

PACKAGE ID - 001312IBMPC00 DNA BP CALLING SW

KWIC TITLE - Automated DNA Base Pair Calling Algorithm

AUTHORS - Yeung, E.S.
Ames Laboratory, IA (United States)

Hall, D.R.
Ames Laboratory, IA (United States)

LIMITATION CODE -COPY **AUDIENCE CODE** - LIM

COMPLETION DATE - 06/12/1996 **PUBLICATION DATE** - 06/12/1999

DESCRIPTION - The procedure solves the problem of calling the DNA base pair sequence from two channel electropherogram separations in an automated fashion. The core of the program involves a peak picking algorithm based upon first, second, and third derivative spectra for each electropherogram channel, signal levels as a function of time, peak spacing, base pair signal to noise sequence patterns, frequency vs ratio of the two channel histograms, and confidence levels generated during the run. The ratios of the two channels at peak centers can be used to accurately and reproducibly determine the base pair sequence. A further enhancement is a novel Gaussian deconvolution used to determine the peak heights used in generating the ratio.

PACKAGE CONTENTS - Media Directory; Software Abstract; Code Outline;
Media Includes Source Code;

SOURCE CODE INCLUDED? - Yes

MEDIA QUANTITY - 1 3.5 Diskette

METHOD OF SOLUTION - Two-base-calling schemes for DNA sequencing are evaluated. Both are based on data collected from two broad-band emission channels derived from either one or two-excitation channels. Standard 4-dye Sanger reaction products are used in conjunction with capillary electrophoretic separation in a polymer matrix. Data acquisition is compatible with high light-throughout imaging and minimal data storage. In one scheme, commercial chromatographic software provides peak recognition and peak heights. The peak height ratios from the two channels provide base calling accuracies of 99.3% and 97.1% through 330 bp and 350 bp, respectively. In another scheme, ratiograms are derived from the two channels. The resulting step-like functions permit calling bases even when successive peaks are not resolved. The base calling accuracy is 99% through 340 bp. Because of simplicity in implementation, either scheme should be readily applicable to high speed high throughput DNA sequencing in capillary arrays.

COMPUTER - IBM PC

PACKAGE ID - 001312IBMPC00 DNA BP CALLING SW

OPERATING SYSTEMS - Windows 3.1 or above

PROGRAMMING LANGUAGES - C++

SOFTWARE LIMITATIONS - Must be compiled before use.

SOURCE CODE AVAILABLE (Y/N) - Y

UNIQUE FEATURES - The procedure solves the problem of calling the DNA base pair sequence from two channel electropherogram separations in an automated fashion.

HARDWARE REQS - Computer capable of operating Windows 3.1 or above.

ABSTRACT STATUS - Released AS-IS 10/12/1999.

SUBJECT CLASS CODE - Z

KEYWORDS -

COMPUTER PROGRAM DOCUMENTATION
D CODES
DATA
DNA

EDB SUBJECT CATEGORIES -
990200

SPONSOR - DOE/DP

PACKAGE TYPE - AS - IS