Identification of catalytic residues in protein structures using a novel feature that integrates the micro-environment and geometrical location properties of residues

Lei Han[1], Jiangning Song[2,3], Ming S. Liu[4,*] and Ziding Zhang[1,*]

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China [2]National Engineering Laboratory for Industrial Enzymes and Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China [3]Department of Biochemistry and Molecular Biology, Faculty of Medicine, Monash University, Melbourne, VIC 3800, Australia [4]CSIRO - Mathematics, Informatics & Statistics, Private Bag 33, Clayton South 3169, Australia

Introduction

Enzymes help to fulfill diverse biochemical functions and play critical roles in almost all cellular processes. Although different mechanisms of some enzymes have been characterized，there is a difficulty in rationalizing the available enzyme sequences/structures with their annotated functions. Identification of catalytic residues (CRs) is the first and important step to characterize the catalytic mechanism and function of an enzyme. Since experimental determination of CRs on large-scale proteome data is still a costly and daunting task, computational methods to predict CRs from enzyme sequences/structures are playing an increasingly important role in complementing the experimental identification. Though there are lots of features for predicting CRs, advancing novel features does not only increase the prediction accuracy but also deepen our understanding of catalytic mechanisms.

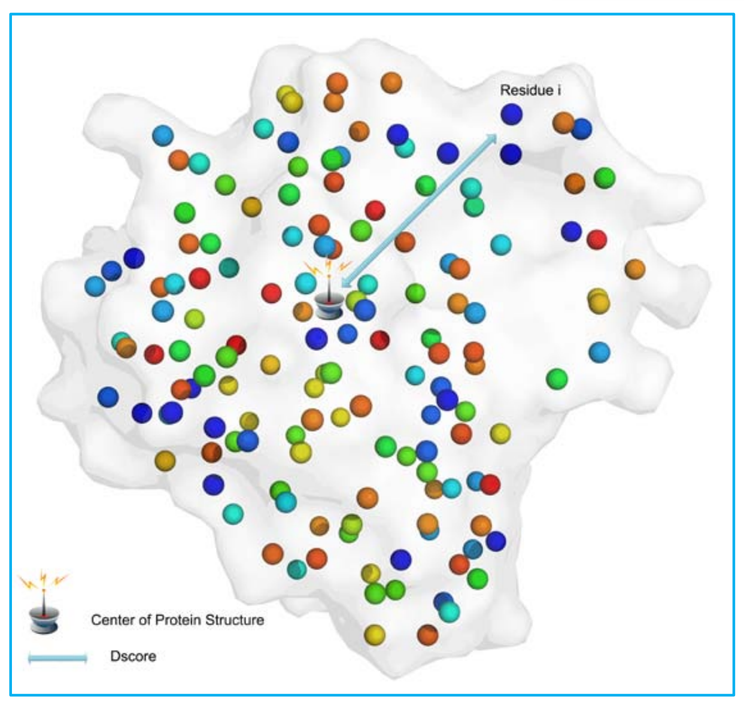
In this work, we developed a novel structural feature called MEDscore to determine the CRs. Firstly, a residue’s micro-environment (ME) was converted into a series of spatially neighboring residue pairs and a ME-based score (i.e. MEscore) was proposed to quantify a residue’s ME information. Secondly, a parameter named Dscore was set up to measure a residue’s global positional information. Finally, MEDscore was defined from an effective nonlinear integration of MEscore and Dscore.

Materials and Methods

Benchmark Enzyme Dataset

The enzyme dataset used in this study was based on the Catalytic Site Atlas (CSA) database (version 2.2.12).After filtering homology sequences and based on other criteria, 223 enzyme structure domains were retained in our final dataset. These structure domains were downloaded from ASTRAL.

