Mol Divers (2008) 12:131–137
DOI 10.1007/s11030-008-9085-9

FULL LENGTH PAPER

# Prediction of compounds' biological function (metabolic pathways) based on functional group composition

**Yu-Dong Cai · Ziliang Qian · Lin Lu · Kai-Yan Feng · Xin Meng · Bing Niu · Guo-Dong Zhao · Wen-Cong Lu**

Received: 22 May 2008 / Accepted: 24 July 2008 / Published online: 14 August 2008

**Abstract** Efficient in silico screening approaches may provide valuable hints on biological functions of the compound-candidates, which could help to screen functional compounds either in basic researches on metabolic pathways or drug discovery. Here, we introduce a machine learning method (Nearest Neighbor Algorithm) based on functional group composition of compounds to the analysis of metabolic pathways. This method can quickly map small chemical molecules to the metabolic pathway that they likely belong to. A set of 2,764 compounds from 11 major classes of metabolic pathways were selected for study. The overall prediction rate reached 73.3%, indicating that functional group composition of compounds was really related to their biological metabolic functions.

**Keywords** Compound · Biological functions · Nearest neighbor algorithm · Functional group composition · Metabolic pathway

## Introduction

Understanding the relationships between human gene and diseases is a fundamental problem in the post genomic era. In order to improve the understanding of disease process, many "omics " sciences have been applied to provide useful biological information for researchers, including genomics, transcriptomics, proteomics and metabonomics. Among all "omics" sciences, metabonomics has been highlighted from the viewpoint of system biology, because metabonomics aims to investigate a biological system as a whole instead of some independent levels [1,2]. Metabonomics deals with the profiles of metabolites of integrated living systems [3]. Since living systems are dynamic, multivariate modulated and non-linear, analyzing metabolites is regarded as efficient way to release information about interactions between different components within high complex living systems such as humans [4].

As an essential part of the metabonomics analysis, metabolomics [2] focuses on all the metabolites with low molecular mass. Correctly and efficiently mapping those small molecules with great biological significance into to their corresponding metabolic pathways may have a positive

**Electronic supplementary material** The online version of this article (doi:10.1007/s11030-008-9085-9) contains supplementary material, which is available to authorized users.

Y.-D. Cai (✉) · X. Meng · G.-D. Zhao
CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China
e-mail: cyd@picb.ac.cn

Y.-D. Cai · K.-Y. Feng
Department of Mathematics, University of Manchester, Institute of Science and Technology, P.O. Box 88, Manchester M60 1QD, UK

Z. Qian
Graduate School of the Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100039, People's Republic of China

Z. Qian
Bioinformatics Center, Key Lab of Molecular Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, People's Republic of China

L. Lu
Department of Biomedical Engineering, Shanghai Jiaotong University, Shanghai 200240, People's Republic of China

B. Niu · W.-C. Lu
School of Materials Science and Engineering, Shanghai University, 149 Yan-Chang Road, Shanghai 200072, People's Republic of China

effect on further metabonomics analysis. Hence, in recent years, small molecule has become one of the top stars in the metabolic pathway analysis [5]. In this paper, efficient in-silico screening approaches could provide valuable hints on biological functions of the compound-candidates, which will help the screening. As an investigation, for the 11 major classes of metabolic pathways [6], this contribution will focus on predicting in which pathways the small molecules (compounds) participate.

A metabolic pathway is composed of a series of coupled, interconnecting chemical reactions. In the recent decades, various methods [7] have been employed to analyze to role of small molecule in metabolic pathways. However, most of the methods are on the basis of biochemical or physical experiments, which lead to the problem that the speed of annotations for new small molecules always lags behind that of the discovery, as more and more high-throughput equipments are being applied (high resolution mass spectrometry, etc.). Fortunately, machine learning method is good at dealing with such problems, which is fast, automated and meets the need of high-throughput data processing. Also, as an essential part of the bioinformatics, machine learning method has proven to be useful in many domains of biological analysis.

