Bioinformatics and Biosystems 2006, Vol. 1, No. 2, pp. 123-127

Local structural alignment and classification of TIM barrel domains

Changwon Keum[1], Jihong Kim[2] and Jongsun Jung[2]

[1]Department of BioSystems, KAIST, Daejeon, Korea
[2]NGRI, KNIH, Seoul, Korea

Abstract

TIM barrel domain is widely studied since it is one of most common structure and mediates diverse function maintaining overall structure. TIM barrel domain's function is determined by local structural environment at the C-terminal end of barrel structure. We classified TIM barrel domains by local structural alignment tool, LSHEBA, to understand characteristics of TIM barrel domain's functionalvariation. TIM barrel domains classified as the same cluster share common structure, function and ligands. Over 80% of TIM barrels in clusters share exactly the same catalytic function. Comparing clustering result with that of SCOP, we found that it's important to know local structural environment of TIM barrel domains rather than overallstructure to understand specific structural detail of TIM barrel function. Non TIM barrel domains were associated to make different domain combination to form a different function. The relationship between domain combination, we suggested expected evolutional history. We finally analyzed the characteristics of amino acids around ligand interface.

Introduction

TIM barrel fold, composed of 8 continuous (β/α) motifs, has been widely studied since it is one of most common folds as nearly 10% of enzyme known so far have TIM barrel domain(Nozomi Nagano 2001). However, their sequence similarity is not high enough to detect their evolutional relationship and structural similarity. TIM barrel fold is also one of most functionally versatile folds(Wierenga 2001). The active sites of TIM barrel always come at the end of C-terminal barrel composed of inside β-sheets. This suggests TIM barrel have evolved from ancestral TIM barrel with divergent evolution(Rayment 2004). The functional diversity of TIM barrel comes from the fact that the TIM barrel function can be changed by slight variation around the active site, loops connecting C-terminal inside β-sheets and outside α-helices. Function variation of TIM barrel is dividedinto two categories, substrate specificity and catalytic activity. Variation at inside barrel changes substrate specificity of the TIM barrels. On the other hand, variation around active site changes catalytic activity (Altamirano 2000). These TIM barrel's two ways of function creation made it an efficient fold to manipulate new enzyme activity.

One fold-many function character of TIM barrel makesit difficult to assign function based on structure(Nozomi Nagano 2002).Just recognizing overall structure of TIM barrel is not enough to understand relationship between structure and function of TIM barrel domains. Many protein structural databases such as SCOP (Alexey G. Murzin 1995) and CATH (CA Orengo and Thornton 1997) classified TIM barrels according to their structures. However, TIM barrel classification by these databases does not detect specific difference among TIM barrel domains with different functions since TIM barrel domains in each family from the database have diverse function, not unique function.

Here we tried to understand the relationship between structure and function of TIM barrel domain by local structural alignment of TIM barrel domains using LSHEBA (Lee 2000), a local structure alignment tool. We classified the TIM barrel structures using the result of structural alignment between TIM barrel domains. We tried to find structural and functional characteristics of each cluster and the relationship between ligand and function of TIM barrel domains. We also found the evolutional history between each clusters by considering domain combination.

Methods

Collection of TIM barrel imbedded enzymes was performed using 1.65 version of SCOP and manual inspection after the

Corresponding Author: Changwon Keum (Email: keum 4988@empal.com)

local structural alignment of TIM barrel structure from PDB files by LSHEBA. A total of 2, 203 TIM barrel imbedded enzymes were collected where a TIM barrel domain is defined that there must be 8 beta strands forming a barrel and at least 6 or more helices surrounded (Douglas H. Juers 1999).

All vs. all TIM barrel domain comparisons were performed and they were clustered by an exhaustive clustering algorithm, CLIQUE in which a cluster is defined as the complete collection of all TIM barrel domains, all of which are related to all others by a chain of neighbor relations; two TIM barrel domains from two different clusters are not related to each other by any chain of neighbor relations. Two TIM barrel domains are defined as a neighbor if they meet both sequence similarity score (mean HSL >= 250) and c-alpha atom match score(Nmat >= 150). Structures over 95% sequence similarity were used for further analysis after clustering.