Definition and Calculation of Dscore



Geometrical center of one protein structure:

$$\left(c_x, c_y, c_z\right) = \left(\frac{\sum_{i=1}^{N} x_i}{N}, \frac{\sum_{i=1}^{N} y_i}{N}, \frac{\sum_{i=1}^{N} z_i}{N}\right)$$

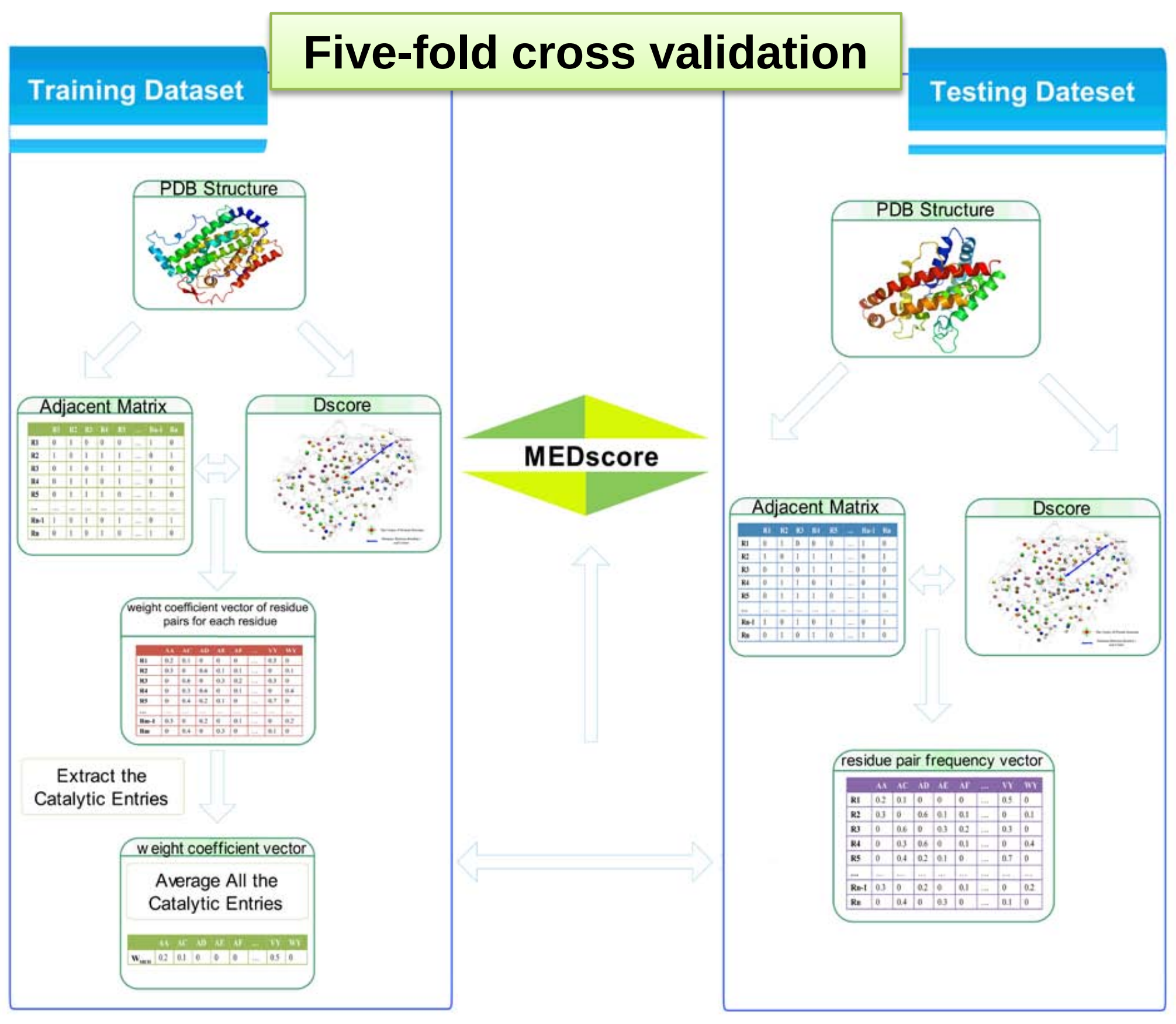
The distance between a residue i and the center of the structure (i.e. Dscore_i)

