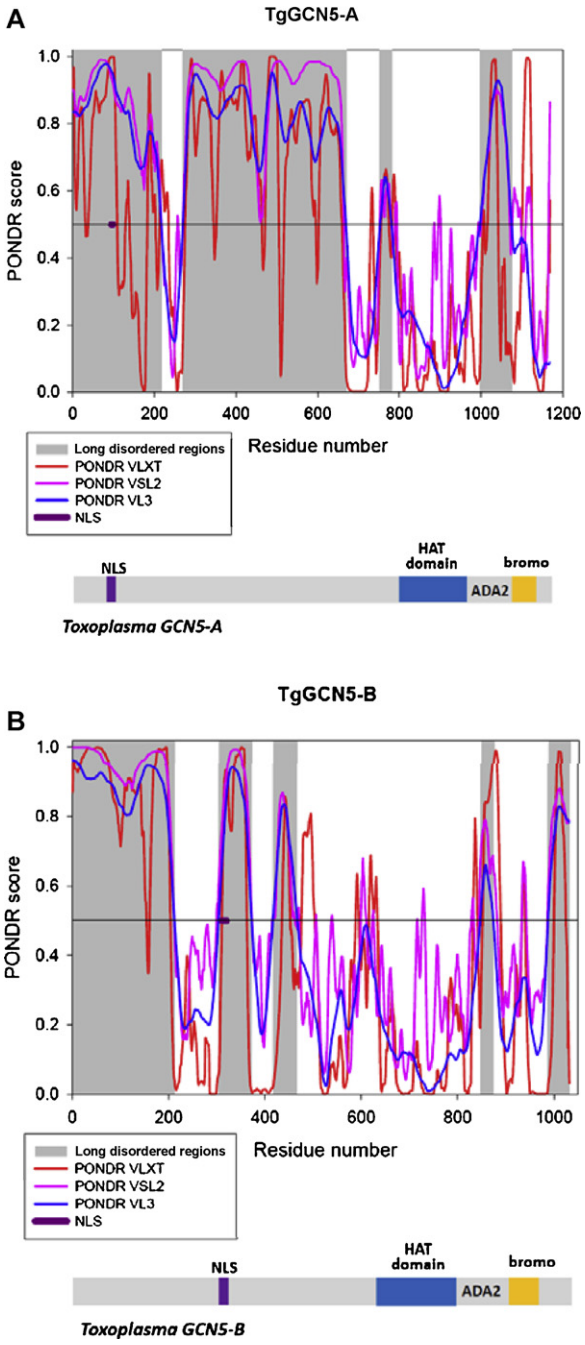

Fig. 1. Intrinsic disorder predictions and domain structures of TgGCN5-A (A) and TgGCN5-B (B). The per-residue propensity for intrinsic disorder was evaluated using a set of *PONDR* algorithms (VL-XT – red lines; VSL2 – purple lines; VL3 – blue lines). In *PONDR* plots (top graphs of each plot), segments with scores above 0.5 correspond to the disordered regions, whereas those below 0.5 correspond to the ordered regions/binding sites. Long regions of predicted disorder are highlighted in gray. Position of the NLS is shown in purple. Below each plot a cartoon showing the key domains of each TgGCN5 protein is shown: HAT domain in blue and bromodomain (bromo) in gold, separated by the ADA2-interacting domain (ADA2). NLS is shown in purple.

The most extensive predicted disorder is located within the N-terminal extension, followed by the ADA2-interacting domain and the extreme C-terminal tail (Fig. 1). Consistent with the idea that unstructured domains engage in protein–protein interactions [4], we have previously verified that the ADA2-interacting domains of TgGCN5-A and -B interact with one or both ADA2 co-activator homologues present in *Toxoplasma* [8]. Additionally, the NLS elucidated for TgGCN5-A (RKRVKR, amino acids 94–99) is embedded in a region of intrinsic disorder (Fig. 1).

