J. Mol. Microbiol. Biotechnol. (1999) 1(2): 303-305.

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 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

Species[a]	Domains		
	GGDEF	EAL	HD-GYP
Escherichia coli	19	18	-
Rickettsia prowazekii	1	1	-
Bacillus subtilis	4	2	-
Mycobacterium tuberculosis	1	2	-
Synechocystis sp.	23	13	2
Borellia burgdorferi	1	1	1
Treponema pallidum	1	-	3
Aquifex aeolicus	11	6	1
Thermotoga maritima	9	-	9
Clostridium acetobutylicum[b]	10	4	8

[a] Genomes of bacteria *Haemophilus influenzae, Helicobacter pylori, Chlamydia trachomatis, C. pneumoniae, Mycoplasma genitalium, M. pneumoniae*, and archaea *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum, Archaeoglobus fulgidus, Pyrococcus horikoshii*, and *Aeropyrum pernix* do not contain any of these domains.
[b] Preliminary data based on unfinished genome sequence (http://www.cric.com/sequence_center/bacterial_genomes)

Received October 4, 1999; accepted October 4, 1999.
*For correspondence. Email galperin@ncbi.nlm.nih.gov; Tel. 301-435-5910; Fax. 301-480-9241.

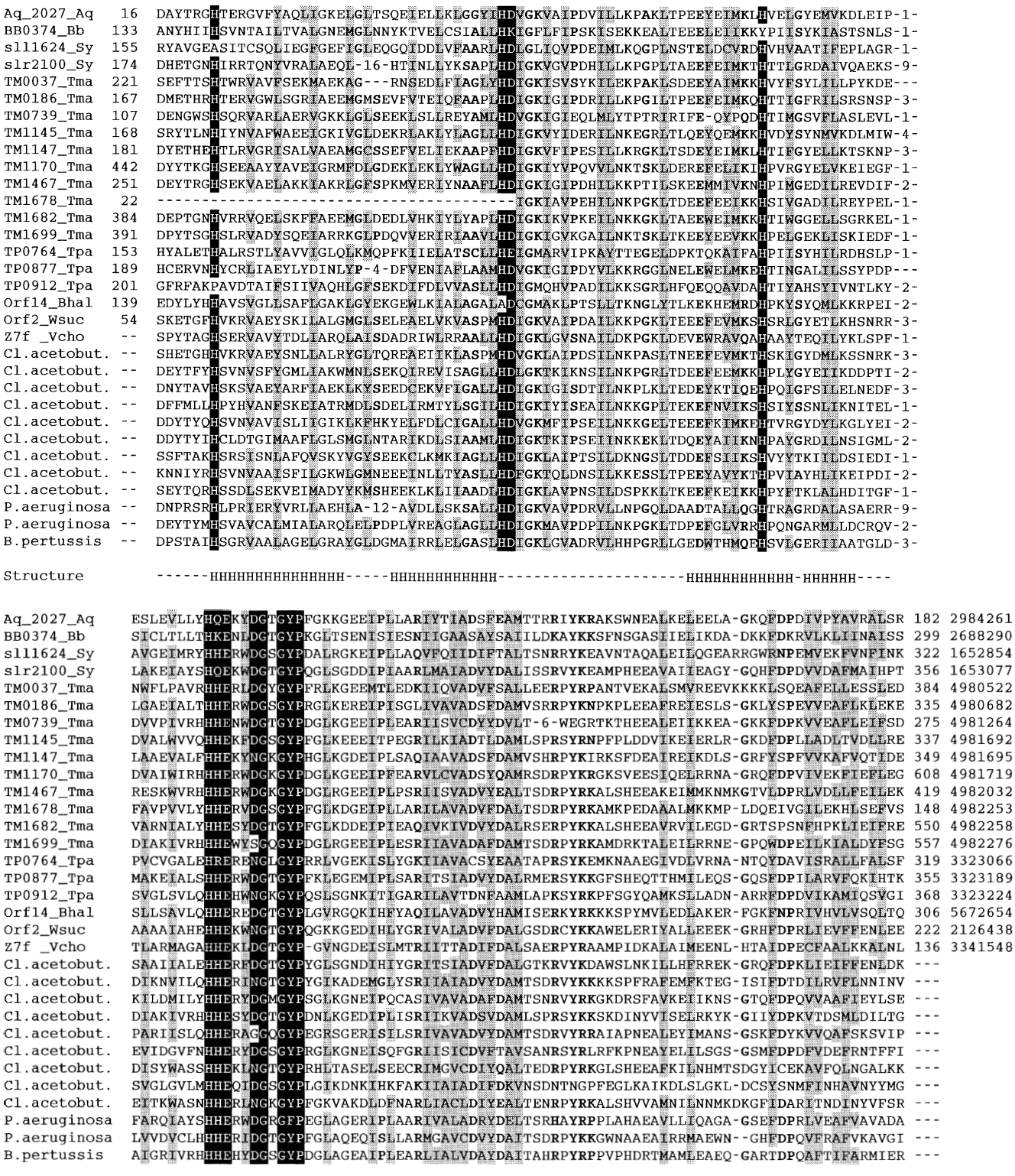

Figure 1. Multiple alignment of HD–GYP domains. The proteins are listed under their names in complete genomes (left column) and their unique gene identification (gi) numbers in the GenBank protein database (right column); the numbers indicate positions of the first and the last residues in each protein, 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.