$$Dscore_i = \sqrt{(c_x - x_i)^2 + (c_y - x_i)^2 + (c_z - x_i)^2}$$

Definition and Calculation of MEDscore

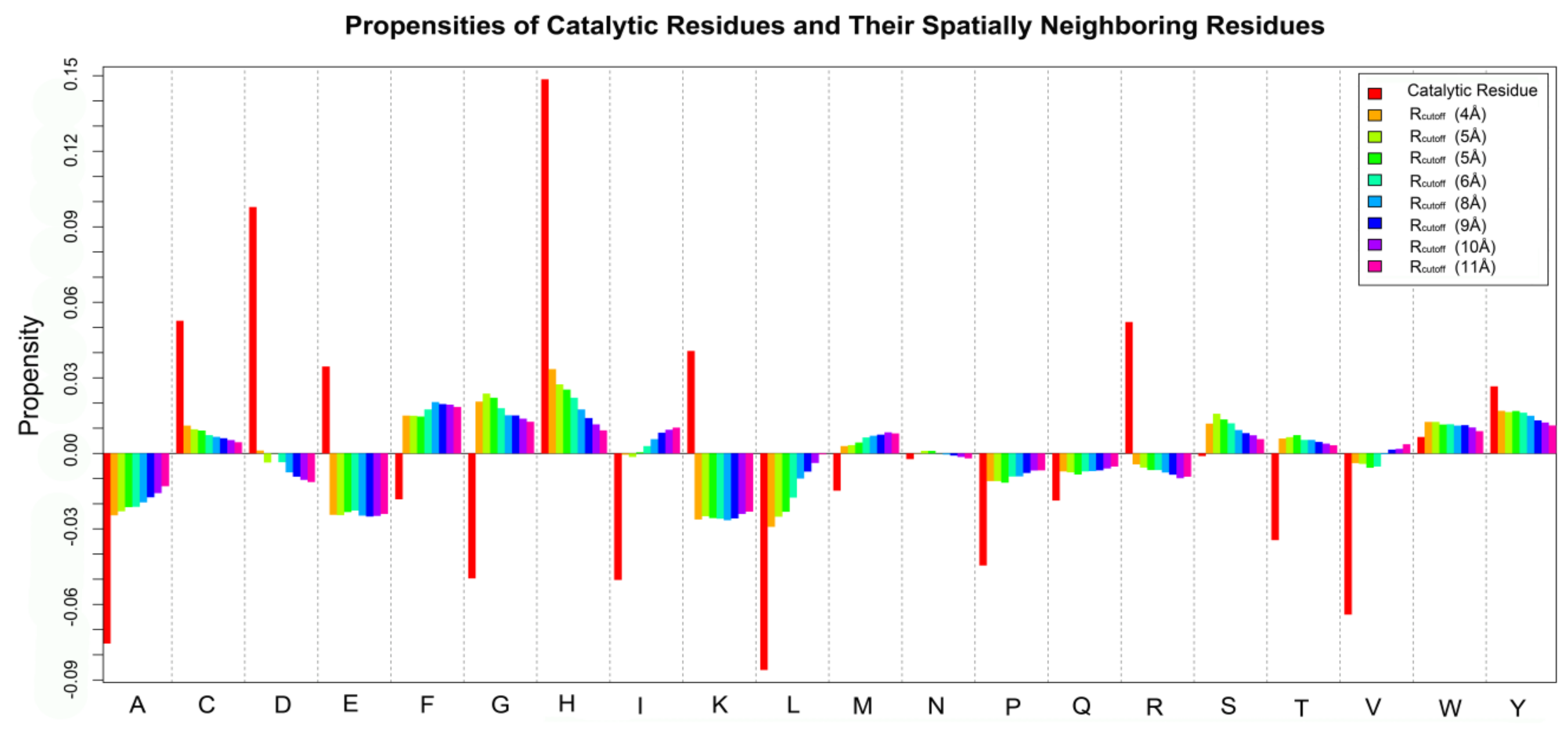
MEDscore is a feature which integrates the micro-environment (ME) and geometrical properties of amino acid residues. The flow-chart of calculating MEDscore is shown at right.

The ME of a residue (MEscore) is represented as a series of spatially neighboring residue pairs in the local of the query residue. The details of calculation MEDscore is similar as the process of measuring MEDscore.

