行政院國家科學委員會專題研究計畫　期中進度報告

（子計畫二）從序列、結構與功能探討酵素家族之反應機制

(1/3)

計畫類別： 整合型計畫
計畫編號： NSC94-2627-B-009-005-
執行期間： 94 年 08 月 01 日至 95 年 07 月 31 日
執行單位： 國立交通大學生物科技學系(所)

計畫主持人： 楊裕雄

計畫參與人員： 李奇翰 徐青辰 蔡郁吟 柯惠菁 陳美春

報告類型： 精簡報告

處理方式： 本計畫可公開查詢

中　華　民　國 95 年 6 月 1 日

從序列、結構與功能探討酵素家族之反應機制

中文摘要

醯胺水解酵素超家族，包括醯亞胺水解酵素、尿囊素水解酵素、二氫乳清酸水解酵素、海因水解酵素…等，這類酵素雖然在一級序列比對上的相似度可能只有不到百分之三十，但都保留相同催化部份的結構。其中，被認爲與酵素催化有非常重要的關聯並高度保留的有四個組胺酸、一個天門冬胺酸、以及一個經過後修飾的離胺酸與金屬離子螯合。此外，由序列與結構的比對也發現亞硫酸基轉移酵素與核苷酸激酶雖然序列相似度在百分之二十以下，但結構上卻有一定程度的相似，其特定結構的保留可能與特殊的受質/配體結合有關。此二類酵素之關連性的研究亦十分值得進一步研究。本計劃的目的在於整合生物資訊、生物化學以及分子生物學三種不同的途徑來更進一步了解這些酵素催化反應中的反應機制與功能相關性。本研究群與總計畫主持人合作所發展出 FAVA (Feature Amplified Voting Algorithm) ，已可預測出醯亞胺水解酵素中與金屬結合的殘基位置。本研究目前已進一步改良 FAVA 並正在結合 GemDock (Generic Evolutionary Method for molecular DOCKing)爲一整合工具來瞭解醯亞胺水解酵素如何選擇反應基質及金屬離子所扮演的角色，尤其在於單金屬離子與雙金屬離子的功能差異探討。由瞭解醯胺水解酵素超家族的催化方式，我們希望能以蛋白質工程的實驗，進而創造全新催化特性的醯胺水解酵素。在本年度的工作，我們先改良了 FAVA 的效能，增加了與 MSA 相容的模組並評估相容後的效能。另一方面，我們經由結合 docking 工具的分析也完成了部分新穎基質與抑制劑的篩選，並以酵素動力學的測試證明之，最後我們也嘗試對不同來源醯亞胺水解酵素進行結晶結構之解析，以期更清楚瞭解控制此酵素反應之關鍵所在。

關鍵字：醯胺水解酵素超家族、醯亞胺水解酵素、亞硫酸基轉移酵素、核苷酸激酶、FAVA (Feature Amplified Voting Algorithm)、 GemDock (Generic Evolutionary Method for molecular DOCKing)

Abstract

Amidohydrolase superfamily includes imidase, allantoinase, dihydroorotase, hydantoinase, and other related enzymes that were proposed based on the rigidly conserved structural domains in identical positions. Although the overall sequence homology within the enzyme superfamily can be lower than 30%, these conserved regions, compose of four histidines, one aspartate, and one modified lysine which is bridging one or two metal ions are involved in the catalysis. In addition, sulfotransferases and nucleotide kinases were also found to be similar in structure but

different in sequence identity (<20%). The conservation of structural features may be correlated to specific functions such as substrate/ligand binding. The correlation between sulfotransferase and nucleotide kinase is very worth study. The goal of this proposal is to understand the mechanism and functional properties of these enzyme families by using combined techniques of bioinformatics, biochemistry and molecular biology. A newly developed bioinformatics tools, FAVA (Feature Amplified Voting Algorithm) has been used to identify critical amino acids for metal binding in amidohydrolase superfamily. In order to knowing how imidase select specific substrate and the role of metals, an upgraded FAVA will be developed to couple with a well-developed tool -- Gem Dock (Generic Evolutionary Method for molecular DOCKing). FAVA and Gem Dock will be used to identify critical amino acids through available sequences and structures, respectively, in the amidohydrolase superfamily. Biochemical and molecular biological studies will follow to confirm the prediction and to further study the mechanism of imidase. In particular, we would like to distinguish the need for either a binuclear or mononuclear metal centre in imide hydrolysis. In this year, we have accomplished the improvement of FAVA, we increased a compatible module of multiple sequence alignment. We also found several novel substrates and inhibitors by combining the docking technique. Finally, we attempted to resolved several imidase crystal from different sources to obtain the keys of controlling imidase reaction.

Keywords: Amidohydrolase Superfamily, Imidase, Sulfotransferase, Nucleotide Kinase, FAVA (Feature Amplified Voting Algorithm), GemDock (Generic Evolutionary Method for molecular DOCKing)

Introduction

For biochemists, rational design and control a reaction of enzymes/proteins is always a fascinated issue. In order to accomplish this goal, researchers used various strategies in discovering the keys of controlling enzyme mechanisms or reactions. As a rule, sequences and structures are considered as materials in studying the protein characteristics especially in detecting the reaction key residues of an enzyme. Increasing number of protein sequences and structures provided opportunities in developing computational tools in analyzing critical residues of protein active sites and substrate binding pockets. The information could be further used in modifying enzyme reactions.

Imide-hydrolyzing enzymes (E.C.3.5.2.2) were first mentioned in 1940s in the literature to hydrolyze hydantoin derivatives in plant and animals (Bernheim and

Bernheim, 1946; Eadie *et al.*, 1949). These enzymes from various origin show wide substrate specificity in imide compounds, such as dihydropyrimidines, hydantoins and phthalimide (Yang *et al*., 1993, Syltdak *et al.*, 1999; Huang and Yang, 2002). Therefore, the imide-hydrolyzing enzyme was classified as an imidase (Syltdak *et al*., 1999). Imidase may come from various sources and is denominated as dihydropyrimidinase or hydantoinase by the substrate it catalyzed (Syltdak *et al*., 1999). Imidase is an important enzyme in industrial application. It was considered to produce the precursors of various antibiotics and drugs.

