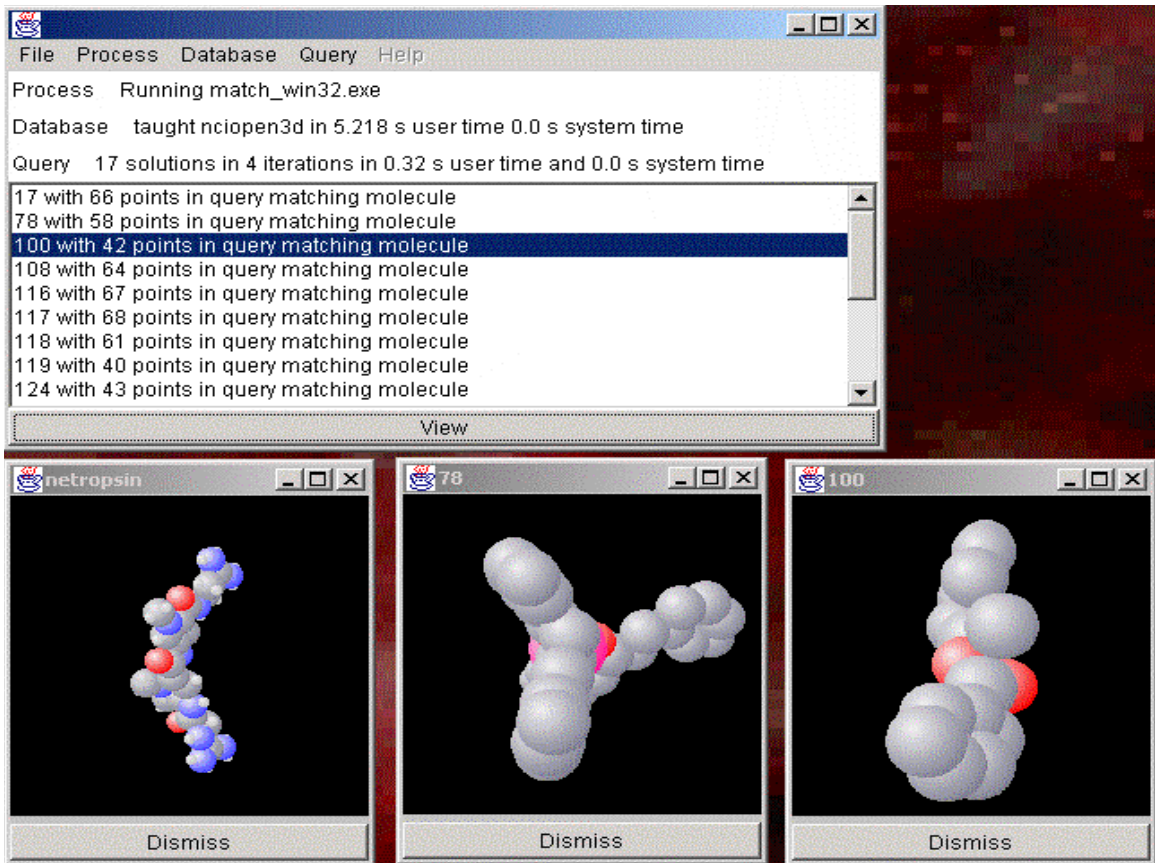
CYBULA
high performance pattern recognition systems

Molecular Database Systems - Chemical Structure Search (AURA-MOL)

Introduction

Cybula Ltd has developed a powerful range of tools based on the **AURA** high performance pattern recognition technology for searching databases containing 3 Dimensional views of complex small molecules (in the order of 60 atoms). The AURA-MOL technology allows a user to search for molecules that have a similar shape to a particular description. The basic technology underpinning this task can be used in a wide variety of problems. Extensions to the basic method allow the system to take into account local properties of the molecules.

The image below shows an example screen shot from a demonstration front end to the system. 200 molecules have been taught from the NCI open database. A query 'netropsin' has been entered and 17 potential matches found in 0.32 seconds (on a 800MHz PC) as displayed in the list. Two example matches have been displayed, along with the query.



Principal features of the technology are:

- The technology allows very large databases of molecules (> 100,000) to be searched.
- Input of new molecules to the database is quick.
- The methods can be run on desktop workstations to supercomputers matching the needs of the user.
- The technology uses the full 3D structure and surface properties of the molecules
- The methods used within the technology are published, allowing full understanding of the methods

The technology consists of a set of C++ functions built on top of the AURA CMM library used in many of our systems. Typical use is through UNIX/linux based programs. A simple JAVA based demonstration system is available to show the operation of the system.

Outline function

The AURA-MOL system describes the surface of the molecule by a set of points. These points are joined in a graph that is then used to search the database of molecules. The nodes in the graph contain attributes that describe the local properties of the molecule at that point. The match engine is composed of a number of CMMs working together through a constraint propagation process. The constraint update procedure has been developed specifically to support CMM based systems, and efficiently searches large databases for potential matches. The results of the process are then supplied to the user with a measure of the similarity to each molecule returned.

Technical details

The AURA-MOL system has been developed at the University of York, Computer Science Department, UK in association with GSK over the last 5 years. The technology is described in detail in a number of papers at the Advanced Computer Architectures web site.

Use of the technology

The AURA-MOL system may be embedded into many applications. The system exists as a C++ library and runs on Linux, NT, Windows 2000 as well as SGI Irix. The system can be run on small PCs to supercomputers. The technology has been licensed to Cybula Ltd and is now available for incorporation into your systems.

For further details contact enquiries@cybula.co.uk. US offices opening shortly.

www.cybula.com

Head Office: Cybula Ltd, IT Centre, York Science Park, Hesligton,
York YO10 5DG, Phone +44(0)1904 567686 Fax +44(0)1904 567685
E-mail: enquiries@cybula.com