How to convert a sample into a numeric vector which is compatible for computer programs processing is a key step to the application of machine learning method. In our work, functional groups [8] were used to code training samples. Each small molecule was converted into a numeric vector as input for machine learning program. There are two merits of using functional groups for coding the small molecular. Firstly, it is much closer to the language used normally by chemists, meaning that most of the information guiding a small molecule into a certain metabolic pathway may hide in that numeric vector. Secondly, it can largely reduce the complexity of machine learning. Therefore, using functional groups to code small molecule is under the considerations of both biochemical background and mathematical computation. In this study, 28 main functional groups in organic chemistry are collected [8]. They are halogen, alcohol, aldehyde, amide, amine, hydroxamic_acid, phosphorus, carboxylate, ester, ether, imine, ketone, methyl, nitro, thiol, sulfonic_acid, sulfoxide, sulfone, sulfonamide, sulfide, ar_5c_ring, ar_6c_ring, non_ar_5c_ring, non_ar_6c_ring, hetero_ar_5c_ring, hetero_ar_6c_ring, hetero_non_ar_5c_ring and hetero_non_ar_6c_ring (Table 1). The concurrency of each of 28 function groups for each compound is used to capture the major chemical property. The same as the domain composition method we have ever used [9,10], this representation approach will facilitate the establishment of feasible computational approach.

Then, does each metabolic pathway have its own featured functional group composition(s)? If it is true, for a given compound, it will be possible to build a feasible predictor to predict the pathways where it participates. In this contribution, the function group composition for each of the 11 major classes of metabolic pathways [6] are calculated (Table 3). The results show the distinguish distribution of function group composition. Then we built predictor based on the Nearest Neighbor Algorithm [11] to classify the compounds into 11 major classes of metabolic pathways. Jackknife cross validation test on the predictor reached 73.3%, which is acceptable.

## Materials and methods

### Numeric coding system for compounds

The small chemical molecules were collected from public available database KEGG compound [ftp://ftp.genome.jp/pub/kegg/release/archive/kegg/42/ligand.tar.gz] release 42.0 [6]. The original dataset contains 14,229 chemical compounds. In order to extract the metabolic pathway information that is strongly related to the biological function of the chemical compounds, the 11 types of the pathways

1. Carbohydrate metabolism
2. Energy metabolism
3. Lipid metabolism
4. Nucleotide metabolism
5. Amino acid metabolism
6. Metabolism of other amino acids
7. Glycan biosynthesis and metabolism
8. Biosynthesis of polyketides and nonribosomal peptides
9. Metabolism of cofactors and vitamins
10. Biosynthesis of secondary metabolites
11. Xenobiotics biodegradation and metabolism

are also downloaded from KEGG pathway database [ftp://ftp.genome.jp/pub/kegg/release/archive/kegg/42/pathway.tar.gz] release 42.0 [6]. After deleting the “multi-function” compounds that participate in more than one metabolic pathway, and those non-informative molecules not existing in any metabolic pathways, 2,764 compounds were obtained for our research dataset (see Table 2 and the supplement material A for details).

### Functional group vector

In order to build a feasible computational method, a small molecule should be represented by a set of numerical values. Similar to functional domain composition for protein sequence analyses [9,10,12,13], we try to identify major functional groups [8] for each small molecule by our in-house C++ program [http://pcal.biosino.org/fc_analyzer.html]. As a result, 28 main functional groups [8] can be identified

**Table 1** Twenty-eight functional groups used to represent small molecules

| General feature | Key group | | | |
|---|---|---|---|---|
| Two dimensional structure | Alcohol | Aldehyde | Amide | Amine |
| | Hydroamic_acid | Phosphorus | Carboxylate | Methyl |
| | Ester | Ether | Imine | Ketone |
| | Nitro | Halogen | Thiol | Sulfonic_acid |
| | Sulfone | Sulfonamide | Sulfoxide | Sulfide |
| Cycle two dimensional structure | ar_5c_ring | ar_6c_ring | Non_ar_5c_ring | Non_ar_6c_ring |
| | hetero_ar_6_ring | hetero_non_ar_5_ring | hetero_non_ar_6_ring | hetero_ar_5_ring |