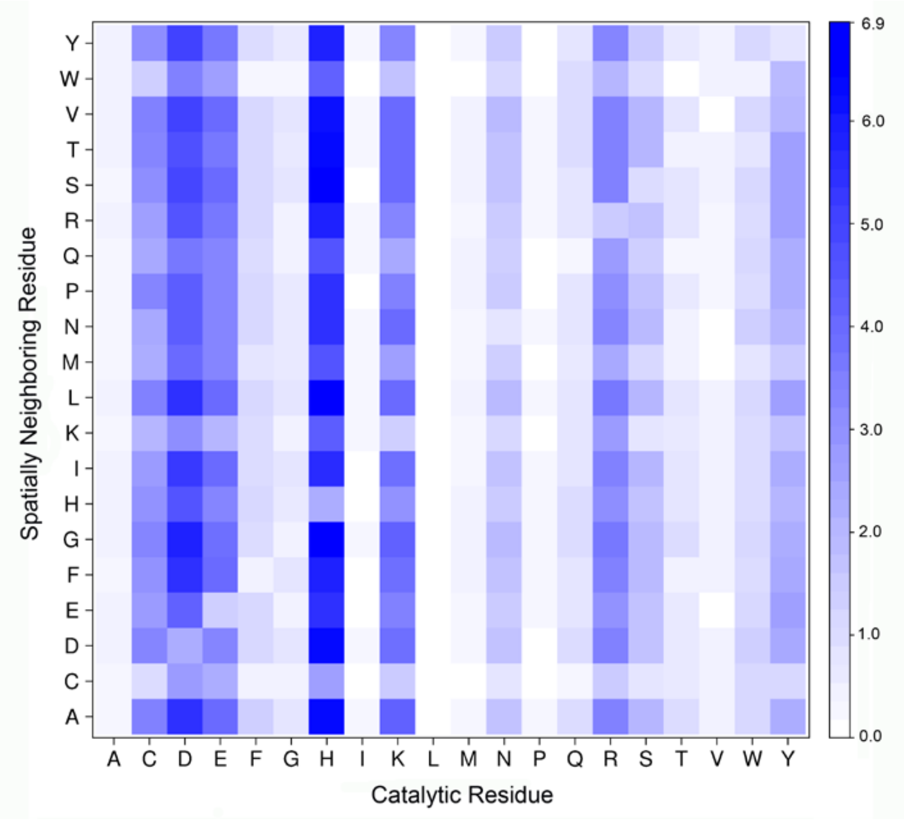

Five-fold cross validation
Training Dataset
Testing Dateset
PDB Structure
Adjacent Matrix
Dscore
weight coefficient vector of residue pairs for each residue
Extract the Catalytic Entries
weight coefficient vector
Average All the Catalytic Entries
MEDscore
PDB Structure
Adjacent Matrix
Dscore
residue pair frequency vector