In this project, we started with a computational tool, FAVA (Feature Amplified Voting Algorithm), was developed for detecting the critical residues of enzymes and proteins via a well clustering protein sequences. In order to evaluate the efficiency of FAVA, we applied this method in several enzyme superfamilies. The prediction results would soon be confirmed with known structure and mutational experimental data. In this subproject, we focus the application of FAVA in amidohydrolase superfamily. The FAVA results were offered the guideline of site-directed mutagenesis. We further combined the docking techniques to screen the novel substrates and inhibitors. In first year, we improve the efficiency of FAVA algorithm. In this year, we have accomplished the improvement of FAVA, we increased a compatible module of multiple sequence alignment. We also found several novel substrates and inhibitors by combining the docking technique. Finally, we attempted to resolved several imidase crystal from different sources to obtain the keys of controlling imidase reaction.

Goals

To understand the action of enzyme families with the focuses on metal ion dependent imide hydrolysis catalyzed by imidase and nucleotide binding in sulfotransferase. In this year, we planed to accomplished the following specific aim:

1. Sequence analysis and identification of critical residues of imidase and related enzymes using Feature Amplified Voting Algorithm (FAVA).

The candidates of those implied in imidase substrate binding would be predicted by FAVA. The information regarding imidase substrates and substrate coordinators would be extracted by analyzing the sequences of amidohydrolase superfamily with FAVA.

2. Structure analysis and molecular docking using Generic Evolutionary Method for molecular docking (GemDOCK) to search for novel substrates and inhibitors of imidase.

Rules for substrate selection of imidase will be formulated by coupling computational docking method and experimental studies.

Literature Review

Imide-hydrolyzing enzymes (E.C.3.5.2.2) were first mentioned in 1940s in the literature to hydrolyze hydantoin derivatives in plant and animals (Bernheim and Bernheim, 1946; Eadie *et al.*, 1949). These enzymes from various origin show wide substrate specificity in imide compounds, such as dihydropyrimidines, hydantoins and phthalimide (Yang *et al*., 1993, Syltdak *et al.*, 1999; Huang and Yang, 2002). Therefore, the imide-hydrolyzing enzyme was classified as an imidase (Syltdak *et al*., 1999). Imidase may come from various sources and is denominated as dihydropyrimidinase or hydantoinase by the substrate it catalyzed (Syltdak *et al*., 1999). Initial investigation indicated that several imide-hydrolyzing enzymes from mammalian and bacterial sources reported in relatively low sequence similarity. Other studies in evolution of metabolic pathway also found that there were several enzymes that catalyzed similar but different substrates by using the same or partially identical mechanism to imidase (Holm and Sanger, 1997). These enzymes include dihydroorotase, allantoinase, urease and amidohydrolases that were found from mammalian, plants and fungus. Some of these enzymes catalyze the compounds with cyclic amide, and were called cyclic amidohydrolase. The metallo-dependent hydrolase superfamily in SCOP was designated as the ‘amidohydrolase superfamily(Holm and Sanger, 1997). Subsequent analyses have revealed that members of this superfamily also catalyze the hydrolysis of a wide variety of substrates at tetrahedral phosphorus and trigonal carbon centers.

One unique feature of this group is that a mononuclear or binuclear metal centre is required to activate a hydrolytic water molecule for nucleophilic attack. In some instances the metal centre also serves to enhance the electrophilic character of the substrate. The members share two conserved domains: an N-terminal domain that determines substrate specificity; and a $(\beta/\alpha)_8$-barrel domain that contains the catalytic groups. The most common structural motif identified thus far for enzymes with divalent metal ions is the one represented by imidase, also known as dihydropyrimidinase (EC 3.5.2.2), dihydropyrimidine hydrase, or dihydropyrimidine amidohydrolase. This identical catalytic center is also found in dihydroorotase (DHO), allantoinase (ALL) and phenylhydantoinase (Kim *et al.*, 2000). The amino acid residues (four histidines, one aspartate and one modified lysine) that coordinate the metal ions in the active site are highly conserved in imidase and other related enzymes. In its active site, the more solvent-shielded metal ion (M_α) is ligated to two histidine residues from the first β-strand and an aspartic acid from the eighth β-strand. The more solvent-exposed metal ion (M_β) is ligated to two histidine residues from the fifth

and sixth β-strands. In addition to these five ligands, the metal ions are bridged by a carbamate functional group formed from the post-translational carboxylation of the ε-amino group of a lysine from the fourth b-strand. Finally the two metal ions are bridged by a hydroxide from solvent.

There are variations about the metal requirement in the amidohydrase superfamily. What is not so clear is the functional requirement for either a binuclear or mononuclear metal centre or why different metal ions are used. The conservation of the metal ligands for these two subfamilies clearly indicates a divergence from a common ancestral parent. One plausible functional requirement may emanate from the reactions catalyzed by these two systems. With adenosine and cytosine deaminases, cleavage of the C–N bond requires the protonation of both products and thus the resting state of the protein may require bound water for the supply of both protons. By contrast, with DHO and URE, only the amino group product requires protonation and thus the resting state of the enzyme may only require a hydroxide bound between the two divalent cations (Gerlt and Raushel, 2003). Because of this highly conserved catalytic domain, it is very suitable for coupling the bioinformatical methods, molecular techniques and biochemical approaches to clarify the role of metal and catalytic domain playing.

The identification of protein cavities is the starting point for many enzyme engineering studies and structure-based drug design applications. Sites of activity in proteins usually lie in cavities, where the binding of a substrate typically serves as a mechanism for triggering some event, such as a chemical modification or conformational change. Although binding site locations are often furnished by x-ray data or fold recognition, tools that automatically predict these locations have become quite popular in structure-based drug design, especially as front-ends to molecular docking or when alternate binding sites are sought (Ringe, 1995). However, in most of proteins or enzymes, when structure information is hardly available, alternative strategies should be considered.

