Human Guanylate Binding Protein 4: Purification, Functional Assay, and Protein Sequence Analysis

Blake Roberts, Taiwo Scholes Adewole, Qian Yin

Abstract

When a host system faces infection, immediate multipurpose responses become necessary, as the host must be ready to defend against any type of pathogen. Ubiquitous defense proteins such as the Guanylate Binding Proteins (GBPs) are vital in this effort. GBPs are interferon-γ induced proteins that bind to guanosine molecules as a substrate and can defend against microbial invasion. Humans have seven paralogs of GBP in their genome, each with an N-terminal GTPase domain, a middle domain, and a C-terminal effector domain. In this study, we have taken a deeper look into human GBP4.

Though all seven paralogs have similar functions, they retain stark differences in their nucleoside-binding capabilities and signaling. To understand the molecular basis for these differences, we examined the protein sequences of each human GBP paralog and compared them to GBP4. In doing so, we can determine relationships between structure and function and discover conserved residues, domain features, and functional motifs across each human GBP. Coinciding with the protein analysis, we have also purified GBP4's GTPase domain and N-terminus to middle domain to homogeneity to assay their ability to bind to different guanosine nucleosides.

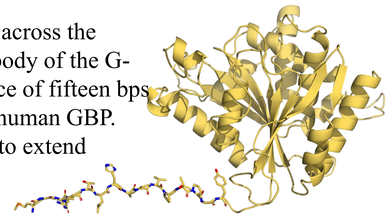
Understanding GBP4's immunological niche is critical, as previous studies have displayed GBP4's relevance in innate immune signaling. Currently, there are no published studies of the biochemical or structural features of GBP4. Characterizing GBP4 is a major step toward comprehending the GBPs in their entirety and utilizing them in future medicine.

Analysis of key motifs

```
GBP4        MGERTLHAAVPTPGYPESESIMMAPICLVENQEEQLTVNSKALEILDKISQPVVVVAIVG  60
GBP6        ----------------MESGPKMLAPVCLVENNNEQLLVNQQAIQILEKISQPVVVVAIVG  45
GBP7        ----------------MASEIHMPGPVCLTENTKGHLVVNSEALEILSAITQPVVVVAIVG  45
GBP5        ----------------MALEIHMSDPMCLIENFNEQLKVNQEALEILSAITQPVVVVAIVG  45
GBP1        ----------------MASEIHMTGPMCLIENTNGRLMANPEALKILSAITQPMVVVAIVG  45
GBP3        ----------------MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVG  45
GBP2        ----------------MAPEINLPGPMSLIDNTKGQLVVNPEALKILSAITQPVVVVAIVG  45
                                                 End of G-domain
GBP4 299    GIIVTGKRLGTLVVTYVDAINSGAVPCLENAVTALAQLENPAAVQRAADHYSQQMAQQLR  358
GBP6 284    GITVTGNRLGTLAVTYVEAINSGAVPCLENAVITLAQRENSAAVQRASDYYSQQMAQRVK  343
GBP7 284    GILVTGNRLGMLVETYLDAINSGATPCLENAMAVLAQCENSAAVQRAANHYSQQMAQQVR  343
GBP5 283    GIMVNGSRLKNLVLTYVNAISSGDLPCIENAVLALAQRENSAAVQKAIAHYDQQMGQKVQ  342
GBP1 285    GIQVNGPRLESLVLTYVNAISSGDLPCMENAVLALAQIENSAAVQKAIAHYEQQMGQKVQ  344
GBP3 283    GIKVNGPRLESLVLTYINAISRGDLPCMENAVLALAQIENSAAVQKAIAHYDQQMGQKVQ  342
GBP2 283    GIPVNGPRLESLVLTYVNAISSGDLPCMENAVLALAQIENSAAVEKAIAHYEQQMGQKVQ  342
                                          End of middle domain
GBP4 479    LQSQVVVEESILQSDKALTAGEKAIAAERAMKEAAEKEQELLREKQKEQQQMMEAQERSF  538
GBP6 464    LESQMVIEESILQSDKALTDREKAVAVDRAKKEAAEKEQELLKQKLQEQQQQMEAQVKSR  523
GBP7 464    LQSQVVIEESILQSDKALTAGEKAIAAKQAKKEAAEKEQELLRQKQKEQQQMMEAQERSF  523
GBP5 463    LKSKESVSHAILQTDQALTETEKKKKEAQVKAEAEKAEAQRLAAIQRQNEQMMQERERLH  522
GBP1 465    LKSKESMTDAILQTDQTLTEKEKEIEVERVKAESAQASAKMLQEMQRKNEQMMEQKERSY  524
GBP3 463    LKSKESVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIKYQQMMEEKEKSY  522
GBP2 463    LESKEDVADALLQTDQSLSEKEKAIEVERIKAESAEAAKKMLEEIQKKNEEMMEQKEKSY  522

GBP6 524    KENIAQLKEKLQMEREHLLREQIMMLEHTQKVQNDWLHEGFKKKYEEMNAEISQFKRMID  583
GBP4 539    QEYMAQMEKKLEEERENLLREHERLLKHKLKVQEEMLKEEFQKKSEQLNKEINQLKEKIE  598
GBP7 524    QENIAQLKKKMERERENYMRELRKMLSHKMKVLEELLTEGFKEIFESLNEEINRLKEQIE  583
GBP5 523    QEQVRQME----IAKQNWLAEQQKMQEQQMQEQAAQLSTTFQAQNRSLLSELQHAQRTVN  578
GBP1 525    QEHLKQLTEKMENDRVQLLKEQERTLALKLQEQEQLLKEGFQKESRIMKNEIQDLQTKMR  584
GBP3 523    QEHVKQLTEKMERERAQLLEEQEKTLTSKLQEQARVLKERCQGESTQLQNEIQKLQKTLK  582
GBP2 523    QEHVKQLTEKMERDRAQLMAEQEKTLALKLQEQERLLKEGFENESKRLQKDIWDIQMRSK  582
```

