BIOINFORMATICS APPLICATIONS NOTE Vol. 18 no. 1 2002
Pages 209–210

PCSB—a program collection for structural biology and biophysical chemistry

A. Hofmann and A. Wlodawer*

Macromolecular Crystallography Laboratory, NCI at Frederick, Frederick, MD 21702, USA

Received on April 24, 2001; revised on August 16, 2001; accepted on August 17, 2001

ABSTRACT

We present the first package of Java classes specifically aimed at the handling of structural and biophysical problems. To enable object-oriented programming a basis of fundamental Java classes is required which deals with basic operations of vectors, matrices, amino acid sequences, crystal symmetries and PDB files. Five classes, which carry out these basic operations, were constructed and bundled together with several utility functions in the PCSB package.

Furthermore, to demonstrate their applicability and to obtain programs handling common tasks in structural laboratories, we present the first six applications of PCSB. All applications are portable to different platforms and require only the Java Runtime Environment to be installed on the system.

Availability: http://www24.brinkster.com/hofmanna/pcsb/

Contact: hofmanna@ncifcrf.gov

Supplementary information: A manual for PCSB describing the Java classes as well as the applications is available as PDF file.

INTRODUCTION

Structural characterization of proteins very often involves a series of spectroscopic experiments that require the recording of 'baseline spectra' as controls, thus increasing the amount of data to be processed. In subsequent analyses, sample spectra have to be corrected against their respective controls and very often the conversion of raw data is necessary as well (for example, for Circular Dichroism (CD) data). Additionally, crystallographic problems frequently require the knowledge of certain parameters such as anomalous scattering factors, solvent content, or symmetry operators. A convenient lookup of these data on the available computer system is another task addressed here. At the same time, modern structural laboratories present multiple computing environments thereby setting a demand for cross-platform tools.

*To whom correspondence should be addressed.

To exploit the entire potential of Java (Sun Microsystems, 2000) we decided to generate fundamental classes (similar to the Java Foundation Classes), which enable the handling of objects within structural biology and biophysics. These classes will simplify algorithm construction for our applications and can also be employed by other groups designing their own programs. The second aspect deals with the applications themselves where a user-friendly design and the addressing of specific 'structural' questions was in focus.

THE CLASSES OF THE PACKAGE PCSB

Three classes provide the mathematical background for geometric problems: Vector, Matrix and Eigenvalue Decomposition. Methods of the class Vector include vector normalization, the calculation of the distance between two points (absolute value of the difference vector) and others. The class Matrix provides methods for the most common matrix operations, among them an inversion procedure for symmetrical matrices by Gauss–Jordan elimination and methods for matrix–vector and vector–matrix multiplication. The class Eigenvalue Decomposition was adapted from the JAMA package (Hicklin *et al.*, 1999) in unchanged form.

The class PDB file defines an object from PDB files and provides methods for all common operations within PDB files. These include read and write methods, coordinate transformation by rotation or translation, fractionalization of orthogonal coordinates and others. Similarly, the class SymLib provides all necessary information about the 230 crystal symmetry classes. SymLib was constructed with the help of the CCP4 symmetry library (CCP4, 1996).

For definition of amino acid sequence objects the class AASequence contains all required features. An object is established simply from a string encoding the sequence. Non-amino acid characters like numbers, spaces or other symbols are filtered out. Sequences can be read in or exported in one- or three-letter coding.

The package PCSB. Utility comprises several classes supporting the applications with commonly used operations. Among others, this package provides classes for

printing, status bar handling and variable conversions (double to int, Integer to double, etc.).

THE APPLICATIONS

Peptides (v2.4) is a small program for calculation of peptide/protein parameters from a given sequence in either the one- or three-letter code. It calculates the number of residues, molecular mass, molecular formula, pl (pK_a and pK_b values from the *Handbook of Chemistry and Physics*, 1985), molar extinction coefficient (Gill and von Hippel, 1989), amino acid content, and a pH–charge correlation.