Extracting useful functional/structural information from a set of amino acid sequences has become imperative for experimental biologist. In general, analyzing the sequence conservation, functional correlation, and structure information of an interested sequence and other sequences that belong to a protein/enzyme family or superfamily is usually the effective way to obtain the desired information, like functional or structural critical residues which can assist biologists for further studies at molecule level. In previous studies, various pairwise alignment and multiple sequence alignment methods were used to detect the conserved residues implied functional role in a set of sequences (Altschul *et al.*, 1990; 1994; 1997), however, this information is not sufficient to indicate which residues were involve in catalysis or

responsible for other roles. Other sequence comparison algorithms like SSEARCH (Smith and Waterman, 1981), FASTA (Pearson and Lipman, 1988), or BLAST (Altschul *et al.,* 1994) were also have the same problems. In stead of alignment methods, several strategies have been proposed to predict functional residues from sequence alone. The most frequently used approaches are the motif-based sequence analysis (Henikoff and Henikoff, 1991; Sonnhammer and Kahn, 1994; Ogiwara *et al*., 1996; Liu *et al*., 2003; Puntervoll *et al.,* 2003). However, the motif-based approaches resulted in too many false-positives to be useful in large-scale protein superfamily. Phylogenomic techniques, like evolutionary trace method in identifying functional important residues (Lichtarge *et al*., 1996), which uses evolutionary information to improve accuracy are particularly useful in large-scale analyses. The Evolutionary Trace Method allows scientists to perform a variety of sequence analyses such as variations across a family of naturally occurring mutants or isoforms, or subtle differences between the same protein from different species or tissues, and automatically relate the results back to a given structure and identify key features structurally clustered around substrate and dimmer interfaces (Lichtarge *et al*., 1997; Innis *et al*., 2000; Lichtarge e*t al*., 2002; Madabushi *et al*., 2004). Although the Evolutionary Trace Method could find out the key residues, the identification of enzyme binding pocket need to more sophisticated analysis.

Voting or voting-like concept has been widely used in different computing algorithms for various purposes. In computational biological applications, voting concept was often integrated with neural network for protein clustering and structure prediction (King *et al*., 2001). In this study we introduced this concept to an enzyme superfamily sequence analysis for searching functional key residues. According to the observation in sequence identity and biochemical properties of the proteins in amidohydrolase superfamily, we developed a feature amplified voting algorithm to identify the functional key residues of a rat imidase. The identified critical residues will be clarified by experimental references and available structural information. In our previous studies, we used a Feature Amplified Voting Algorithm (FAVA) to find out the metal coordinators of mammalian imidase (Lee *et al.,* submitted to *Bioinformatics*) and the toxicity caused residues of RNase signal peptide. In this study, we will try to upgrade this algorithm to explore the imidase binding pocket by analyzing the sequences of cyclic amidohydrolase superfamily.

Dimental center was observed in bacterial enzymes which include DHO, ALL, and hydantoinase, and it was found that the carboxylated lysine residue bridges the two metal ions. However, mammalian imidase and the fish imidase contain only single divalent metal ions in the active site. Similar requirement was observed for some adenosine and cytosine deaminase (Kennedy and Gibney, 2001). There is an

exception for a bacterial (*Aquifex aeolicus*) DHO which was reported to contain single divalent metal ions in the active site recently (Ahuja *et al.*, 2004). The first crystal structure of hydantoinase from *Thermus sp.* was solved recently (Abendroth *et al.*, 2002). A proposed mechanism was based this structure, hydantoin enters the active–site cavity and bind with carbonyl oxygen atom to Zn1. Upon substrate binding, a flexible flap closes the active site, leading to a rearrangement of residues, which now build up a strong hydrogen bond network with hydantoin. In a concerted reaction, the amino group approaches Zn2. This brings the bridging hydroxide ion in close proximity to the activated carbonyl carbon atom, enabling its nucleophilic attack.

On the other hand, the mammalian or fish imidase that contains single metal ion seems to have different catalytic mechanism. The function of the metal ion in imidase is to coordinate substrate and maintain a suitable active site environment but not directly involve the ionization of a water molecule. Instead, a histidine may be responsible for the ionization of water and aspartic acid is also found to be important for imide hydrolysis and may function as a proton donor/acceptor. The metal coordinated water may be activated by ionization, polarization, or poised for displacement. Imidase metal ion may function as a Lewis acid for catalysis to stabilize the developing negative charge of imide in the transition state or to fix the orientation of the metalbound substrate so that it is in position for nucleophilic attack by water or a hydroxide ion activated by a histidine. Although several catalytic mechanisms were proposed as described, none of them provide any direct evidence to explain some fundamental questions such as how the enzyme select different substrates and why the number of metals and the type of metals used by various similar enzymes. We plan to begin our study by surveying the huge amount of the available sequence and structure information. An upgrading FAVA program will attempt to predict the residues of substrate binding pocket then verify the predicting results by site-directed mutagenesis following by a variety of biochemical analysis. The new version FAVA will increase the sensitivity in exploring the connective functional patterns. The obtained results may provide a clue to elucidate molecular mechanism of imidase and help to create newly catalytic abilities for novel substrates.

Methods

The overall experimental scheme of this research is summarized and shown in Scheme I. Four approaches will be taken to study the action of imidase at molecular level.

