

# Analysis of the Evolutionary Relationships of the PDZ Domains

Masaki Yamamura

masaki@kuicr.kyoto-u.ac.jp

Tatsuya Akutsu

takutsu@kuicr.kyoto-u.ac.jp

Susumu Goto

goto@kuicr.kyoto-u.ac.jp

Institute for Chemical Research (ICR), Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

**Keywords:** PDZ domain, sequence analysis

## 1 Introduction

The PDZ domains are well-known globular fold module of many scaffolding proteins. The domains are involved in protein-protein interaction and play a central role in organizing diverse cell signaling assemblies [2]. Recently the motifs of their binding sites have been researched in biochemical and structural studies [1, 3], but it was pointed out the difficulty to classify the PDZ domains on the basis of the chemical characteristics of binding sites residues [4]. Another one of the specific known features is that the domain appears several times in a mature protein. The result of sequence analysis of the PDZ domain strongly suggested that the PDZ domains were derived from the domain duplication rather than the vertical conservation or the gene duplication.

## 2 Materials and Methods

The sequence data of the PDZ domain were collected from SMART database [5] and the total number was 1951. We calculated all to all pairwise alignments scores for those 1951 domains, using the Smith-Waterman algorithm. The complete linkage clustering was applied with the score 100 as the threshold.

## 3 Results

The number of best-best relation pairs was 248. The number of clusters was 150 and in a cluster the frequency of appearance of PDZ domains included in the same proteins was 929/1951. Those relatively high score gave the idea that the PDZ domains were derived from the domain duplication rather than the vertical conservation or the gene duplication.