Conservation ratio of structurally aligned TIM barrels were calculated as : $\frac{1}{n} N_i \sum_{i=1}^{20} \frac{N_i}{n}$, i = 20 amino acids.

EC numbers, enzyme classification number, are assigned to each PDB file (Bairoch 2000). We assigned EC number of TIM barrel protein to TIM barrel domain. Some TIM barrel proteins lack EC number and some has incomplete EC number. We basically used complete EC number to analyze TIM barrel domain function.

Ligand information was from HETATM header of PDB files of TIM barrels. Ligand interface was defined a set of amino acids within 4.5 angstrom from ligand.

Domain definition of TIM barrel and the other domain associated with TIM barrel domain used here was come from 1.65 version of SCOP.

Result

2,203 TIM barrel chains among total 2,480TIM barrel chains determined by LSHEBA, recognized as TIM barrel domains in SCOP, were used for classification. Of them, we focused on 1438 chains which were classified into 45 distinct clusters, each with over 10 member chains. EC numbers were not assigned to 116 TIM barrels among 1438TIM barrel domains in 45 clusters (Table 1). We analyzed TIM barrel function only EC number annotated TIM barrel domains.. SCOP classified the 1438 TIM barrel domains in LSHEBA cluster into 31 distinct structural families. SCOP cluster we mentioned in this paper means SCOP family of those 1438 TIM barrel domains, not entire SCOP family.

Conservation of function in the clusters

EC number is a way of function annotation scheme composed of 4 levels, e.g., 4.2.4.32 (Bairoch 2000). Catalytic function is described by EC number from top level to 3rd level. 4thlevel indicates substrate specificity. It can be potential problem that enzymes share the same EC number when they catalyzethe same reaction even if the reactions were mediated by different mechanism (Doig 2005). We ignored this problem since we assumed that TIM barrel domains shared conserved structure to perform the same function in each cluster where TIM barrel domains share local structure.

118 and 42 different EC numbers at 4th level and 3rd level respectively assigned to total 2,203 TIM barrel chains. For TIM barrels in 45 clusters, 56 and 24 different EC numbers were assigned at 4th and 3rd level respectively. 39 clusters among total 45 clusters have uniqueEC number at 3rd level and 36 clusters have unique EC number at 4thlevel while 25 and 20 clusters among 31 SCOP clusters have unique EC number at 3rd and 4th level, respectively (Table 1). LSHEBA classified TIM barrel domains more strictly than SCOP so that it is possible to classify them into functionally related clusters. Incomplete EC numbers at 3rd and 4th were EC numbers assigned at top and 2nd level, but not at 3rd or 4th. However, TIM barrel domains with incomplete EC numbers were not classified as distinct cluster.

Table 1. EC number conservation.

	SCOP	LSHEBA
Complete EC 3rd	25/31(80.6%)	39/45(86.6%)
Incomplete EC 3rd	22/31(71%)	36/45(80%)
Complete EC 4th	20/31(64.5%)	36/45(80%)
Incomplete EC 4th	18/31(58%)	34/45(75.5%)

Functional diversity

56 different EC numbers from total 118 EC numbers were classified into structurally related clusters by LSHEBA (Figure 1). EC number diversity is increasing from 3rd level to 4th level. 4th level indicates substrate specificity of enzymes.