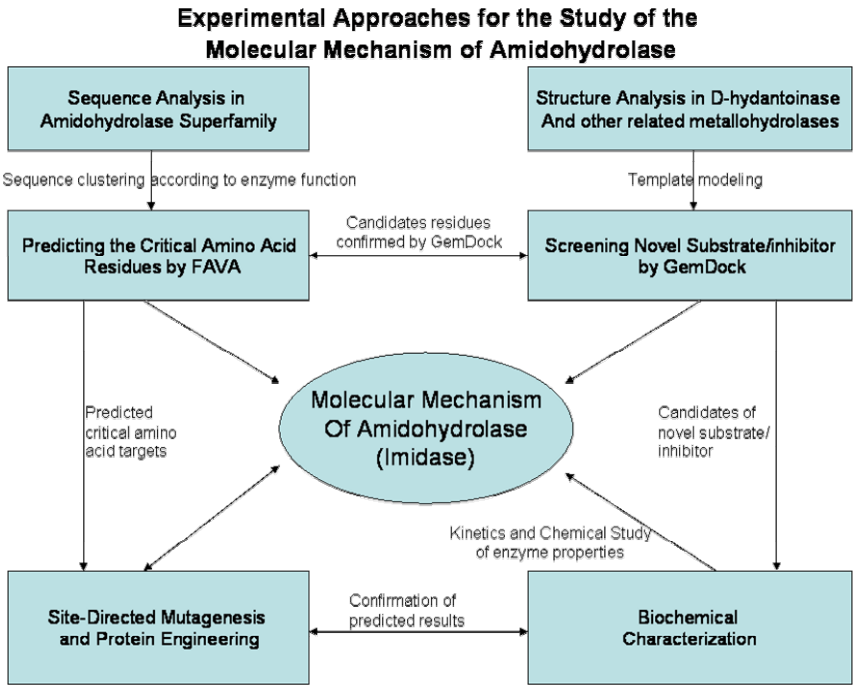

Figure 1 Scheme for the experimental approaches

In the first year of this project, we will focus on the development of bioinformatics tools, updated FAVA and GemDOCK, and sequence and structure analysis will be followed. Novel sequence will also be obtained following cloning and expression of fish liver imidase. Site-directed mutagenesis will follow right after critical amino acids were identified by FAVA.

Using FAVA to identify critical residues responsible for the action of imidase

The sequences of amidohydrolase superfamily were obtained from Protein Information Resource (PIR). The downloaded sequences would further classified into groups according to their substrate spectrum. The clustered sequences then input the Feature Amplified Voting Algorithm and the desired enzyme binding pocket would be predicted by the algorithm. After comparing the output candidates with substrate, the dependence of substrate and the pattern of binding pocket would be revealed. The Feature Amplified Voting Algorithm was progressed in two steps. The first step was to align the target, substrate identical enzymes (property A proteins) and substrate related proteins (~A proteins) by a three-optimal alignment. The three-optimal alignment was designed to align three sequences in space, such alignment would help to get best alignment result for each vote among the three protein groups (Target, ~A and A). When alignment was done, we expect to identify amino acid residues critical for target enzyme substrate binding pockets exist in target and property A proteins, but absent in ~A proteins. In the second step, a voting score (V-score) was given based on

previous assumption then the V-score was sum up in each comparison.

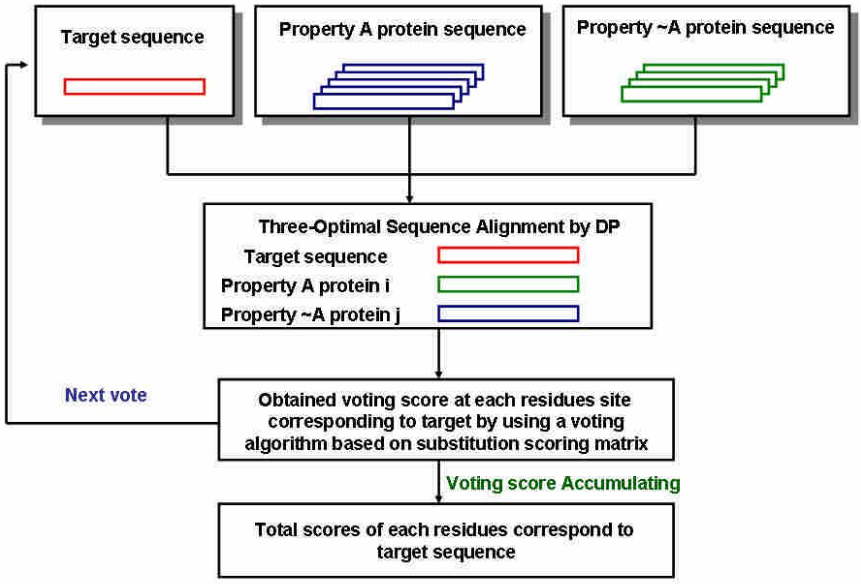

Figure 2 Flowchart of Feature Amplified Voting Algorithm. In this study, the target sequence would first align simultaneously to property A proteins and ~A proteins, then obtained voting score (V-score) by voting algorithm based on Blosum62 scoring matrix. Finally, the scores of each vote would be accumulated at each corresponded residues in target sequence and obtained a total score (T-score).

3.3.2 Using GemDOCK to screen for novel substrate and inhibitor of imidase

Currently, four X-ray crystal structures of hydantoinase were resolved. According to the sequence identity of hydantoinase from *Agrobacterium tumefaciens* and those with crystal structure, hydantoinase of *Burkholderia pickettii* (PDB code 1NFG) was selected as template. The sequence identity of both sequences is 92.6%, and different amino acids are mostly located in the C-terminal, which is not directly involved in catalysis. Insight II is used to model the three dimensional structure of *Agrobacterium tumefaciens* hydantoinase. First, homology modeling is used to obtain the overall structure of *Agrobacterium tumefaciens* hydantoinase form the template structure. Then, posttranslational modification of carboxylysing on lysine is modified by computation. Finally, Insight II is used to minimze the energy of the modeled structure. The constructed structure of *Agrobacterium tumefaciens* hydantoinase is then used as the docking protein in GemDOCK.

GemDOCK is a flexible ligand docking program, that uses a Generic Evolutionary Method and combines a scoring function. According to the interaction between small compound and protein, the best fit energy will be generated and produce a suitable orientation for small compounds docking within the protein. Compounds pick up by GemDOCK will be confirmed through experimental assay. A rapid spectrophotometric assay was used as the standard assay. Briefly, the decrease in absorbency at 298 nm was measured upon hydrolysis of phthalimide as the substrate at 25℃. To start the reaction, the enzyme solution was added into 1 ml solution, containing 1 mM phthalimide and 100 mM Bis-Tris proprane at pH 7.0. Under these

conditions, a change in A298 of 2.26 represents the hydrolysis of 1 mol of phthalimide. The hydrolysis of imide was monitored with a UV/VIS spectrophotometer (Hitachi U3300). HPLC was used for assay novel substrates of hydantoinase. The column was RP18-reverse phase and mobile phase was 100% H_2O. UV-Vis detector was used at 214 nm. Substrates were incubated at 25℃ with enzymes and 100 mM Bis-Tris propane at pH 7.0. Blank reaction was made only different in without enzyme. Novel inhibitors will be examined using standard assay conditions, except that substrate concentration is 0.5 mM phthalimide.

