

# AAindex: Amino Acid index database

Shuichi Kawashima\* and Minoru Kanehisa

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Received September 29, 1999; Accepted October 4, 1999

## ABSTRACT

**AAindex is a database of amino acid indices and amino acid mutation matrices. An amino acid index is a set of 20 numerical values representing various physico-chemical and biochemical properties of amino acids. An amino acid mutation matrix is generally  $20 \times 20$  numerical values representing similarity of amino acids. AAindex consists of two sections: AAindex1 for the collection of published amino acid indices and AAindex2 for the collection of published amino acid mutation matrices. Each entry of either AAindex1 or AAindex2 consists of the definition, the reference information, a list of related entries in terms of the correlation coefficient and the actual data. The database may be accessed through the DBGET/LinkDB system at GenomeNet (<http://www.genome.ad.jp/aaindex/>) or may be downloaded by anonymous FTP (<ftp://ftp.genome.ad.jp/db/genomenet/aaindex/>).**

## INTRODUCTION

Each of the 20 amino acids has multifaceted properties that are responsible for the specificity and diversity of protein structure and function. A large body of experimental and theoretical research has been performed to characterize different kinds of properties of individual amino acids and to represent them in terms of the numerical index. In 1988 Nakai *et al.* collected 222 amino acid indices from research papers and investigated their interrelationships by the hierarchical cluster analysis (1). They identified four major clusters: (i)  $\alpha$ -helix and turn propensities, (ii)  $\beta$ -strand propensity, (iii) hydrophobicity and (v) physicochemical properties. Tomii and Kanehisa (2) increased the size of the collection to include 402 indices and re-performed the clustering. The result was generally in good agreement with the previous work. In addition to amino acid indices, they also collected 42 amino acid mutation matrices from the literature. The AAindex database is based on these works first developed by Nakai *et al.* and then extended by Tomii and Kanehisa. The AAindex database is continuously updated by the present authors (3).

## THE CURRENT DATABASE

The AAindex database can be retrieved by the DBGET/LinkDB

system (4) of the Japanese GenomeNet service (5) at <http://www.genome.ad.jp/aaindex/>. The DBGET/LinkDB system integrates most of the major molecular biology databases and is especially suited for using hyperlinks to related entries within the AAindex database as well as to other databases. The AAindex database consists of two sections: AAindex1 for amino acid indices and AAindex2 for amino acid mutation matrices. The content of the two sections is as follows.

### AAindex1

The AAindex1 section currently contains 437 amino acid indices. Each entry consists of an accession number, a short description on the index, the reference information and the numerical values for the property of 20 amino acids. In addition, it contains neighbor information, namely, cross-links to similar indices with an absolute value of the correlation coefficient of  $\geq 0.8$ .

### AAindex2

The AAindex2 section currently contains 71 amino acid mutation matrices: 52 symmetric matrices and 19 non-symmetric matrices. The format of the entry is almost the same as that of AAindex1 except that it contains 210 numerical values (20 diagonal and  $20 \times 19/2$  off-diagonal elements) for a symmetric matrix or 400 or more numerical values for a non-symmetric matrix (some matrices include a gap or distinguish two states of cysteine).

## ACKNOWLEDGEMENTS

We thank Drs Kenta Nakai and Kentaro Tomii for the initial developments of the AAindex database. This work was supported in part by a Grant-in-Aid for Scientific Research on the Priority Area 'Genome Science' from the Ministry of Education, Science, Sports and Culture of Japan. The computational resource was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

## REFERENCES

1. Nakai, K., Kidera, A. and Kanehisa, M. (1988) *Protein Eng.*, **2**, 93–100.
2. Tomii, K. and Kanehisa, M. (1996) *Protein Eng.*, **9**, 27–36.
3. Kawashima, S. and Kanehisa, M. (1999) *Nucleic Acids Res.*, **27**, 368–369.
4. Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y. and Kanehisa, M. (1998) *Pacific Symp. Biocomput.* 1998, 683–694.
5. Kanehisa, M. (1997) *Trends Biochem. Sci.*, **22**, 442–444.

\*To whom correspondence should be addressed. Tel: +81 774 38 3270; Fax: +81 774 38 3269; Email: [shuichi@kuicr.kyoto-u.ac.jp](mailto:shuichi@kuicr.kyoto-u.ac.jp)