**Table 2** Distribution of compounds in 11 classes of metabolic pathways

| No | Pathway | Number |
|---|---|---|
| 1 | Carbohydrate metabolism | 321 |
| 2 | Energy metabolism | 35 |
| 3 | Lipid metabolism | 413 |
| 4 | Nucleotide metabolism | 100 |
| 5 | Amino acid metabolism | 375 |
| 6 | Metabolism of other amino acids | 94 |
| 7 | Glycan biosynthesis and metabolism | 52 |
| 8 | Biosynthesis of polyketides and nonribosomal peptides | 245 |
| 9 | Metabolism of cofactors and vitamins | 218 |
| 10 | Biosynthesis of secondary metabolites | 456 |
| 11 | Xenobiotics biodegradation and metabolism | 455 |
| | Total number | 2,764 |

(Table 1) and used to represent the sample of a compound (substrate or product); i.e.,

$$S = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_i \\ \vdots \\ g_{28} \end{pmatrix} \tag{1}$$

where $g_i$ is the occurrence number of the $i$th functional group in Table 1 in the concerned compound.

The detailed functional group vectors for the 2,764 compounds can be found in supplement material B. And the computational program fc_analyzer can be downloaded from our website http://pcal.biosino.org/fc_analyzer.html for free.

## The nearest neighbor algorithm

In our research, NNA (Nearest Neighbor Algorithm) [11] was adopted to predict the category of querying compound. It predicts the querying compound into the category as that of its nearest neighbor in the training dataset $s \in \{s_1, s_2, \ldots, s_N\}$. It works well especially in the situation with extremely large dimensional. In this contribution, the querying small molecule $q$ was compared to each of molecule of the training dataset using the following distance measurement

$$d(q, s) = 1 - \frac{q \cdot s}{\|q\| \cdot \|s\|} \tag{2}$$

where $q \cdot s$ is dot product of vector $q$ and $s$, $\|q\|$ is modulus of, and $\|s\|$ is modulus of $s$. The smaller of $d(q, s)$, the more similar between querying sample $q$ and $s$. Especially, when $d(q, s) = 0$, they are identical. The nearest neighbor of $q$ in the training set, can be identified by

$$d(q, s_1) = \min\{d(q, s_1), d(q, s_2), \ldots, \\ \times d(q, s_i), \ldots, d(q, s_N)\} \tag{3}$$

If there is a tie, saying there are two molecule in the training dataset, $d(q, s_i)$ and $d(q, s_j)$, both are the nearest neighbor of querying molecule $q$, we randomly select one of them as the nearest neighbor. After we got the nearest neighbor, the querying sample $q$ is predicted to the category of its nearest neighbor $s_i$. The high performance multi-threading NNA implementation can also be downloaded from our webpage [http://pcal.biosino.org/NNA.html].

The NNA has also been used in the previous work, such as predicting protein subcellular localization [14], transcription factor DNA-binding preference [15] and achieve good performance.

## Results and discussion

The computation was performed on a Dell Optiplex 260 machine with an a Intel 2.6 GHZ CPU and 2G RAM.

According the function group based coding procedure mentioned above, each of the 2,764 small molecules can be represented with a 28 dimensional vector, of which each component indicates the occurrence of each type of the 28 functional groups. The numeric vectors can then be used as input of the Nearest Neighbor Algorithm and further be cataloged into one of the 11 major pathways (classes).