3.2.3 Using biochemical and molecular biological techniques to study the properties of imidase

Protein purification and enzyme assay. The fish imidase activity were obtained through ammonium sulfate precipitation, octyl, chelating, DEAE, and hydroxyapatite chromatography and it was Approximately 6000-fold purification and a 40% yield. A rapid standard spectrophotometric assay was devised in which the decrease in absorbance at 298 nm was measured upon hydrolysis of phthalimide as substrate. Steady-state kinetic constants K_m and k_{cat} will be determined for wild-type and each of the mutants.

Result and Discussion

Improvement of FAVA

FAVA was previously used in finding functional constraints in enzyme superfamily. We successfully used this method to find out the metal coordinators of mammalian imidase and predicted the toxicity caused residues of Rnase signal peptide. The results from FAVA suggest the candidates of an interested sequence by given scores at each residues from accumulating each voting score among target sequence, property A sequence and property ~A sequence. However, considering the widely used of MSA tools, it is necessary to incorporate the MSA results into FAVA. Therefore, we redesigned FAVA with a MSA compatible version, the scheme of new version FAVA was showed in Figure 3. We also evaluated the new version FAVA in detecting the metal coordinators in cyclic amidohydrolase superfamily and catalytic determining residues in protein tyrosine phosphatase family the detail results was showed in Figure 4 and 5. The residues list was showed in Table 1 and 2. An example of FAVA detail report was showed in appendix. The results of FAVA prediction in the two superfamilies were confirmed by available structure and literatures implied the new version FAVA indeed detected the critical residues of proteins. A comparison of different alignment modules in FAVA were also showed in figure 6. The result indicated when sequences of this superfamily diverse the three-optimal alignment module have higher sensitivity.

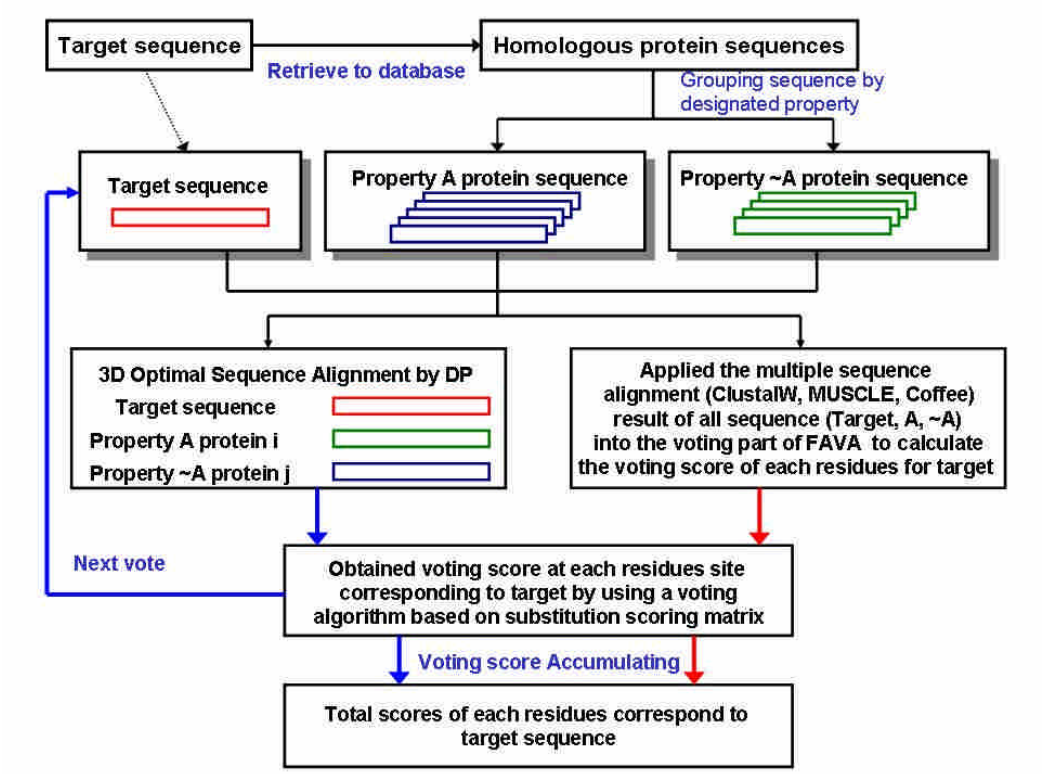

Fig. 3. **The new version Feature Amplified Voting Algorithm**

This figure show the scheme of new version feature amplified voting algorithm (FAVA). The selected target sequence was used to retrieve homologous protein sequences from databases. The homologous sequence were then clustered into property A group (property orthologs) and property ~A group (property paralogs) and applied to FAVA. The sequences could process in two different alignment procedures (Three optimal alignment and MSA represented in blue root and red one, respectively) and accumulating the FAVA score of each residue for target sequence.

