[Cell Cycle 7:4, 542-544; 15 February 2008]; ©2008 Landes Bioscience

Letter to the Editor

Uncharacterized DUF1574 leptospira proteins are SGNH hydrolases

Lukasz Knizewski,[1] Kamil Steczkiewicz,[1] Krzysztof Kuchta,[1] Lucjan Wyrwicz,[2] Dariusz Plewczynski,[1] Andrzej Kolinski,[3] Leszek Rychlewski[2] and Krzysztof Ginalski[1,*]

[1] Interdisciplinary Centre for Mathematical and Computational Modelling; University of Warsaw; Warsaw, Poland;[2] BioInfoBank Institute; Poznan, Poland;[3] Faculty of Chemistry; University of Warsaw; Warsaw, Poland

Abbreviations: DUF, domain of unknown function

Key words: DUF1574, SGNH hydrolases, GDSL serine esterases/lipases, distant homology detection, *Leptospira borgpetersenii, Leptospira interrogans,* leptospirosis

Leptospira borgpetersenii and *Leptospira interrogans* are nonsporulating bacteria responsible for leptospirosis, which is presumed to be the most widespread zoonosis in the world. This water-borne pathogen is usually transmitted to humans through the contact with contaminated soil or water, via infected animal tissue or rat bites. Consequently, farmers, miners, fishermen, sewer workers, participants of 'adventure tourism' and military personnel are at greater risk for infection. The usual portal of entry is through abrasions or cuts in the skin or by inhalation of contaminated water aerosol, however infection may also take place via intact skin after prolonged immersion in water. The incidence is considerably higher in tropical and developing countries what is associated with a longer survival of bacteria with increased possibility for exposure, however the cases of leptospirosis are also reported in developed countries of various geographical latitudes, e.g., USA, Canada and the Netherlands. The great majority of infections caused by these *Spirochetes* are either subclinical or very mild in severity (anticteric leptospirosis), although the fatal cases of much more severe icteric leptospirosis with renal failure or hemorrhagic complications were reported.[1]

DUF1574 family of bacterial proteins catalogued in PFAM[2] database as a domain of unknown function (accession number: PF07611) consists of completely uncharacterized proteins unique only to *Leptospira* species: *Leptospira borgpetersenii* and *Leptospira interrogans*. Consequently, these proteins may play an important role in general infection mechanism or create an important virulence factor and might be a potential drug target. Using a combination of advanced sequence profile searches, fold recognition and comparative modeling we predicted that DUF1574 proteins belong to the SGNH hydrolase superfamily.

Initial extensive PSI-BLAST[3] searches in the NCBI non-redundant protein database identified 12 uncharacterized DUF1574 family sequences of 3 serovars from two *Leptospira* species. These sequences can be grouped into four subfamilies of orthologs with less than 30% sequence identity between groups. DUF1574 proteins are characterized by the presence of the N-terminal signal peptide (as predicted by SignalP[4] and Phobius[5]) and three distinct blocks in predicted (with PSI-PRED[6]) secondary structure pattern: N-terminal ~130aa α/β region (αααβαβαβ), central ~110aa α-helical region (6 α-helices) and C-terminal ~100aa α/β region (αβααβα).

Further searches with DUF1574 sequences using standard sequence comparison tools such as CDD[7] or SMART[8] did not reveal any significant similarity to other proteins or domains of known function. However, top-of-the-line method for remote homology detection, Meta-BASIC,[9] suggested distant similarity of DUF1574 proteins to SGNH hydrolase superfamily members (pdb|1bwp,[10] Z-score 18.9; pdb|1deo,[11] Z-score 13.7) with confident Z-scores above 12 (corresponding to less than 0.05 probability of being incorrect). On the other hand, the consensus of fold recognition, 3D-Jury,[12] mapped DUF1574 family sequence (gi|116332481) both to chelatase-like metal receptors (whole query sequence, 3D-Jury score 59) and SGNH hydrolases (either N- or C-terminal α/β regions of query sequence, 3D-Jury score 46) with border-line confidence scores.[13] Although higher scores were assigned to the metal receptors superfamily, in contrast to obtained SGNH hydrolase predictions, critical residues were not conserved in the proposed alignments, and mappings of secondary structure elements were highly inconsistent. The excision of central α-helical region (137–250 in gi|116332481) and resubmission of remaining α/β segments to 3D-Jury server resulted in a complete mapping of query sequence onto the SGNH hydrolase structures with notable increase of 3D-Jury reliability scores (pdb|1bwp, 3D-Jury score 103; pdb|1fxw,[14] 102; pdb|1vjg, 74), while no hits to chelatase-like metal receptors were observed. Presence of critical active site residues and good mapping of secondary structure elements and hydrophobic profiles strongly supported the prediction that DUF1574 proteins belong to SGNH hydrolase superfamily. Additional searches conducted for the central α-helical region using Meta-BASIC and 3D-Jury did not provided any significant hits to known protein families or structures.