TIM barrel fold can change its substrate specificity retaining its catalytic activity by changing a few amino acids at C-terminus inside barrel. Over 80% of TIM barrels with EC class 4 and 5 were remained in LSHEBA clusters, however their EC number diversity were near 50%. Remaining 20% TIM barrels

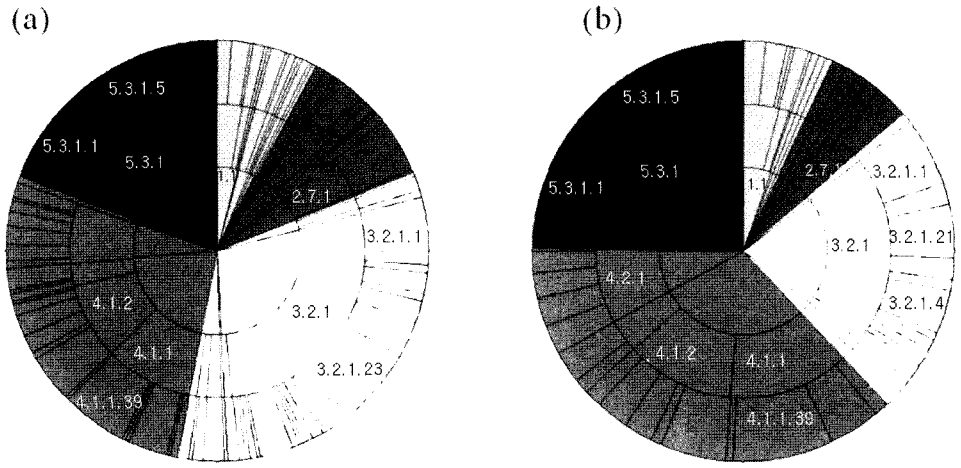

Figure 1. Concentric pie chart (EC wheel) shows the distribution of TIM barrel functions. Different EC levels, 2nd ,3rd and 4th, are represented by circles from inner to outer (a) for all 2,203 TIM barrels and (b) for 1,438 TIM barrels in clusters.

made 50% functional diversity. This showsthat TIM barrel has evolved to have different function from structural ancestor TIM barrel. Catalytic activity has been maintained to act on various substrates. This may be the reason why TIM barrel fold is most functionally diverse functional framework. EC class 3 has total 43 different EC numbers and 692 TIM barrel domains in total TIM barrel chains, however only 15 EC numbers and 323 Tim barrels were classified into distinct clusters. TIM barrels with catalytic function lyase were most functionally diverse, at the same time, most structurally versatile. This suggests that TIM barrels achieved their function by divergent evolution.

Domain combination

Majority of proteins consist of at least two domains (Christine Vogel 2004). Domain duplication and shuffling are the two major sources to generate a new domain combination. Domain recombination is also a major source of function differentiation. Each domain in a protein could have its own role when it functions. Combined domain could form different active site or recognize different substrate (Douglas H. Juers).

57 TIMbarrel domains in total 184 TIM barrel proteins with 95% sequence similarity threshold from total 1438 are multi domain protein. There are total 12 different non TIM barrel domains in multi domain TIM barrels. But there's only one duplicated TIM barrel. It has duplicated TIM barrel domain, c.1.2.4. Duplicated TIM barrel has different EC number, 5.3.1.24, while single TIM barrel domain protein, c.1.2.4, has EC number 4.1.1.48. In case of TIM domain c.1.8.5, three different domains were associated to form 3 different domain combinations. These proteins with different domain combination maintain the same function, 3.2.1.14. Protein composed of single TIM barrel c.1.8.5 also has the same catalytic function, 3.2.1, but different substrate specificity, 3.2.1.96. The relationship of combination between these 4 different domain from TIM barrel's point could be visualized(Figure 2). At first, a single TIM barrel domain function 3.2.1.14. Later, when it needed

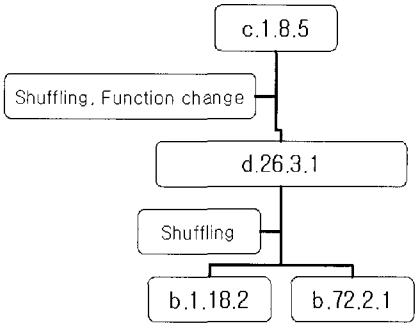

Figure 2. Suggested domain combination from c.1.8.5

to function on different substrate, another domain,d.26.3.1, was recombined to acquire modified function. Continuous shuffling with b.1.18.2 and b.72.2.1 had occurred without function change. As shown the example, function change might or might not occurred correspond with domain recombination.