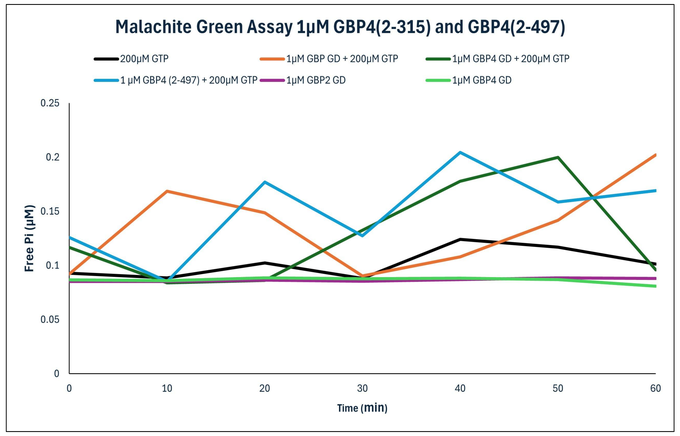
The G-domain is a highly conserved region across the human GBPs, with minor variations in the body of the G-domain. However, GBP4 contains a sequence of fifteen bps positioned before the start site of any other human GBP. This mildly hydrophobic motif is predicted to extend outside of the main G-domain structure

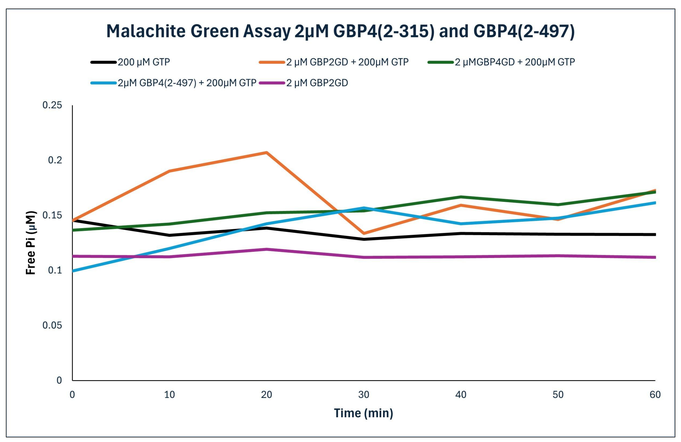


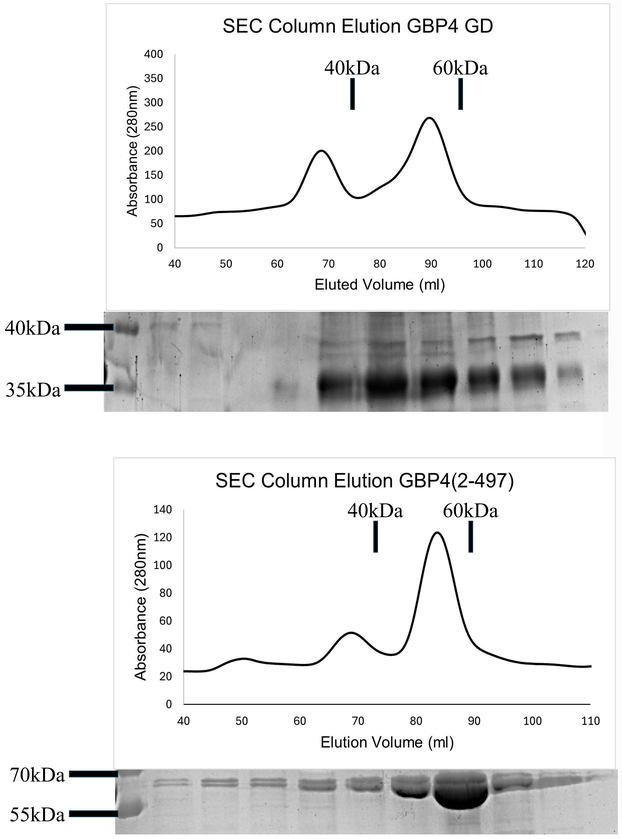
At the start of the middle domains, similar conservation can be observed, however, the variation increases as the sequence gets closer to the effector domain. The effector domains of each GBP are highly variable, especially among different "branches" of GBPs in the phylogeny.

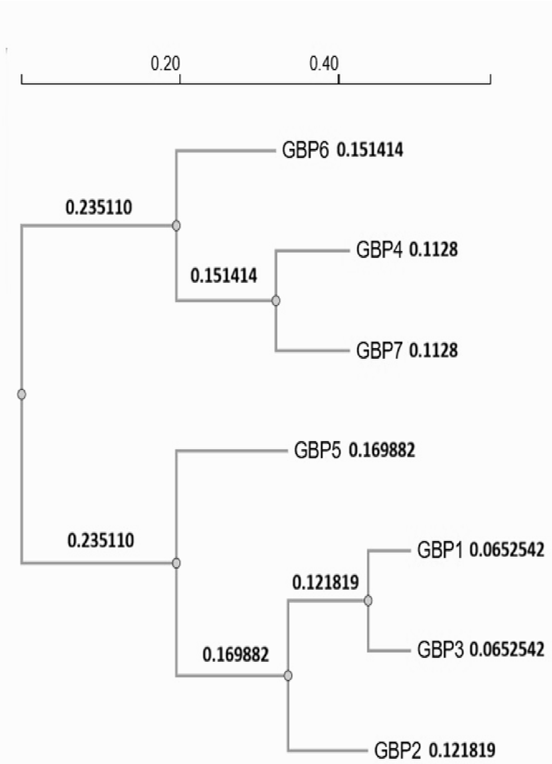
The effector domain becomes disarrayed and specialized across the GBPs, however, a pattern can be observed even in this highly unconserved region: GBP4 and GBP7 still retain high fidelity to each other, as do GBP1 and GBP3. This relatedness is reflected in the GBP phylogeny, which also reflects that these pairs motifs happen to be the most unrelated to each other as well.

Purification and Hydrolysis Assay









Next Steps

- The unique 15 bp at the start of GBP4 could be examined on further to discover functional characteristics.
- The results of the sequence alignment will be further corroborated with positional information of conserved and different sequences.
- The full length of GBP4 cannot be successfully expressed in bacterial cells. To circumvent this, insect cells will be utilized to grow the protein for further assays and the creation of a confirmable structural model.
- Further trials of the Malachite Green Assay will be performed to achieve more stable and conclusive results