Results and Discussion

Propensities of residues in the ME surrounding the CRs

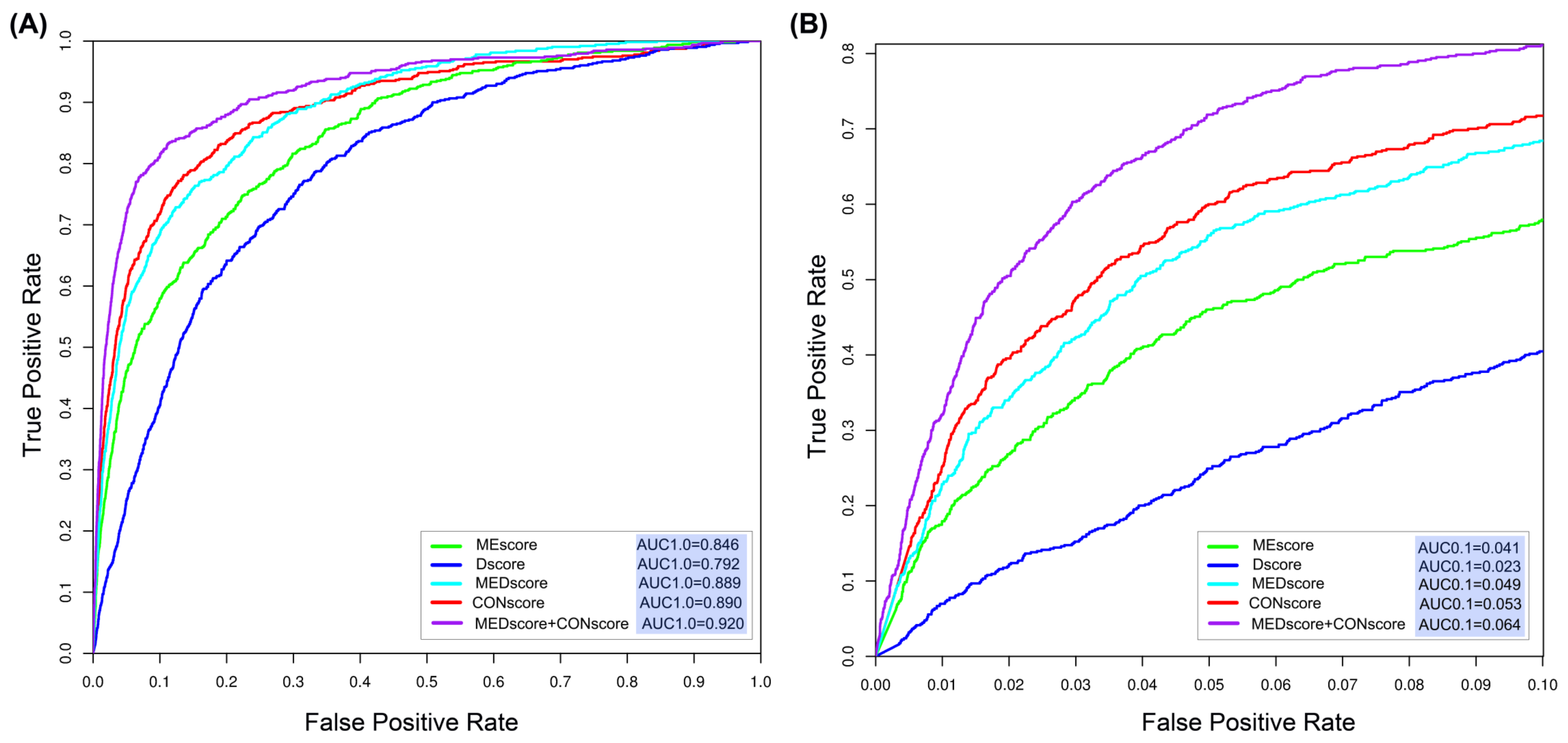


The weight coefficients of spatially neighboring residue pairs in the MEs of CRs

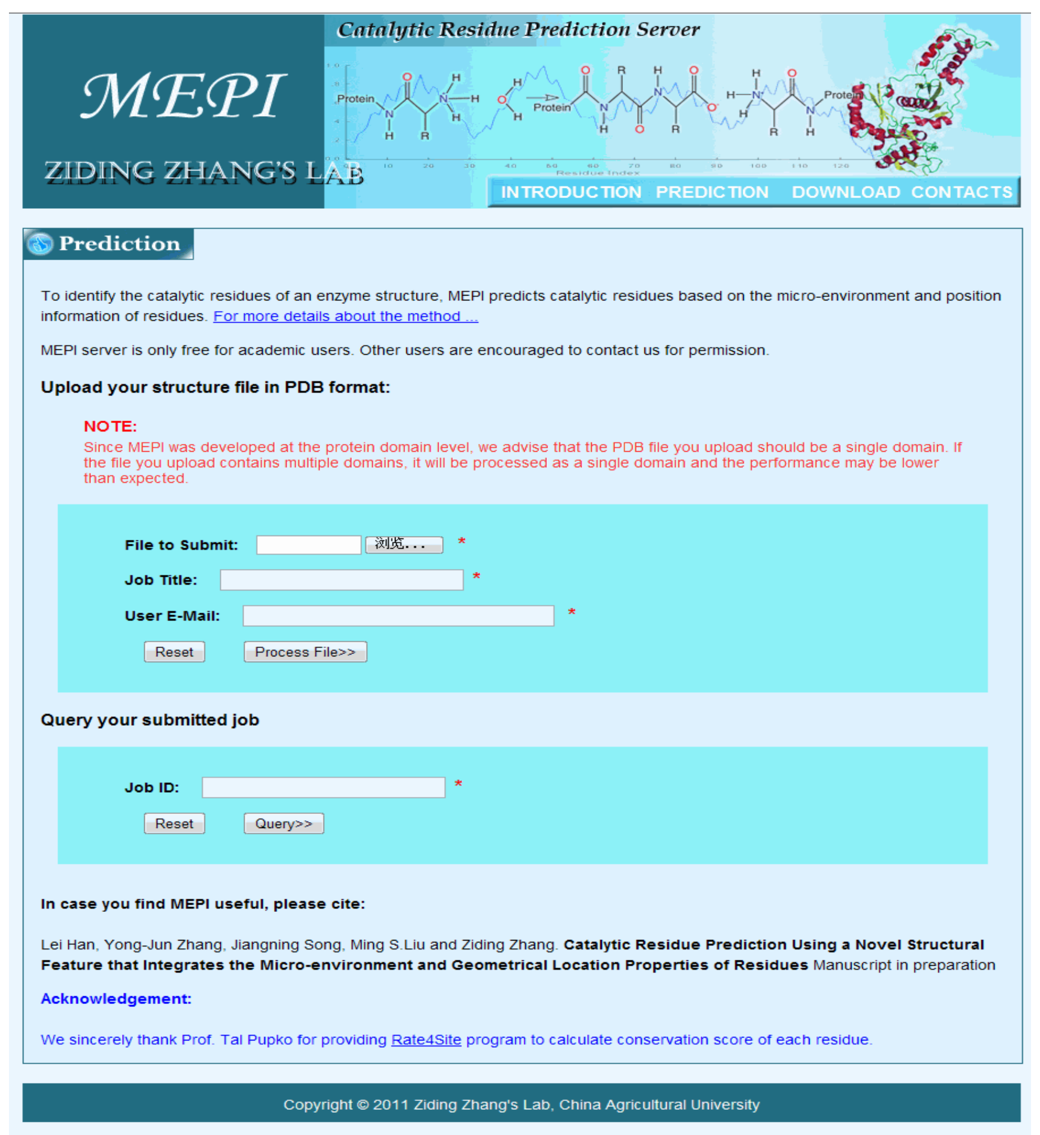


As shown in the left figure, different residue pairs exhibit the scaled propensities in the ME of catalytic residues, providing important insights into the molecular mechanism of enzymatic catalysis.

Performance comparison of each feature



Web server for CRs prediction (http://protein.cau.edu.cn/mepi/)


Catalytic Residue Prediction Server
MEPI
ZIDING ZHANG'S LAB
INTRODUCTION PREDICTION DOWNLOAD CONTACTS
Prediction
浏览...
Reset
Process File>>
Reset
Query>>

To identify the catalytic residues of an enzyme structure, MEPI predicts catalytic residues based on the micro-environment and position information of residues. For more details about the method ...

MEPI server is only free for academic users. Other users are encouraged to contact us for permission.

Upload your structure file in PDB format:

NOTE:
Since MEPI was developed at the protein domain level, we advise that the PDB file you upload should be a single domain. If the file you upload contains multiple domains, it will be processed as a single domain and the performance may be lower than expected.

