A Specialized Version of the HD Hydrolase Domain Implicated in Signal Transduction

Michael Y. Galperin*, Darren A. Natale, L. Aravind, and Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Recently, a superfamily of proteins containing a previously undetected domain with predicted metal-dependent phosphohydrolase activity has been described and designated the HD superfamily, after the principal conserved residues implicated in catalysis (Aravind and Koonin, 1998). In the course of our analysis of ancient conserved regions in microbial genomes (Koonin et al., 1998), we found a distinct version of this domain which is encoded in one to three copies in the genomes of *Aquifex aeolicus*, *Borrelia burgdorferi*, *Synechocystis* sp. and *Treponema pallidum*, but is dramatically expanded in the genomes of *Thermotoga maritima* and *Clostridium acetobutylicum* (Figure 1). Compared with the consensus HD domain (Aravind and Koonin, 1998), this version contains a number of additional highly conserved residues; hereinafter we refer to it as the HD-GYP domain, after the characteristic sequence signatures (Figure 1). This domain was also detected in previously uncharacterized proteins from *Wolinella succinogenes* (Kreis-Kleinschmidt et al., 1995), *Bacillus halodurans* (Takami et al., 1999), *Pseudomonas aeruginosa* and *Bordetella pertussis* (Figure 1). The HD-GYP domain is missing in *E. coli* and *B. subtilis*. Remarkably, however, in other γ-proteobacteria, such as *Shewanella putrefaciens* and *Vibrio cholerae*, it is present in up to 8 copies (data not shown).

While none of the proteins that contain the HD-GYP domain has ever been characterized experimentally, the spectrum of the domains that are associated with HD-GYP in multidomain proteins (Figure 2) suggests that it is probably involved in signal transduction. In *Synechocystis* sp., both copies of the HD-GYP domain are found in proteins that also contain CheY-like receiver domains of the two-component signal transduction system (Pao and Saier, 1995; Volz and Matsumura, 1991). A similar CheY–HD-GYP domain organization is found in two *T. maritima* proteins, TM0186 and TM1147 (Figure 2). Two other proteins from *T. maritima*, TM1170 and TM1682, combine the HD-GYP domain with extracytoplasmic ligand-binding domains, which are closely related, respectively, to periplasmic solute-binding protein components of the ATP-dependent transport systems (Tam and Saier, 1993) and the extracytoplasmic part of methyl-accepting chemotaxis proteins of *Bacillus subtilis*, such as McpA and McpB (Hanlon and Ordal, 1994). Such a combination of an extracytoplasmic ligand-binding domain and a cytoplasmic HD-GYP domain, connected by a transmembrane segment, has the same topology as methyl-accepting proteins and many sensor kinases, which further supports the participation of the HD-GYP domain in signal transduction. Finally, in the Aq_2027 protein from *A. aeolicus*, the HD-GYP domain is found together with the GGDEF domain (Figure 2). The latter domain has been recently identified in diguanylate cyclases and phosphodiesterases involved in the regulation of cellulose synthesis in *Acetobacter xylinum* (Tal et al., 1998) and in a variety of bacterial signalling proteins in combination with CheY, PAS, and HAMP domains (Hecht and Newton, 1995; Aravind and Ponting, 1999). The GGDEF domain is often associated with another uncharacterized domain, EAL, in particular, in diguanylate cyclases and phosphodiesterases (Tal et al., 1998; Aravind and Ponting, 1999). Remarkably, however, the combination of the HD-GYP and EAL domains is not seen in any of the currently available microbial genomes. Moreover, the number of copies of the HD-GYP domain in complete genomes generally correlates with prevalence of the GGDEF domain over the EAL domain (Table 1). Furthermore, the HD-GYP family of HD proteins so far is lacking in archaea and eukaryotes and so are the GGDEF and EAL domains. This suggests that HD-GYP domain might be also involved in cyclic diguanylate-mediated signaling. The HD superfamily is related to the cAMP/cGMP phosphodiesterases that are involved in eukaryotic signalling (Aravind and Koonin, 1998). Therefore it is seems plausible that the HD-GYP family proteins are likely to possess a diguanylate phosphodiesterase activity and complement the function that, in the characterized diguanylate phosphodiesterases, is performed by the EAL domain.

![Table 1. Distribution of Three Domains Implicated in Signal Transduction in Complete Microbial Genomes](http://www.cric.com/sequence_center/bacterial_genomes)
Figure 1. Multiple alignment of HD–GYP domains. The proteins are listed under their names in complete genomes (left column) and their unique gene identification (gi) numbers in the GenBank protein database (right column); the numbers indicate positions of the first and the last residues in each protein, where available, and the distances between the aligned segments. Species name abbreviations are as follows: Aq, Aquifex aeolicus; Bb, Borrelia burgdorferi; Bhal, Bacillus halodurans; Cl.acetobut., Clostridium acetobutylicum; Sy, Synechocystis sp.; Tma, Thermotoga maritima; Tpa, Treponema pallidum; Vcho, Vibrio cholerae; Wsuc, Wolinella succinogenes. Reverse shading indicates most conserved amino acid residues that are probably involved in metal and/or substrate binding. Grey shading indicates conserved uncharged amino acid residues, other conserved residues are in bold. The secondary structure of the HD domain is as predicted by PHDsec program (Rost and Sander, 1993); H indicates predicted α-helical segments, dash indicates a loop or the absence of confident prediction.
Such a function is compatible with the high sequence conservation of this domain as well as its unusual expansion in certain genomes. Indeed, cyclic diguanylate stimulates cellulose synthesis in *Acetobacter xylinum* in response to the lack of oxygen (Ross et al., 1991). Similarly, multiple HD-GYP domains in *T. maritima*, *C. acetobutylicum*, *S. putrefaciens* and *V. cholerae* might be involved in signaling the availability of various electron acceptors, including iron and sulfur (Nealson and Saffarini, 1994; Vargas et al., 1998). Remarkably, sugar metabolism in *Thermotoga neapolitana* has been found to be subject to catabolite repression (Galperin et al., 1997; Vargas and Noll, 1996), although this organism is devoid of the PTS system (Galperin et al., 1996; Nelson et al., 1999) and contains negligible amounts of cAMP (Vargas and Noll, 1996). Thus preferential utilization of certain sugars (e.g., glucose) in *Thermotoga* should be regulated by an elaborate regulatory system different from those found in, for example, *E. coli* or *B. subtilis*. Whatever its exact function, the HD-GYP domain is likely to play a crucial role in this novel regulatory mechanism.

Acknowledgements

Analysis of unfinished genome sequences was made possible by generous submission to the public databases of preliminary sequence data by Genome Therapeutics (C. acetobutylicum), the Sanger Centre (B. pertussis), The Institute for Genome Research (S. putrefaciens, V. cholerae) and the Pseudomonas sequencing project (P. aeruginosa).

References


Vargas, M., and K. M. Noll. 1996. Catabolite repression in the hyperthermophilic bacterium *Thermotoga neapolitana* is independent of cAMP. Microbiology. 142: 139-144.


Galperin et al.