ACDP (v2.1) and AFDP (v2.1) stand for Automated CD/Fluorescence Data Processing and share the layout and program philosophy of automated baseline correction from a series of CD or fluorescence spectra. An ASCII file for each sample spectrum is written out, containing the original and derived data. Two wavelengths can be selected to be monitored in each spectrum and their molar ellipticities (ACDP) or fluorescence intensities (AFDP) as well as their ratios are tabulated in the results file. Filters are implemented to read in data files from AVIV or JASCO instruments (ACDP) and in case of AFDP, files from FLWinLab (Version 1.10, Perkin–Elmer) and DataMax (Version 2.1, Fluoromax) are recognized.

Align (v1.0) allows the structural alignment of two molecules by calculating the superposition operator (Kabsch, 1976) from either all, the backbone or the Cα atoms in orthogonal and fractional coordinates contained in two PDB files (see Figure 1). A radial distance criterion can be applied which excludes atom pairs being too distant from each other, and a histogram panel provides graphical representation of the positional and *B*-factor displacement as a function of the individual atoms.

Given the PDB coordinates from N molecules related by a proper rotation, the program Symmetry (v2.0) calculates parameters of the N-fold rotation axis, like the centroid of the rotation, the director of the axis and the average rotation angle. The angles of the rotation axis against the *x*-, *y*- and *z*-axis of the coordinate system are also calculated.

The program Tables (v1.3) is a 'library' program and can perform three different tasks: look-up of anomalous scattering factors, display of symmetry operators of a given spacegroup and calculation of the crystal solvent content (Matthews' coefficient; Matthews, 1968).

CONCLUSION

We present the first Java implementation of algorithms related to structural biology and biophysical chemistry in fundamental Java classes. The package PCSB provides mathematical algorithms for vector and matrix operations as well as the handling of PDB files, symmetry operations and amino acid sequences. Updates as well as new releases will also be described on the web. The package PCSB as well as the applications are freely available from the authors together with a detailed manual.

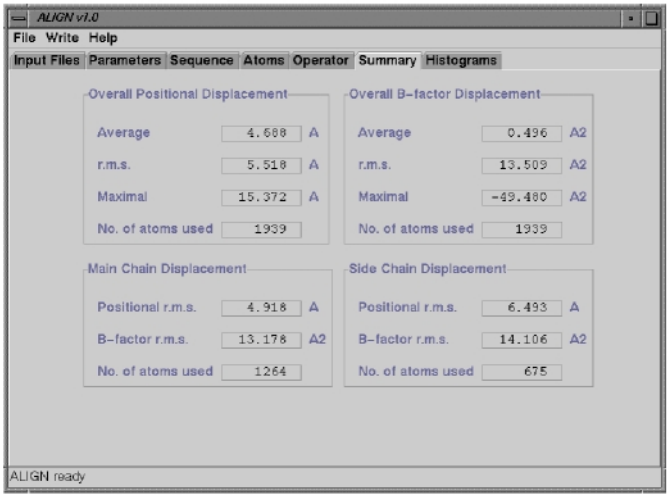

Fig. 1. Screen shot of the summary section of the program Align. The coordinates of the second (aligned) molecule can be written out as a PDB file. All results can be printed or saved as an ASCII file.

ACKNOWLEDGEMENTS

This project was inspired by the achievements of Hans-Georg Beisel and Andreas Bergner.

REFERENCES

CCP4 (1996) Collaborative Computational Project, Number 4; the CCP4 suite: programs for protein crystallography (vs. 3.0). *Acta Cryst.*, D **50**, 760–763.

Gill,S.C. and von Hippel,P.H. (1989) Calculation of protein extinction coefficients form amino acid sequence data. *Anal. Biochem.*, **182**, 319. (1990) Erratum. *Anal. Biochem.*, **189**, 283.

Weast,R.C., Astle,M.J. and Beyer,W.H. (eds) (1985) *Handbook of Chemistry and Physics*, 66th edn, CRC Press, Boca Raton, FL.

Hicklin,J., Moler,C., Webb,P., Boisvert,R.F., Miller,B., Pozo,R. and Remington,K. (1999) *JAMA: a Java Matrix Package*. http://math.nist.gov/javanumerics/jama/.

Kabsch,W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A **32**, 922–923.

Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.

Sun Microsystems (2000) Java 2 Platform Standard Edition Version 1.3. http://java.sun.com/j2se/1.3/.