Characteristics of TIM barrel at ligand interface

Ligand interface is especially important region of the structure since it corresponds with active site of the protein(Stephen J Campbell 2003). Ligand interface of a protein might have been more conserved than other region for functional conservation.

Amino acid composition of TIM barrels around ligand interface is not much different from that around non interface (Figure 3). On the other hand, there's significant difference between them where aligned region with conservation ratio over 0.6 (Figure 4).

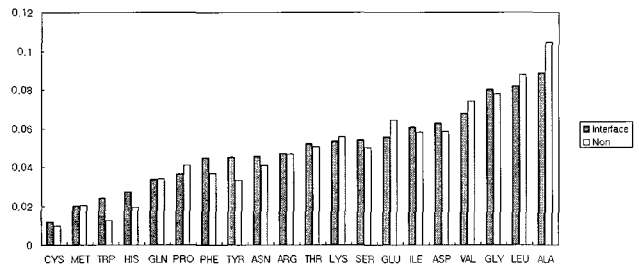

Figure 3. Relative ratio of amino acids around ligand interface

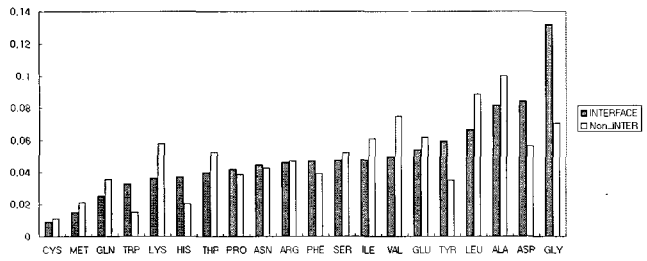

Figure 4. Relative ratio of amino acids around ligand interface with conservation ratio over 0.6

Negatively charged amino acids, GLU and ASP, were more frequently observed than positively charged ones, ARG and LYS. It represents that there are more positively charged ligands than negatively charged ones since to recognize charged ligand, opposite charged amino acids are needed. Most of GLY located in loop connecting helix and sheet (Figure 5). This showed that conservation of GLY residues located in loop were important to conserve function among structural relatives. Charged amino acids were located in helix or loop rather than sheet. As in the other structures , charged amino acids are located around interface region in TIM barrel, so that charged amino acids in TIM barrel seldom occurred in sheets which form inside barrel structure which is not located around interface in TIM barrel. However, conservation ratio of ligand interface does not distinguishable according to each structure (Figure 6). It seemed that amino acids around ligand interface were equally under conserving pressure.

Discussion

We have analyzed functional diversity of TIM barrel domains in terms of their structure. Structural environment of the same cluster was well conserved among TIM barrel domains even when they have different domain combination. As domain combined to have a new domain combination, function diverges. The change can be of complete catalytic function or of substrate specificity conserving catalytic function. This suggested that TIM barrel domains had evolved to have required function by combining with other domains. Ligand interface is another source of function variation of TIM barrels. Variation of amino acids around ligand interface corresponds to active site of TIM barrel, C-terminus of inside barrel. Especially, frequently occurring amino acids at ligand interface such as Gly and negatively charged amino acids could be key to analyze or predict the function of TIM barrel. In case of Gly, since it extremely located in loop region where catalytic residues of TIM barrel form, aligned ratio of Gly at ligand interface could be important indicator of function conservation. Negatively charged amino acids are also functionally important since they directly interact with ligand and substrate for functioning. When these amino acids are structurally differ from already known TIM barrels, it might leads to modification of the function.