We first accumulated function group frequencies for each group. As shown in Table 3, each pathway also has its own characteristic functional group composition, for example, pathway 1 (carbohydrate metabolism) is specific in carboxylate and pathway 3 (lipid metabolism) is abundance both in sulfonamide and carboxylate. The similarity between each pair of pathways is measured by the cosine distance. The smaller the cosine distance, the lower similarity is between

**Table 3** Accumulated function group composition of 11 functional categories of metabolic pathways

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Halogen | 60 | 4 | 151 | 0 | 82 | 3 | 3 | 401 | 46 | 96 | 68 |
| Alcohol | 5 | 2 | 31 | 0 | 135 | 1 | 1 | 320 | 58 | 389 | 447 |
| Aldehyde | 183 | 12 | 139 | 19 | 334 | 64 | 48 | 44 | 337 | 117 | 248 |
| Amide | 127 | 11 | 241 | 139 | 189 | 20 | 15 | 91 | 332 | 304 | 94 |
| Amine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hydroxamic_acid | 12 | 1 | 11 | 2 | 39 | 4 | 1 | 6 | 16 | 37 | 33 |
| Phosphorus | 0 | 0 | 160 | 0 | 0 | 1 | 0 | 4 | 8 | 91 | 14 |
| Carboxylate | 1077 | 25 | 647 | 141 | 241 | 37 | 206 | 564 | 176 | 442 | 202 |
| Ester | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ether | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 5 | 241 |
| Imine | 24 | 0 | 386 | 0 | 6 | 3 | 0 | 6 | 1 | 215 | 29 |
| Ketone | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Methyl | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 14 |
| Nitro | 208 | 2 | 298 | 3 | 10 | 4 | 66 | 247 | 3 | 233 | 22 |
| Thiol | 64 | 4 | 76 | 79 | 166 | 7 | 10 | 390 | 192 | 584 | 362 |
| Sulfonic_acid | 228 | 10 | 176 | 121 | 108 | 12 | 72 | 63 | 72 | 77 | 101 |
| Sulfoxide | 107 | 8 | 112 | 114 | 113 | 33 | 160 | 202 | 178 | 97 | 118 |
| Sulfone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sulfonamide | 155 | 23 | 947 | 12 | 206 | 36 | 334 | 1107 | 660 | 844 | 185 |
| Sulfide | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ar_5c_ring | 4 | 2 | 2 | 2 | 16 | 12 | 1 | 0 | 12 | 0 | 10 |
| ar_6c_ring | 0 | 0 | 2 | 0 | 1 | 6 | 0 | 0 | 3 | 0 | 9 |
| non_ar_5c_ring | 29 | 0 | 59 | 0 | 22 | 8 | 19 | 121 | 28 | 121 | 20 |
| non_ar_6c_ring | 55 | 11 | 66 | 63 | 204 | 75 | 28 | 73 | 123 | 70 | 83 |
| hetero_ar_5c_ring | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hetero_ar_6c_ring | 21 | 4 | 65 | 2 | 41 | 6 | 0 | 23 | 15 | 39 | 63 |
| hetero_non_ar_5c_ring | 85 | 10 | 111 | 124 | 156 | 25 | 7 | 21 | 261 | 109 | 146 |
| hetero_non_ar_6c_ring | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 0 |

**Table 4** Cosine distances among 11 functional categories of metabolic pathways

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.213 | | | | | | | | | |
| 3 | 0.318 | 0.151 | | | | | | | | |
| 4 | 0.341 | 0.245 | 0.510 | | | | | | | |
| 5 | 0.375 | 0.111 | 0.346 | 0.266 | | | | | | |
| 6 | 0.481 | 0.188 | 0.447 | 0.377 | 0.116 | | | | | |
| 7 | 0.346 | 0.150 | 0.115 | 0.503 | 0.373 | 0.407 | | | | |
| 8 | 0.439 | 0.221 | 0.130 | 0.602 | 0.350 | 0.520 | 0.131 | | | |
| 9 | 0.557 | 0.135 | 0.212 | 0.398 | 0.151 | 0.267 | 0.213 | 0.218 | | |
| 10 | 0.462 | 0.235 | 0.146 | 0.487 | 0.269 | 0.494 | 0.241 | 0.0849 | 0.180 | |
| 11 | 0.564 | 0.383 | 0.519 | 0.464 | 0.214 | 0.461 | 0.552 | 0.375 | 0.371 | 0.245 |