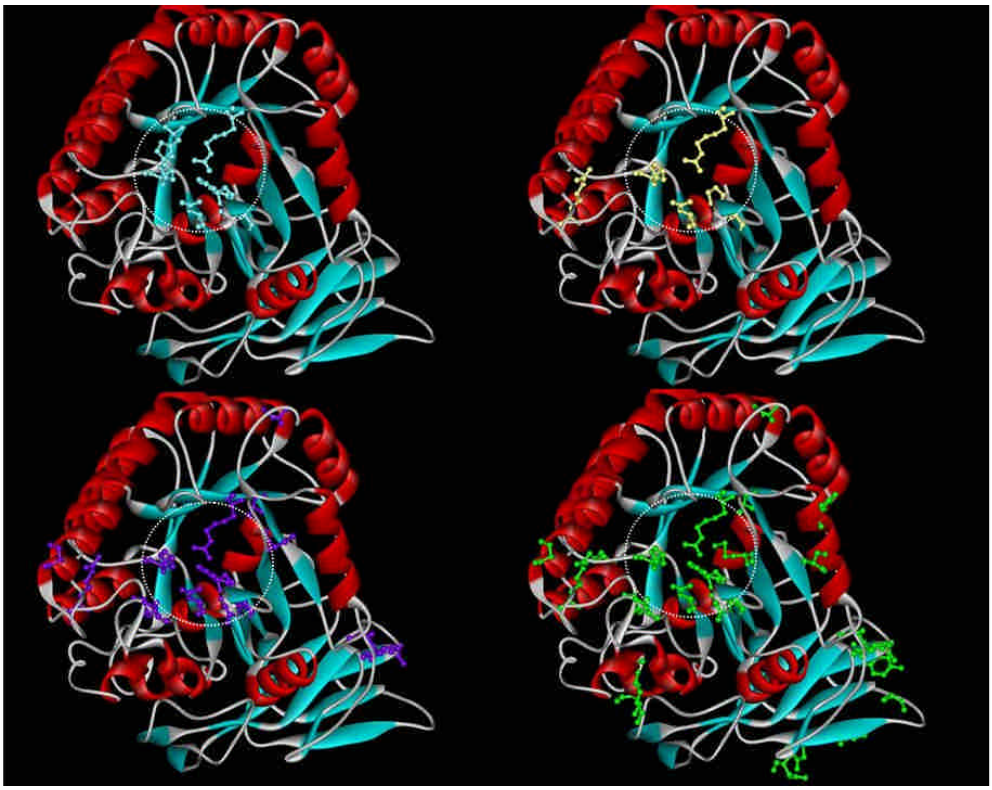

Fig. 4. The FAVA prediction results in the case study of amidohydrolase superfamily. (A) The metal coordinators of a known structure 1GKP. (B)(C)(D) Showed the different default cutoff (1%, 3% and 5%, respectively) of FAVA prediction, the cutoff could be considered as a limiting condition in distinguishing the potential determinants of designated property. All positions refer to pdb structure 1GKP, a bacterial hydantoinase.

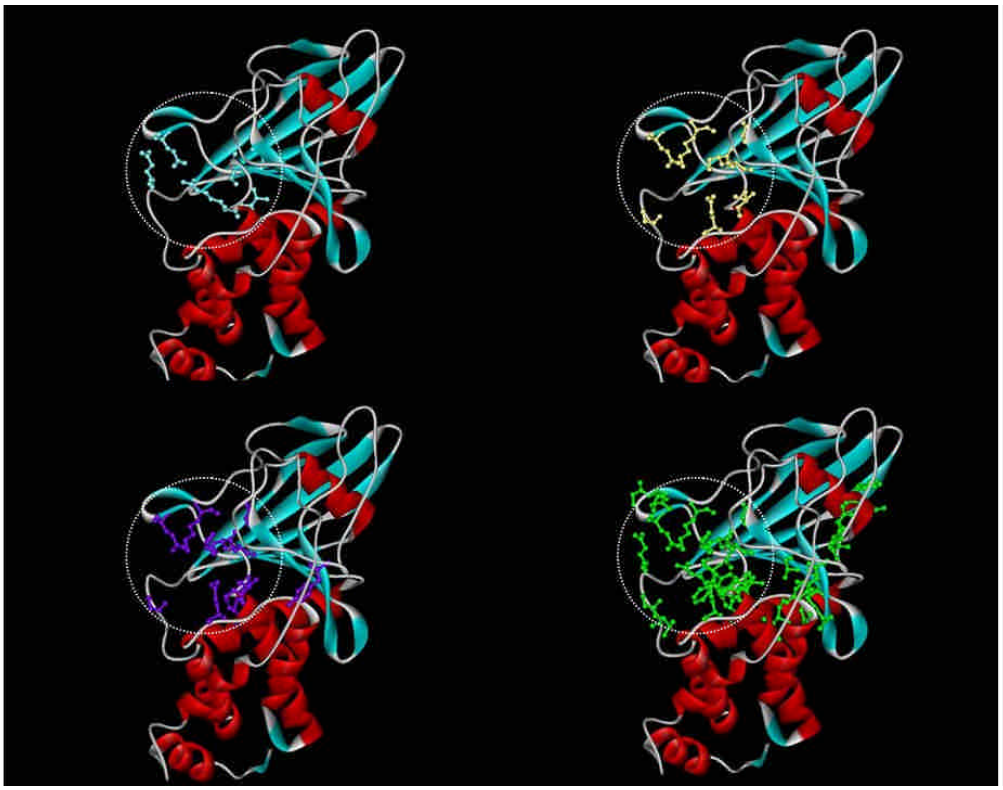

Fig.5. The FAVA prediction results in the case study of classical PTP family. (A) The phosphatase active site of human LAR D2 domain (PDB 1LAR). (B)(C)(D) Showed the different default cutoff (3%, 5% and 10%, respectively) of FAVA prediction, the cutoff could be considered as a limiting condition in distinguishing the potential determinants of phosphoryl-substrate binding.

Fig.4. An efficiency evaluation in applied different alignment strategies in FAVA. The three different alignment methods were applied in case study I (amidohydrolase superfamily). The result showed the three-alignment and advance alignment method (MUSCLE) have higher sensitivity in potential determinant detecting in amidohydrolase superfamily.