SGNH hydrolases (also called GDSL serine esterases/lipases[15]) form a large superfamily of hydrolytic enzymes with multifunctional lyase activities (e.g., thioesterase, protease, arylesterase,

*Correspondence to: Krzysztof Ginalski; Interdisciplinary Center for Mathematical and Computational Modelling; Warsaw University; Pawinskiego 5a; 02-106 Warsaw, Poland; Tel.: +48.22.874.9100; Fax: +48.22.874.9115; Email: kginal@icm.edu.pl

Submitted: 11/20/07; Accepted: 11/25/07

Previously published online as a *Cell Cycle* E-publication: http://www.landesbioscience.com/journals/cc/article/5386

Figure 1. Multiple sequence alignment for representative DUF1574 family members and selected SGNH hydrolase structures. Sequences are labeled according to the NCBI gene identification (gi) number or PDB code and an abbreviation of the species name and serovar: Lbh, *Leptospira borgpetersenii* serovar Hardjo-bovis; Lic, *Leptospira interrogans* serovar Copenhageni; Lil, *Leptospira interrogans* serovar Lai. The first and last residue numbers are indicated before and after each sequence with total sequence length of DUF1574 proteins following in square bracket. The numbers of excluded residues are specified in parentheses. Residue conservation in DUF1574 family is denoted with following scheme: uncharged, highlighted in yellow; charged or polar highlighted in grey; small, letters in red. Critical active site residues (catalytic triad) are highlighted in black, while remaining less conserved glycine and asparagine of SGNH signature motif are highlighted in blue. SGNH hydrolase superfamily characteristic sequence blocks I, II, III, V are labeled. Multiple sequence alignment for inserted α-helical potential substrate recognition domain is shown in the red box at the top. Locations of predicted (gi|116332481) and observed (pdb|1bwp) secondary structure elements (E, β-strand; H, α-helix) are marked above the corresponding sequences. Alignment for DUF1574 sequences was prepared using PCMA[25] program followed by some manual adjustments. Sequence-to-structure mapping between DUF1574 family and SGNH hydrolase structures was built manually using 3D assessment procedure,[26] taking into account predicted secondary structure and hydrophobic profile of the family.

lysophospholipase) and broad regio- and substrate specificities. Structural classification of proteins SCOP[16] currently defines six different families in SGNH hydrolase superfamily: (1) esterase that hydrolyses ester bonds in lipid cover of the plant tubers (associated with a scrab disease),[17] (2) haemagglutinin esterase responsible for binding and lysing membrane receptors,[18] (3) acetylhydrolase regulating platelet-activating factors through hydrolysis of phospholipids (signal termination),[19] (4) rhamnogalacturonan acetylesterase critical for degradation of polysaccharides,[11] (5) thioesterase I, TAP[20] with lisophospholipase activity and (6) a putative lipase alr1529.