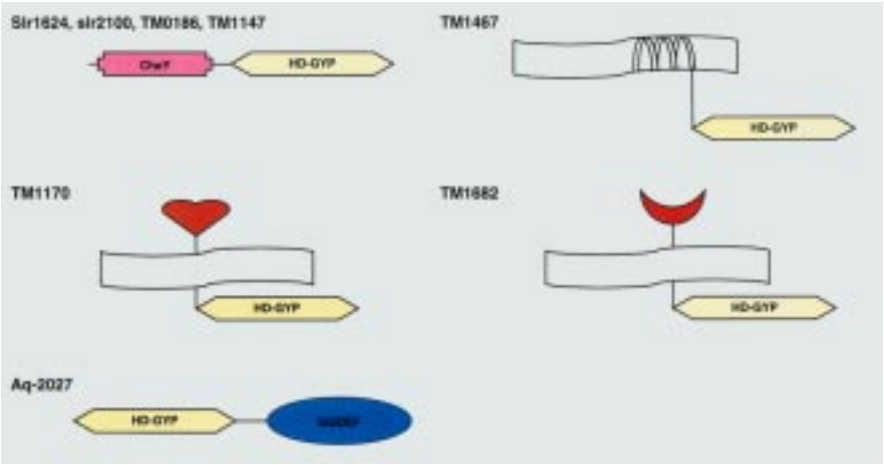

Figure 2. Association of HD-GYP domain with other signaling domains. CheY domain (Pao and Saier, 1995; Volz and Matsumura, 1991) and the periplasmic ligand-binding domain of TM1170 (Vyas *et al*., 1988) are well characterized; GGDEF domain (Hecht and Newton, 1995; Tal *et al*., 1998) and the MCP-like extracellular ligand-binding domain of TM1682 (Hanlon and Ordal, 1994) are less studied. The transmembrane portion of TM1467 does not show significant similarity to any characterized membrane protein.

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 signalling 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

Aravind, L., and E. V. Koonin. 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem. Sci. 23: 469-472.

Aravind, L., and Ponting, C.P. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol. Lett. 176: 111-116.

Galperin, M. Y., K. M. Noll, and A. H. Romano. 1996. The glucose transport system of the hyperthermophilic anaerobic bacterium *Thermotoga neapolitana*. Appl. Environ. Microbiol. 62: 2915-2918.

Galperin, M. Y., K. M. Noll, and A. H. Romano. 1997. Coregulation of beta-galactoside uptake and hydrolysis by the hyperthermophilic bacterium *Thermotoga neapolitana*. Appl. Environ. Microbiol. 63: 969-972.

Hanlon, D. W., and G. W. Ordal. 1994. Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. J. Biol. Chem. 269: 14038-14046.

Hecht, G. B., and A. Newton. 1995. Identification of a novel response regulator required for the swarmer- to-stalked-cell transition in *Caulobacter crescentus*. J. Bacteriol. 177: 6223-6229.

Koonin, E. V., R. L. Tatusov, and M. Y. Galperin. 1998. Beyond the complete genomes: from sequences to structure and function. Curr. Opin. Struct. Biol. 8: 355-363.

Kreis-Kleinschmidt, V., F. Fahrenholz, E. Kojro, and A. Kroger. 1995. Periplasmic sulphide dehydrogenase (Sud) from *Wolinella succinogenes*: isolation, nucleotide sequence of the sud gene and its expression in *Escherichia coli*. Eur. J. Biochem. 227: 137-142.

Nealson, K. H., and D. Saffarini. 1994. Iron and manganese in anaerobic respiration: environmental significance, physiology, and regulation. Annu. Rev. Microbiol. 48: 311-343.

Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, and C. M. Fraser. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature. 399: 323-329.

Pao, G. M., and M. H. Saier, Jr. 1995. Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. J. Mol. Evol. 40: 136-154.

Ponting, C. P., J. Schultz, F. Milpetz, and P. Bork. 1999. SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Res. 27: 229-232.

Ross, P., Mayer, R., and Benziman, M. 1991. Cellulose biosynthesis and function in bacteria. Microbiol. Rev. 55: 35-58.

Rost, B., and C. Sander. 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232: 584-599.

Takami, H., N. Masui, K. Nakasone, and K. Horikoshi. 1999. Replication origin region of the chromosome of alkaliphilic *Bacillus halodurans* C-125. Biosci. Biotechnol. Biochem. 63: 1134-1137.

Tal, R., H. C. Wong, R. Calhoon, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana, and M. Benziman. 1998. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. J. Bacteriol. 180: 4416-4425.

Tam, R., and M. H. Saier, Jr. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol. Rev. 57: 320-346.

Vargas, M., K. Kashefi, E. L. Blunt-Harris, and D. R. Lovley. 1998. Microbiological evidence for Fe(III) reduction on early Earth. Nature. 395: 65-67.

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

Volz, K., and P. Matsumura. 1991. Crystal structure of Escherichia coli CheY refined at 1.7-A resolution. J. Biol. Chem. 266: 15511-15519.

Vyas, N. K., M. N. Vyas, and F. A. Quiocho. 1988. Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. Science. 242: 1290-1295.