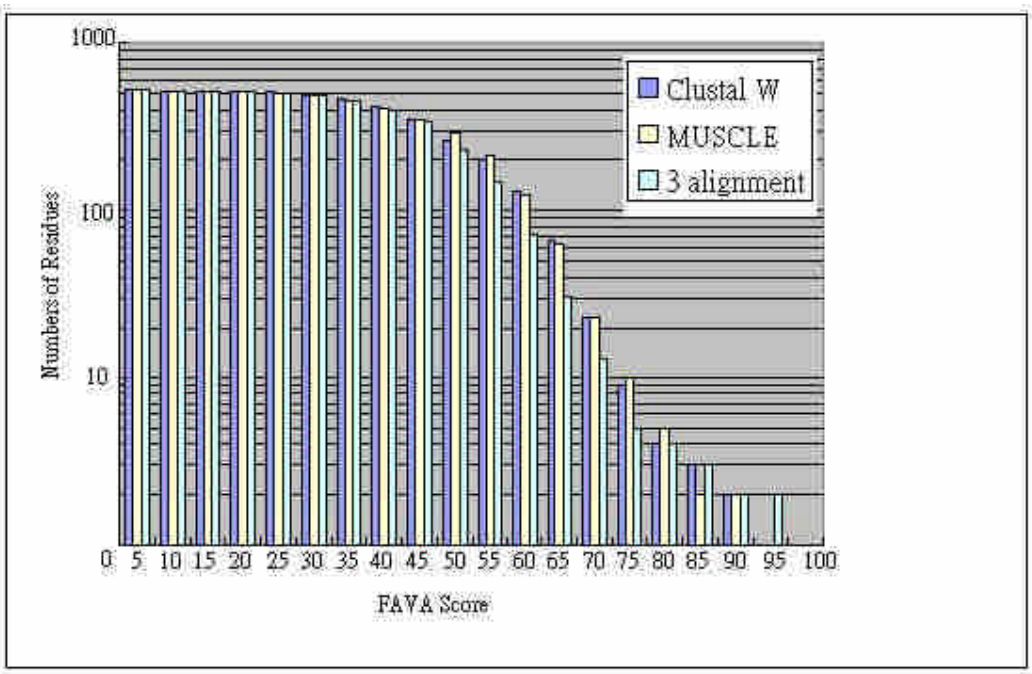

Fig.6. An efficiency evaluation in applied different alignment strategies in FAVA. The three different alignment methods were applied in case study I (amidohydrolase superfamily). The result showed the three-alignment and advance alignment method (MUSCLE) have higher sensitivity in potential determinant detecting in amidohydrolase superfamily.

Table 1

FAVA prediction result in different cut off range of amidohydrolase superfamily

Cut off	Figure	FAVA Prediction Sites
Metal coordinator	**2A**	**H59, H61, K150, H183, H239, D315**
1%	**2B**	**H239, D315, H61**, R292, **K150**
3%	**2C**	**H239, D315, H61**, R292, **K150**, M287, **H59** W448, R30, M93, A126, S199, A333, L172
5%	**2D**	**H239, D315, H61**, R292, **K150**, M287, **H59** W448, R30, M93, A126, S199, A333, L172
10%		**H239, D315, H61**, R292, **K150**, M287, **H59** W448, R30, M93, A126, S199, A333, L172, K284, D439, S354, Q306, L108, E104, D45, V43, A37,Y22

Note—The bold residues were identified as metal coordinators in pdb structure 1GKP.

All positions were referred to the structure 1GKP.

Table 2

FAVA prediction result in different cut off range in classical PTP family

Cut off	Figure	FAVA Prediction Sites
Active site reisdues	**3A**	**E126, C222, S223, R228, T229**
1%		A224, **S223**
3%	**3B**	A224, **S223**, V226, K131, Q266, G190, G97, **E126**
5%	**3C**	A224, **S223**, V226, K131, Q266, G190, G97, **E126**, E268, G227, T267, Y59, F201, T122, I233, Y92, Q260
10%	**3D**	A224, **S223**, V226, K131, Q266, G190, G97, **E126**, E268, G227, T267, Y59, F201, T122, I233, Y92, Q260, Y60, A94, T184, **T229, R228**, F183, V191, R176, **C222**, G128, W111, I218, R129, R63

Note—The bold residues were identified which involved in protein tyrosine phosphatase activity in pdb structure 1LAR. All positions were referred to the structure 1LAR D2 domain.

Effect of metal ions on the catalytic properties of recombinant bacterial hydantoinase

After obtained the FAVA prediction, we used the prediction result to approach the metal effects in imidase. Hydantoinase is a metalloenzyme that catalyzes the hydrolytic cleavage of dihydropyrimidines and 5'-monosubstituted hydantoins. The crystal structures have shown that bacterial hydantoinase possesses a binuclear metal center in which two metal atoms are bridged by post-carboxylated lysine. Here, the catalytic properties of bacterial hydantoinase were further investigated. The activity of mutants for the metal binding site, H57A, H59A, K148A, H181A, H237A, and D313A, was completely abolished. The activity of K148A could be chemically rescued by short-chain carboxylic acids and cobalt, indicating the role of the post-carboxylated lysine was to coordinate the binuclear ion within the active site of hydantoinase. The mutant D313E resulted in a different pH profile compared with that of wild-type hydantoinase. Based on the experimental results presented in this study and the crystal structure of hydantoinase, the kinetic mechanism of the enzyme was proposed.

Enzyme	Cobalt content per enzyme subunit	Specific activity[b] (μmol/min/mg)
wild-type	2.5 ± 0.1	1.1 ± 0.1
H57A	1.2 ± 0.1	$< 10^{-4}$
H59A	1.2 ± 0.1	$< 10^{-4}$
K148A	0.5 ± 0.1	$< 10^{-4}$
K148C	0.8 ± 0.2	$< 10^{-4}$
K148D	0.8 ± 0.2	$< 10^{-4}$
K148S	0.9 ± 0.0	$< 10^{-4}$
H181A	1.2 ± 0.1	$< 10^{-4}$
H237A	1.2 ± 0.1	$< 10^{-4}$
D313A	1.4 ± 0.1	$< 10^{-4}$
D313E	2.2 ± 0.1	0.05 ± 0.01

Reference

Abendroth, J., Niefind, K., and Schomburg, D., (2002) *J. Mol. Biol.* **320**, 143-156.

Ahuja, A., Purcarea, C., Ebert, R., and Sadecki, S., (2004) *J. Biol. Chem.*, in press.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman D. J., (1990) *J. Mol. Biol.*, **215,**

403-410.

Altschul, S. F., Boguski, M. S., Gish, W., and Wooten, J. C., (1994) *Nature Genet.* **6**, 119-129.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res,* **25**, 3389-3402.

Bernheim, F., and Bernheim, M. L. C., (1946) *J. Biol. Chem.* **163**, 683-685.

Chen W.T., Liu M.C., Yang Y.S. (2005) *Anal. Biochem.*, in press.

Eadie, G. S., Bernheim, F., and Bernheim, M. L. C., (1949) *J. Biol. Chem.* **181**, 449-458.