File to Submit: *
Job Title: *
User E-Mail: *

Query your submitted job

Job ID: *

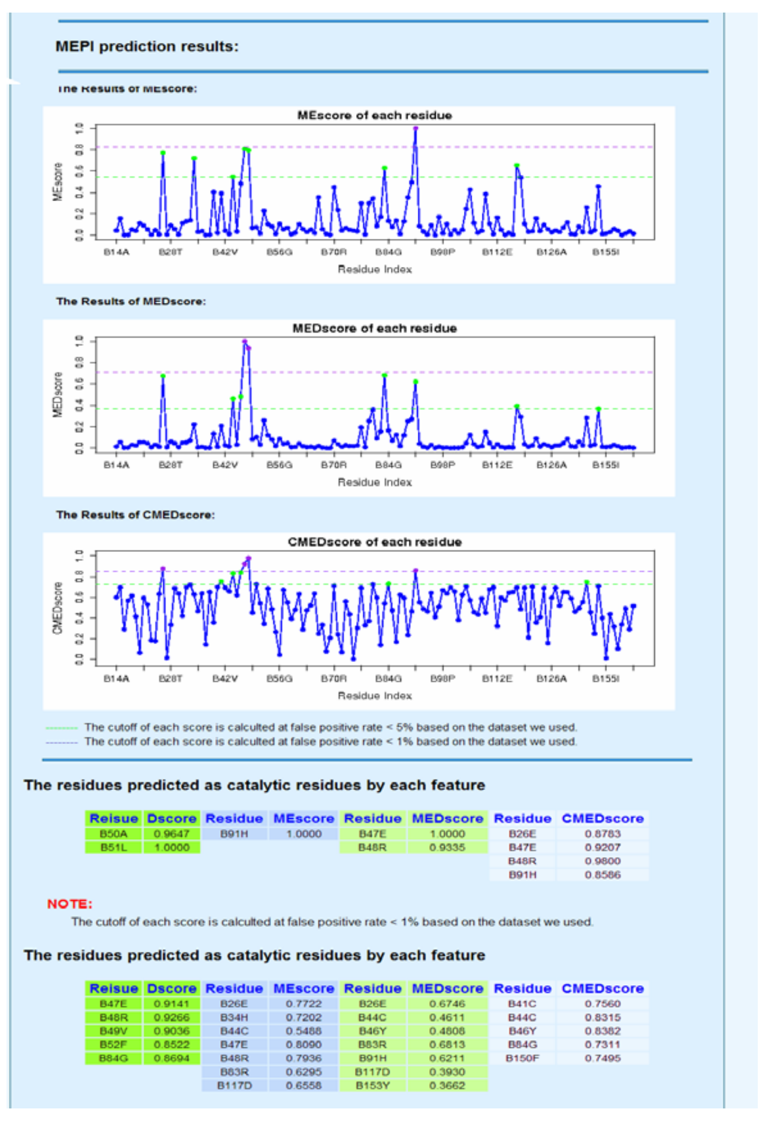
In case you find MEPI useful, please cite:

Lei Han, Yong-Jun Zhang, Jiangning Song, Ming S.Liu and Ziding Zhang. **Catalytic Residue Prediction Using a Novel Structural Feature that Integrates the Micro-environment and Geometrical Location Properties of Residues** Manuscript in preparation

Acknowledgement:

We sincerely thank Prof. Tal Pupko for providing Rate4Site program to calculate conservation score of each residue.

Copyright © 2011 Ziding Zhang's Lab, China Agricultural University



MEPI prediction results:

The Results of MEscore:

The Results of MEDscore:

The Results of CMEDscore:

The cutoff of each score is calculted at false positive rate < 5% based on the dataset we used.
The cutoff of each score is calculted at false positive rate < 1% based on the dataset we used.

The residues predicted as catalytic residues by each feature

Reisue	Dscore	Residue	MEscore	Residue	MEDscore	Residue	CMEDscore
B50A	0.9647	B91H	1.0000	B47E	1.0000	B26E	0.8783
B51L	1.0000			B48R	0.9335	B47E	0.9207
						B48R	0.9800
						B91H	0.8586

NOTE:
The cutoff of each score is calculted at false positive rate < 1% based on the dataset we used.

The residues predicted as catalytic residues by each feature

Reisue	Dscore	Residue	MEscore	Residue	MEDscore	Residue	CMEDscore
B47E	0.9141	B26E	0.7722	B26E	0.6746	B41C	0.7560
B48R	0.9266	B34H	0.7202	B44C	0.4611	B44C	0.8315
B49V	0.9036	B44C	0.5488	B46Y	0.4808	B46Y	0.8382
B52F	0.8522	B47E	0.8090	B83R	0.6813	B84G	0.7311
B84G	0.8694	B48R	0.7936	B91H	0.6211	B150F	0.7495
		B83R	0.6295	B117D	0.3930		
		B117D	0.6558	B153Y	0.3662		

Acknowledgements

The authors thank Dr. Zhi-Ping Feng at the Walter and Eliza Hall Institute of Medical Research (WEHI), Australia for helpful discussions. We also thank Prof. Tal Pupko (Tel Aviv University) for providing the Rate4Site program. This work was supported by grants from the National Key Basic Research Program of China (2009CB918802 and 2012CB1141004) and the National Natural Science Foundation of China (30700137).