We sought to determine the NLS for TgGCN5-B by examining basic-rich stretches contained within a disordered region (analogous to the TgGCN5-A NLS). Residues 316–320 (KKRGR) best fit these criteria so we generated plasmids designed to express truncated, FLAG-tagged forms of recombinant TgGCN5-B in *Toxoplasma*, as we did previously to map the NLS for TgGCN5-A [10]. FLAG-tagged TgGCN5-B lacking the first 320 amino acids ($_{FLAG}$GCN5-BΔ320) showed cytoplasmic localization (supplemental Fig. S1A). However, FLAG-tagged TgGCN5-B lacking the first 315 amino acids ($_{FLAG}$GCN5-BΔ315), which retains the KKRGR motif, still displayed cytoplasmic localization (data not shown). Moreover, KKRGR fused to *E. coli* β-galactosidase (β-gal) expressed in *Toxoplasma* failed to gain access to the nucleus (data not shown), suggesting KKRGR is necessary but not sufficient for nuclear localization. We hypothesized that additional residues upstream of this basic-rich stretch are required for proper compartmentalization of TgGCN5-B. A new construct lacking the first 304 residues, $_{FLAG}$GCN5-BΔ304, supports this idea (supplemental Fig. S1B). To define the minimal NLS motif, additional deletion constructs were designed that lacked either the first 310 ($_{FLAG}$GCN5-BΔ310) or 313 ($_{FLAG}$GCN5-BΔ313) amino acid residues. While the construct $_{FLAG}$GCN5-BΔ313 was cytoplasmic (Fig. 2A), $_{FLAG}$GCN5-BΔ310 was able to enter the parasite nucleus (Fig. 2B). When we excised just the 10 amino acids from 311 to 320 (RPAENKKRGR) from full-length TgGCN5-B ($_{FLAG}$GCN5-BΔNLS), the recombinant protein was restricted to the parasite cytoplasm (Fig. 2C), validating that these 10 residues are necessary for nuclear localization. To demonstrate that the elucidated NLS is sufficient for nuclear localization, we attached the RPAENKKRGR residues onto the C-terminus of *E. coli* β-galactosidase (β-gal) followed by a FLAG tag and monitored the distribution of the fusion protein within the parasites. While β-gal is normally restricted from the parasite nucleus (Fig. 2D), attachment of the TgGCN5-B NLS resulted in virtually all of the fusion protein translocating to the nucleus (Fig. 2E). To rule out the possibility that the FLAG-epitope following the NLS was contributing to the redistribution of β-gal, we replaced it with an HA tag. Results show that β-gal-NLS$_{HA}$ was also nuclear (supplemental data Fig. S2). We conclude that the 10 residue stretch, RPAENKKRGR (amino acids 311–320), is necessary and sufficient to serve as an NLS in *Toxoplasma*.

Searches of the NLS database [12] did not reveal an entry identical to the RPAENKKRGR NLS of TgGCN5-B, revealing that it is a novel monopartite NLS. We investigated the importance of the upstream RP residues on the function of the NLS by creating further mutations in $_{FLAG}$GCN5-BΔ310. Point mutation of either the Arg or Pro to Ala did not hinder nuclear localization (data not shown); however, when both residues were mutated to alanines, nuclear localization was significantly attenuated (Fig. 2F). These studies suggest that the RP residues upstream of the basic core cluster are critical for efficient nuclear localization of TgGCN5-B. This result is in marked contrast to what was observed for TgGCN5-A, whose basic cluster of RKRVKR residues is necessary and sufficient to operate as an NLS [10].

To determine the utility of the TgGCN5-B NLS as a predictor for nuclear localization of other *Toxoplasma* proteins, the ToxoDB was searched for gene predictions harboring a similar motif. TgGCN5-B was the only protein in possession of the exact 10 residue NLS. When permutations were allowed for residues that were not basic or proline (RPxxxKKRxR, with "x" being any amino acid), three predicted proteins were identified: two hypothetical proteins (TGGT1_113380 and TGME49_091900) and one with a PHD-finger domain (TGGT1_071200), commonly found on