SGNH hydrolases have flavodoxin-like fold structure that is characterized by the presence of a central five-stranded parallel β-sheet (with β2β1β3β4β5 strand order) flanked by α-helices on both sides. The active site is situated at the bottom of open cleft that is easily accessible to large substrates. This site is relatively flexible and susceptible to conformational changes, thus exhibiting highly versatile functional roles such as protease, thioesterase, acetylhydrolase and acetylesterase. Conserved residues from characteristic four sequence motifs (blocks I, II, III, V; numbering nomenclature follows ref. 21) are critical for the formation of the active site and catalysis in SGHN hydrolase superfamily. Specifically, the active site is defined by the catalytic triad: serine (block I, motif GXS), aspartic acid and histidine (block V, motif DXXH), which are usually supported by glycine (block II) and asparagine (block III) residues (e.g., S47, G71, N104, D192, H195 in pdb|1bwp[10]). In contrast to canonical α/β hydrolases (e.g., common lipases), members of SGNH

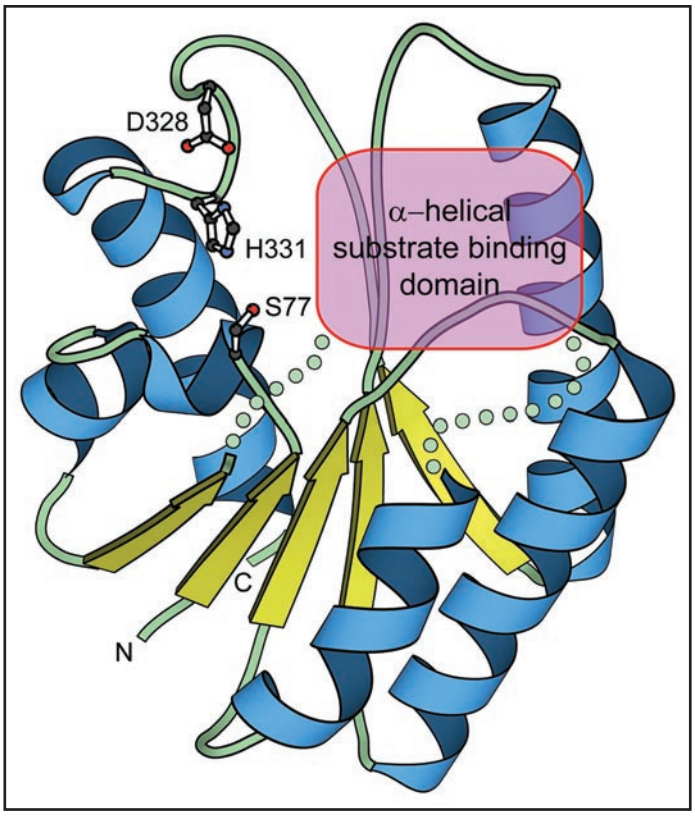

Figure 2. 3D model for one of the DUF1574 family members (gi|116332481) generated with MODELLER[27] using distantly related structures of SGNH hydrolases (pdb|1bwp, pdb|1vjg and pdb|1u8u) as templates. Active site catalytic triad (S77, D328, H333) is shown. Inserted α-helical substrate recognition domain was not modeled, however its location was denoted with violet box.

superfamily do not have "nucleophile elbow" (motif GXSXG). Invariant serine (block I) serves as the nucleophile and a proton donor to the oxyanion hole, histidine (block V) acts as a base to make serine more nucleophilic by deprotonating its hydroxyl group, while aspartic acid (block V) promotes formation of $HisH^+$. Less conserved residues (e.g., absent in pdb|1u8u[20]) in blocks II (Gly) and III (Asn) serve as two additional proton donors to the oxyanion hole. During the general catalytic mechanism, a non-covalent Michaelis complex with substrate is formed, while serine hydroxyl group attacks the substrate carbonyl groups. This is followed by formation of acyl-enzyme intermediate and hydrophilic attack of positioned water molecule on ester carbonyl group resulting in restoration of free enzyme. In order to support substrate binding, the cleft is also equipped with positively charged residues (e.g., R141 in pdb|1bwp) to interact with the substrate carboxylate groups.