the two different pathways. The cosine distance is defined as follow:

$$\cos\left(p_i, p_j\right) = \frac{p_i^{(1)}\cdot p_j^{(1)}+p_i^{(2)}\cdot p_j^{(2)}+\cdots+p_i^{(28)}\cdot p_j^{(28)}}{\sqrt{\left(p_i^{(1)}\right)^2+\left(p_i^{(2)}\right)_2+\cdots+\left(p_i^{(28)}\right)^2}\cdot\sqrt{\left(p_j^{(1)}\right)^2+\left(p_j^{(2)}\right)^2+\cdots+\left(p_j^{(28)}\right)^2}}$$

$$(i \in \{2, 3, \ldots, 11\},\ j \in \{1, 2, \ldots, 10\},\ i > j) \qquad (4)$$

where, $p_i^{(1)}$, $p_i^{(2)}$ and $p_i^{(28)}$ are the 1st, 2nd, and 28th function group frequency in $i$th pathway, $p_j^{(1)}$, $p_j^{(2)}$ and $p_j^{(28)}$ for $j$th pathway respectively.

When all of the 11 classes of pathways compares with each other, there are 55 different cosine distances to be calculated. The 55 different values are listed in Table 4. Figure 1 is the statistical distribution of the 55 values in Table 4. Firstly, we discretized the range of cosine distance into 20 bars with width of 0.05. Secondly, we counted the number of values that belong to the range of each bar. For example, the 1st bar is with the range of 0 to 0.05. When looking up Table 4, there is no value belonging to this rang, so the 1st bar is with frequency equaling to 0. For the 2nd bar, its range is 0.05 to 0.1. Only the cosine distance between the 8th and 10th pathways belongs to the range of the 2nd bar, which has the value of

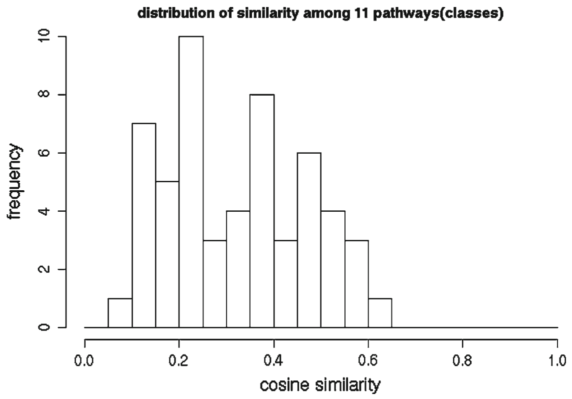

**Fig. 1** Distribution of the similarity among 11 pathways

0.0849. So that, the value of the 2nd bar is 1. The values of other bars are calculated as the above mean.

As shown in Table 4 and Fig. 1, most of the cosine value is less than 0.5, indicating low similarity of functional group composition among 11 classes of pathways. This means the functional group information of compound is really related to their metabolic pathways. Therefore, the following machine learning method based on functional group composition is reasonable.

Then, we developed a NNA predictor based on functional group composition. Jackknife cross-validation test, which has been used in many previous work [12–18] , was performed on the collected dataset to evaluate the performance of NNA predictor. In this contribution, jackknife test can be briefed as the following steps. For each of small molecule $s_i$ in the dataset, we find the nearest neighbor $s_k$ in the rest dataset excluding $s_i$, noted by $S\backslash s_i$. Then, $s_i$ is classified into one of the 11 types of the metabolic pathways as same as that of $s_k$. At last, the predicted pathway type of $s_i$ is compared to the truth. The prediction is success if the prediction agrees with the truth, otherwise failed. Table 5 gives the prediction accuracy on 11 pathways classes. The highest accuracy reached 91.4% for biosynthesis of polyketides and nonribosomal peptides, which is fairly good for a multi-classification problem. For pathways with too few samples, such as energy metabolism and metabolism of other amino acids, the prediction accuracy is not as good as others. The prediction performance is anticipated to be improved if more samples could be collected. Totally, the overall successful rate reaches 73.3%.