## 4 Discussion

Many proteins which include the PDZ domains also include the GuKc domains. The GuKc domains are also well known and involved in construction of protein complexes. It may be interesting to investigate other proteins with single GuKc domain and multiple PDZ domains. Applying evolutionary phylogenetic analysis to those data, we could discuss the hypothesis more clearly.

```

Complete linkage clustering:
  +- >smart|PDZ-sptrembl|Q9KUF5|Q9KUF5/268-350 Protease D0. Vibrio cholerae
  +-| 312.0
  | | +- >smart|PDZ-sptrembl|Q9KRY7|Q9KRY7/247-324 Tail-specific protease. Vibrio cholerae
  +-| 295.0
  | | +- >smart|PDZ-sptrembl|Q9KUF5|Q9KUF5/375-448 Protease D0. Vibrio cholerae
  +-| 157.0
  | | +- >smart|PDZ-sptrembl|Q9L1S8|Q9L1S8/92-169 Putative carboxy-terminal processing protease. Streptomyces coelicolor
  | | +-| 525.0
  | | | +- >smart|PDZ-sptrembl|Q9KUF6|Q9KUF6/255-337 Protease DegS. Vibrio cholerae
  +-| 505.0
  | | +- >smart|PDZ-sptrembl|Q9L5V8|Q9L5V8/786-857 MdsD protein precursor. Prevotella
  +-| 150.0
  | | +- >smart|PDZ-sptrembl|Q9LK70|Q9LK70/272-358 Similarity to serine protease. Arabidopsis thaliana
  +-| 396.0
  | | +- >smart|PDZ-sptrembl|Q9LBK0|Q9LBK0/396-466 Heat shock protein HtrA. Shigella sonnei
  +-| 247.0
  | | +- >smart|PDZ-sptrembl|Q9LBK0|Q9LBK0/289-371 Heat shock protein HtrA. Shigella sonnei
  +-| 128.0
  | | | +- >smart|PDZ-sptrembl|Q9N531|Q9N531/1293-1366 Hypothetical 152.4 kDa protein. Caenorhabditis elegans
  | | | +-| 541.0
  | | | | +- >smart|PDZ-sptrembl|Q9N531|Q9N531/1194-1264 Hypothetical 152.4 kDa protein. Caenorhabditis elegans
  | | | +-| 419.0
  | | | | +- >smart|PDZ-sptrembl|Q9NAN2|Q9NAN2/166-249 T26E3.3 protein (PAR-6). Caenorhabditis elegans
  +-| 336.0
  | | | | +- >smart|PDZ-sptrembl|Q9N530|Q9N530/1326-1399 Hypothetical 155.9 kDa protein. Caenorhabditis elegans
  | | | +-| 461.0
  | | | | +- >smart|PDZ-sptrembl|Q9N530|Q9N530/1227-1297 Hypothetical 155.9 kDa protein. Caenorhabditis elegans
  +-| 269.0
  | | +- >smart|PDZ-sptrembl|Q9LK71|Q9LK71/301-394 Cb|AAF23847.1. Arabidopsis thaliana
  +-| 107.0
  | | +- >smart|PDZ-sptrembl|Q9NF04|Q9NF04/141-222 Y105E8C.b protein. Caenorhabditis elegans
  +-| 513.0
  | | +- >smart|PDZ-sptrembl|Q9NF04|Q9NF04/54-131 Y105E8C.b protein. Caenorhabditis elegans
  +-| 105.0
  | | | +- >smart|PDZ-sptrembl|Q9NBV3|Q9NBV3/373-448 INAD. Drosophila melanogaster
  | | | +-| 449.0
  | | | | +- >smart|PDZ-sptrembl|Q9NBV3|Q9NBV3/258-333 INAD. Drosophila melanogaster
  | | | +-| 434.0
  | | | | +- >smart|PDZ-sptrembl|Q9NBV3|Q9NBV3/497-576 INAD. Drosophila melanogaster
  +-| 242.0
  | | | | +- >smart|PDZ-sptrembl|Q9NBV3|Q9NBV3/26-106 INAD. Drosophila melanogaster
  +-| 171.0
  | | | | +- >smart|PDZ-sptrembl|Q9NDP4|Q9NDP4/563-642 Tyrosine phosphatase. Ciona intestinalis
  +-| 248.0
  | | | | +- >smart|PDZ-sptrembl|Q9NBV3|Q9NBV3/593-664 INAD. Drosophila melanogaster
  +-| 156.0
  | | | | | +- >smart|PDZ-sptrembl|Q9NB04|Q9NB04/736-815 Discs LOST. Drosophila melanogaster
  | | | | | +-| 319.0
  | | | | | | +- >smart|PDZ-sptrembl|Q9NB04|Q9NB04/568-648 Discs LOST. Drosophila melanogaster
  | | | | | +-| 194.0
  | | | | | | +- >smart|PDZ-sptrembl|Q9NB04|Q9NB04/326-403 Discs LOST. Drosophila melanogaster
  | | | | | +-| 344.0
  | | | | | | +- >smart|PDZ-sptrembl|Q9NB04|Q9NB04/157-229 Discs LOST. Drosophila melanogaster
  +-| 125.0
  | | +- >smart|PDZ-sptrembl|Q9NF04|Q9NF04/407-478 Y105E8C.b protein. Caenorhabditis elegans
  +-| 81.0
  | | +- >smart|PDZ-sptrembl|Q9KNC0|Q9KNC0/96-169 Protease-related protein. Vibrio cholerae
  +-| 247.0
  | | +- >smart|PDZ-sptrembl|Q9KJN6|Q9KJN6/415-486 Putative serine protease D0-like. Myxococcus xanthus
  +-| 247.0

```

Figure 1: A part of dendrogram of sequence analysis of the PDZ domain.

The data was listed in SMART database (<http://smart.embl-heidelberg.de/>). The total number of entries was 1951. Shown numbers of separation points mean Smith-Waterman score.

## References

- [1] Bezprozvanny, I. and Maximov, A., Classification of PDZ domains, *FEBS Letters*, 509:457–462, 2001.
- [2] Harris, B.Z. and Lim, W.A., Mechanism and role of PDZ domains in signaling complex assembly, *Journal of Cell Science*, 114(18):3219–3231, 2001.
- [3] Reina, J., Lacroix, E., Hobson, S.D., Fernandez-Ballester, G., Rybin, V., Schwab, M.S., Serrano, L., and Gonzales, C., Computer-aided design of a PDZ domain to recognize new target sequences, *Nature Structural Biology*, 9(8):621–627, 2002.
- [4] Vaccaro, P. and Dente, L., PDZ domains: Troubles in classification, *FEBS Letters*, 512:345–346, 2002.
- [5] <http://smart.embl-heidelberg.de/>