Gerlt, J. A., and Raushel, F. M., (2003) *Curr. Opin. in Chem. Biol*. **7**, 252-264.

Hartley, B.S., and Kilby, B.A., (1954) *Biochem. J*. **56**, 288-293.

Henikoff, S. and Henikoff, J.G.., (1991) *Nucleic Acids Res.* **19**(23), 6565-6572.

Holm, L., and Sander, C., (1997) *Proteins* **28**, 72-82.

Huang, C. Y, and Yang, Y. S., (2002) *Biochem. Biophys. Res. Commun.* **297**, 1027-1032.

Huang, C. Y., Chiang, S. K., Yang, Y. S., and Sun, Y. J., (2003) *Acta Crystallogr. D* **59**, 643-945.

Huang, C. Y., and Yang, Y. S., (2003) *Biochem. Biophys. Res. Commun.* **312**, 467-472.

Huang, C. Y., and Yang, Y. S., (2004) *Protein Expr. Purif. Accepted.*

Huang, C. Y., Chao, Y. P., and Yang, Y. S., (2003) *Protein Expr. Purif.* **30**, 134-139.

Ireton, G. C., McDermott, G., Black, M. E., and Stoddard, B. L., (2002) *J. Mol. Biol.* **315**, 687-697.

Kakuta Y., Sueyoshi T., Negishi M., Pedersen L.C. (1999) *J Biol Chem.* **274**: 10673-10676.

Kennedy, M.L., and Gibney, B.R., (2001) *Curr. Opin. Struc. Biol*. **11**, 485-490.

Kim, G. J., Lee, D. E., and Kim, H. S., (2000) *J. Bacteriol*. **184**, 7021-7028.

King, R. D., Karwath, A., Clare, A., and Dehaspe, L., (2001) *Bioinformatics* **17**, 445–454.

Koch, M.A., Breinbauer, R. and Waldmann, H., (2003) *Biol. Chem.*, 384, 1265 – 1272.

Lichtarge, O., Bourne, H. R., and Cohen, F. E., (1996) *J Mol Biol* **257**(2), 342-358.

Lichtarge, O., Yamamoto, K. R, and Cohen F. E., (1997) *J Mol Biol* **274**(3), 325-337.

Innis, C. A., Shi, J., and Blundell, T.L., (2000) *Protein Eng* **13**(12), 839-847.

Lichtarge, O., Sowa, M. E., and Philippi, A., (2002) *Methods Enzymol* **344**, 536-556.

Lin, E.-S. and *Yang, Y.-S. (2000) *Biochem. Biophys. Res. Comm*., **271**, 818-822.

Hsiao, Y.-S. and *Yang, Y.-S. (2002) *Biochemistry*, **41**, 12959-12966.

Lin, E.-S., Yang, J.-M. and *Yang, Y.-S. (2003) *J Chinese Chemical Soc*, **50**, 655-663.

Liu, A.H., Zhang, X., Stolovitzky, G.A., Califano, A. and Firestein, S.J., (2003) *Genomics* **81**(5), 443-456.

Madabushi, S., Gross, A. K., Philippi, A., Meng, E. C., Wensel, T. G.., and Lichtarge, O., (2004) *J Biol Chem* **279**(9), 8126-8132.

McGuigan, C., Gudavalli, R., Letunic, I., Bork, P., Rychlewski, L., Kuster, B., Helmer, C. M., Hunter, W. N., Aasland, R., and Gibson, T. J., (2003) *Nucleic Acids Res.* **31**(13), 3625-3630.

Ogiwara, A., Uchiyama, I., Takagi, T., and Kanehisa M., (1996) *Prot. Sci*. **5,** 1991-1999.

Pearson, W. R. and Lipman, D. J., (1988) *Proc. Natl. Acad. Sci.*, **85**(8), 2444–2448.

Puntervoll, P., Linding, R., Gemund, C., Chabanis, D. S., Mattingsdal, M, Cameron, S., Martin D. M., Ausiello, G., Brannetti, B., Costantini, A., Ferre, F., Maselli, V., Via, A., Cesareni, G., Diella, F., Superti, F. G., Wyrwicz, L., Ramu, C., McGuigan, C., Gudavalli, R., Letunic, I., Bork, P., Rychlewski, L., Kuster, B., Helmer, C. M., Hunter, W. N., Aasland, R., and Gibson, T. J., (2003) *Nucleic Acids Res.* **31**(13), 3625-3630.

Ogawa, J., Soong, C.L., Honda, M., and Shimizu, S. (1999), *Appl. Environ. Microbiol.* **65**, 1459-62.

Ringe, D., (1995) *Cur. Op. Struct. Biol.* **5**, 825.

Runser, S. M., and Meyer, P.C., (1993) *Eur. J. Biochem*. **231**, 1315-1324.

Ruppert, J., Welch, W. ,and Jain, A. N., (1997) *Prot. Sci*. **6**, 524.

Smith, T. F. and Waterman, M. S., (1981) *J. Mol. Biol*. **147**, 195-197.

Su, T. M., and Yang, Y. S., (2000) *Protein Expr. Purif.* **19**, 289-297.

Sonnhammer, E. and Kahn, D., (1994) *Prot. Sci*. **3**, 482-492.

Sugahara T., Yang Y.S., Liu C.C., Pai T.G., Liu M.C. (2003) *Biochem J*. **375**: 785-791.

Su, T.-M. and *Yang, Y.-S. (2003) *Biochemistry*, **42**, 6863-6870.

Syldatk, C., May, O., Altenbuchner, J., Mattes, R., and Siemann, M., (1999) *Appl. Microb. Biotechnol.* **51**, 293-309.

Wallach, D.P. and Grisolia, S., (1957) *J. Biol. Chem.* **163**, 277-288.

Syldatk, C., and Pietzsch, M., Hydrolysis and formation of hydantoins. In: K. Drauz, H. Waldmann (Eds), Enzyme Catalysis in Organic Synthesis. VCH, Weinheim. Germany, pp.409-431

Yang, Y. S., Ramaswamy, S., and Jakoby, W. B., (1993) *J. Biol. Chem.* **268**, 10870-10875.