**Table 5** Performance of jackknife cross validation test on the multi-classifier

| No | Pathway | Success rate |
|---|---|---|
| 1 | Carbohydrate metabolism | 243/321 = 75.7% |
| 2 | Energy metabolism | 9/35 = 25.7% |
| 3 | Lipid metabolism | 342/413 = 82.8% |
| 4 | Nucleotide metabolism | 76/100 = 76.0% |
| 5 | Amino acid metabolism | 219/375 = 58.4% |
| 6 | Metabolism of other amino acids | 21/94 = 22.3% |
| 7 | Glycan biosynthesis and metabolism | 33/52 = 63.5% |
| 8 | Biosynthesis of polyketides and nonribosomal peptides | 224/245 = 91.4% |
| 9 | Metabolism of cofactors and vitamins | 142/218 = 65.1% |
| 10 | Biosynthesis of secondary metabolites | 370/456 = 81.1% |
| 11 | Xenobiotics biodegradation and metabolism | 347/455 = 76.3% |
|  | Overall | 2,026/2,764 = 73.3% |

The above result indicates that metabolic pathways really have their own featured functional group composition and introducing machine learning method (NNA) based on functional group composition to the analysis of metabolic pathways is feasible. That may be due to the fact that the reactions in metabolic pathways are relative “stable”, that is, the “near” molecules may belong to the same classes of metabolic pathways. Take the classical glycolysis pathway for example. This pathway is composed of ten-step sequential reactions, where one glucose is split and converted into two pyruvates (http://www.genome.jp/kegg/pathway/map/map00010.html) [19]. There are 11 compounds and 10 reactions that participate in the glycolysis pathway (See Table 6). As can be seen in Table 6, the functional group composition of most compounds are similar, especially compounds that are in the coterminous reactions.

However, as one kind of machine learning methods, Nearest Neighbor Algorithm can not avoid the limitation of machine learning methods either. The analysis based on Nearest Neighbor Algorithm is qualitative instead of quantitative, because Nearest Neighbor Algorithm can not reveal the relationship between metabolic pathways (target-variable) and functional group composition (feature-variable) without any apriori knowledge. Therefore, in our future work, we will develop wonderful computational methods to evolve our predictor into analyzer, which can extract some quantitative information from the interaction between target-variable and feature-variable.

## Conclusions

In this research, we try to predict the participating metabolic pathways of small chemical molecule by analyzing their function group information. The predictor achieved acceptable performance on the jackknife cross-validation test. Therefore, our predictor maybe used for finding the new compounds which participate in metabolic procedures, and furthermore for supporting the extending the unknown part of the metabolic networks.

**Acknowledgements** The work was funded by the Grant from Chinese Academy of Science: KSCX2-YW-R-112.

**Table 6** The 11 compounds in the glycolysis pathway

| Compound | Compound ID | Compounds' functional group composition (Halogen, alcohol, aldehyde, amide, amine, hydroxamic_acid, phosphorus, carboxylate, ester, ether, imine, ketone, methyl, nitro, thiol, sulfonic_acid, sulfoxide, sulfone, sulfonamide, sulfide, ar_5c _ring, ar_6c_ring, non_ar_5c_ring, non_ar_6c_ring, hetero_ar_5c_ring, hetero_ar_6c_ring, hetero_non_ar_5c_ring, hetero_non_ar_6c_ring) |
|---|---|---|
| D-Glucose | C00031 | 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 |
| | ↓Reaction 1 | |
| alpha-D-Glucose 6-phosphate | C00668 | 0 6 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 |
| | ↕Reaction 2 | |
| beta-D-Fructose 6-phosphate | C05345 | 0 3 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 |
| | ↓Reaction 3 | |
| beta-D-Fructose1,6-bisphosphate | C05378 | 0 2 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 |
| | ↕Reaction 4 | |
| D-Glyceraldehyde 3-phosphate | C00118 | 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↕Reaction 5 | |
| 3-Phospho-D-glyceroyl phosphate | C00236 | 0 1 0 0 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↕Reaction 6 | |
| 2,3-Bisphospho-D-glycerate | C01159 | 0 0 0 0 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↕Reaction 7 | |
| 3-Phospho-D-glycerate | C00197 | 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↕Reaction 8 | |
| 2-Phospho-D-glycerate | C00631 | 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↕Reaction 9 | |
| Phosphoenolpyruvate | C00074 | 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| | ↓Reaction 10 | |
| Pyruvate | C00022 | 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |

## References

1. Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov 1:153–161. doi:10.1038/nrd728
2. Nicholson JK, Wilson ID (2003) Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nat Rev Drug Discov 2:668–676. doi:10.1038/nrd1157
3. Nicholson JK, Lindon JC, Holmes E (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29: 1181–1189. doi:10.1080/004982599238047
4. Nicholson JK, Holmes E, Lindon JC, Wilson ID (2004) The challenges of modeling mammalian biocomplexity. Nat Biotechnol 22:1268–1274. doi:10.1038/nbt1015
5. Vastrik I, D'Eustachio P, Schmidt E, Joshi-Tope G, Gopinath G, Croft D et al (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8:R39. doi:10.1186/gb-2007-8-3-r39
6. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34(database issue): D354–D357
7. Burkart MD (2003) Metabolic engineering—a genetic toolbox for small molecule organic synthesis. Org Biomol Chem 1:1–4. doi:10.1039/b210173d
8. Marchand-Geneste N, Watson KA, Alsberg BK, King RD (2002) New approach to pharmacophore mapping and QSAR analysis using inductive logic programming. Application to thermolysin inhibitors and glycogen phosphorylase B inhibitors. J Med Chem 45:399–409. doi:10.1021/jm0155244
9. Cai YD, Chou KC (2005) Predicting enzyme subclass by functional domain composition and pseudo amino acid composition. J Proteome Res 4:967–971. doi:10.1021/pr0500399
10. Cai YD, Doig AJ (2004) Prediction of *Saccharomyces cerevisiae* protein functional class from functional domain composition. Bioinformatics 20:1292–1300. doi:10.1093/bioinformatics/bth085
11. Salzberg S, Cost S (1992) Predicting protein secondary structure with a nearest-neighbor algorithm. J Mol Biol 227:371–374. doi:10.1016/0022--2836(92)90892-N
12. Jia P, Qian Z, Zeng Z, Cai Y, Li Y (2007) Prediction of subcellular protein localization based on functional domain composition. Biochem Biophys Res Commun 357:366–370. doi:10.1016/j.bbrc.2007.03.139
13. Lu L, Qian Z, Cai YD, Li Y (2007) ECS: an automatic enzyme classifier based on functional domain composition. Comput Biol Chem 31:226–232. doi:10.1016/j.compbiolchem.2007.03.008
14. Chou KC, Cai YD (2005) Predicting protein localization in budding yeast. Bioinformatics 21:944–950. doi:10.1093/bioinformatics/bti104
15. Qian Z, Lu L, Liu X, Cai Y-D, Li Y (2007) An approach to predict transcription factor DNA binding site specificity based upon gene and transcription factor functional categorization. Bioinformatics 23:2449–2454. doi:10.1093/bioinformatics/btm348
16. Mondal S, Bhavna R, Mohan Babu R, Ramakumar S (2006) Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification. J Theor Biol 243:252–260. doi:10.1016/j.jtbi.2006.06.014
17. Wong YH, Lee TY, Liang HK, Huang CM, Wang TY, Yang YH, Chu CH, Huang HD, Ko MT, Hwang JK (2007) KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. Nucleic Acids Res 35(Web Server issue): W588–W594

18. Xie D, Li A, Wang M, Fan Z, Feng H (2005) LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. Nucleic Acids Res 33(Web Server issue): W105–W110

19. Trudy McKee JRM (1999) Biochemistry: an introduction. 2nd edn. McGraw-Hill Companies, Inc