As shown in Figure 1, the catalytic triad (S77, D328, H333 in gi|116332481) is fully conserved in DUF1574 proteins. Additionally, the predicted secondary structure pattern is consistent with that of SGNH hydrolase fold core (βαβαβαβαβα). Less conserved proton donors in SGNH hydrolase superfamily (glycine in block II and asparagine in block III) are not present in all DUF1574 family members. To analyze details of the active site architecture and the surrounding cleft, a homology model of DUF1574 SGNH hydrolase domain was built (Fig. 2). Inspection of the active site cleft revealed that positively charged residues, which are observed in other SGNH hydrolases, are missing in DUF1574 proteins in this region (these residues migrated to α-helical domain, see below). Although α-helical region was not modeled, its position in relation to catalytic SGNH hydrolase domain suggests that it may constitute substrate binding domain (Fig. 2). The predicted α-helical domain possesses several positively charged residues, conserved within the subfamilies (e.g., R158, K183, R192, K194, K205 in gi|116332481), that might interact with substrate carboxylate groups.

The genomic context analysis using STRING[22] suggests functional associations for DUF1574 proteins with alginate o-acetyltransferase (confident neighborhood score of 0.8). Genes encoding DUF1574 family members are located adjacent to the alginate o-acetyltransferase encoding gene, however the functional link between these two genes cannot be confirmed (e.g., acetyltransferase/acetylase modification system). Extracellular polysaccharide alginate biosynthesis in bacteria has been shown to lead to the establishment of biofilm,[23] thus genes of alginate synthesis pathway have potential important role for pathogenesis and virulence during infection. [24] Because alginate is acetylated in the periplasmic space right before the product is transported thought the membrane, DUF1574 hydrolases as potentially secreted proteins might be involved in processing of the surface carbohydrate layer of cell. The observed sequence divergence within the analyzed family may suggest an existence of variance in substrate specificity among these proteins. The future experimental verification of probable role of DUF1574 proteins in biology of *Leptospira* and the pathogenesis of leptospirosis is an emerging task.

Acknowledgements

This work was supported by NIH (1R01GM081680-01), European Commission (LSHG-CT-2004-503567, COOP-CT-2004-508399), EMBO Installation and FNP (FOCUS) grants.

References

1. Levett PN. Clin Microbiol Rev 2001; 14:296-326.
2. Finn RD, et al. Nucleic Acids Res 2006; 34:D247-51.
3. Altschul SF, et al. Nucleic Acids Res 1997; 25:3389-402.
4. Bendtsen JD, et al. J Mol Biol 2004; 340:783-95.
5. Kall L, et al. J Mol Biol 2004; 338:1027-36.
6. Jones DT. J Mol Biol 1999; 292:195-202.
7. Marchler Bauer A, et al. Nucleic Acids Res 2005; 33:D192-6.
8. Letunic I, et al. Nucleic Acids Res 2006; 34:D257-60.
9. Ginalski K, et al. Nucleic Acids Res 2004; 32:W576-81.
10. Ho YS, et al. Protein Eng 1999; 12:693-700.
11. Molgaard A, et al. Structure 2000; 8:373-83.
12. Ginalski K, et al. Bioinformatics 2003; 19:1015-8.
13. Ginalski K, et al. Nucleic Acids Res 2003; 31:3291-2.
14. Sheffield PJ, et al. Protein Eng 2001; 14:513-9.
15. Akoh CC, et al. Prog Lipid Res 2004; 43:534-52.
16. Andreeva A, et al. Nucleic Acids Res 2004; 32:D226-9.
17. Wei Y, et al. Nat Struct Biol 1995; 2:218-23.
18. Rosenthal PB, et al. Nature 1998; 396:92-6.
19. Derewenda ZS, et al. Cell Mol Life Sci 1998; 54:446-55.
20. Lo YC, et al. Biochemistry 2005; 44:1971-9.
21. Upton C, et al. Trends Biochem Sci 1995; 20:178-9.
22. von Mering C, et al. Nucleic Acids Res 2005; 33:D433-7.
23. Gacesa P. Microbiology 1998; 144:1133-43.
24. Nascimento AL, et al. J Bacteriol 2004; 186:2164-72.
25. Pei J, et al. Bioinformatics 2003; 19:427-8.
26. Ginalski K, et al. Proteins 2003; 53:410-7.
27. Sali A, et al. J Mol Biol 1993; 234:779-815.