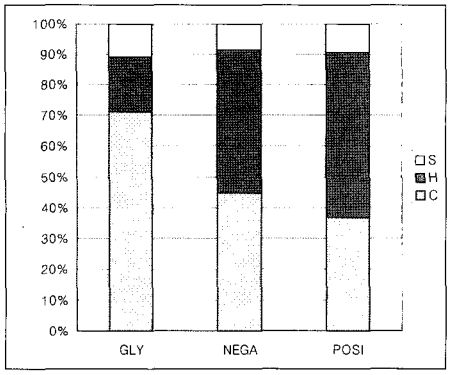

Figure 5. Relative frequency of Gly, positively charged and negatively charged amino acids.

Classification of TIM barrel domains using LSHEBA structue alignment failed to classified all TIM barrel domains into distinct function, however, the result of classification showed local structural similarity of TIM barrel is a basic functional unit of TIM barrel domains. In case of nabonin, LSHEBA correctly distinguished it as distinct functional cluster from others. It also suggests function prediction for folds with many function such as TIM barrel fold can be possible by detecting local structure at active site.

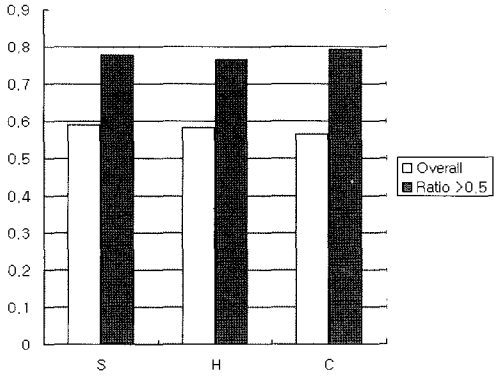

Figure 6. Average conservation ratio according to secondary structure. S represents sheet, H represents helix and C represents coil(loop).

References

[1] Alexey G. Murzin, S. E. B., Tim Hubbard and Cyrus Chothia (1995). "SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures." Journal of Molecular Biology 247: 536-540.

[2] Altamirano, M. M. (2000). "Directed evolution of new catalytic activity using the a/b-barrel scaffold." Nature 403: 617-622.

[3] Bairoch, A. (2000). "The ENZYME database in 2000." Nucleic Acids Research 28: 304-305.

[4] CA Orengo, A. M., S Jones, DT Jones, MB Swindells and and J. Thornton (1997). "CATH −a hierarchic classification of protein domain structures." Current Biology 5: 1093-1108.

[5] Christine Vogel, M. B., Nicola D Kerrison, Cyrus Chothia and Sarah A Teichmann (2004). "Structure, function and evolution of multidomain proteins." Current Opinion in Structural Biology 14: 208-216.

[6] Cyrus Chothia, J. G., Christine Vogel, Sarah A. Teichmann (2003). "Evolution of the protein repertoire." Science 300: 1701-1703.

[7] Doig, P. D. D. a. A. J. (2005). "Predicting Enzyme Class From Protein Structure

[8] Without Alignments." Journal of Molecular Biology 345(187-199).

[9] Douglas H. Juers , R. E. H. a. B. W. M. (1999). "Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between [beta]-galactosidase and other glycohydrolases." Protein science 8: 122-136.

[10] Lee, J. J. a. B. (2000). "Protein structure alignment using environmental profiles " Protein engineering design & selection 13: 535-543.

[11] Nozomi Nagano, J. M. T. (2001). "The (ba)8 glycosidases: sequence and structure analysis suggest distant evolutionary relationships." Protein Eng. 14: 845-855.

[12] Nozomi Nagano, J. M. T. (2002). "One fold with many functions : The evolutionary relationships between TIM barrel families based on their sequences, structures and functions." JMB 321: 741-765.

[13] Rayment, E. L. W. a. I. (2004). "Understanding the importance of protein structure to nature's routes for divergent evolution in TIM barrel enzymes." Acc. Chem. Res. 37: 149-158.

[14] Stephen J Campbell, N. D. G., Richard M Jackson and David R Westhead (2003). "Ligand binding : functional site location, similarity and docking." Current Opinion in Structural Biology 13: 389-395.

[15] Wierenga, R. K. (2001). "The TIM-barrel fold : a versatile framework for efficient enzymes." FEBS 492: 193-198.