CAVEAT: 3D Database For Designing Molecules

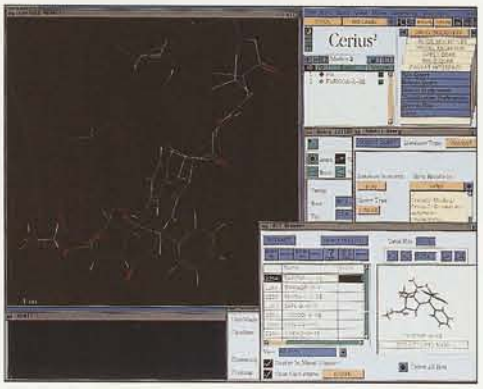
by Dr. Georges Lauri, University of California at Berkeley

CAVEAT is an interactive 3D database search program developed specifically to help organic chemists design molecules. This program was originated in the research group of Professor Paul A. Bartlett at UC Berkeley. While other programs seek to match pharmacophores in their search and to identify intact molecules for immediate testing, CAVEAT aids chemists in devising entirely new structures.

CAVEAT does this by finding molecules and structural fragments with bonds that satisfy a specified geometric relationship. These bonds are treated as vectors, embodying not only the location of substituents but also their orientation. This powerful abstraction can be applied to a diverse array of problems.

CAVEAT identifies molecular frameworks that can serve as alternative skeletons for peptide or other mimetic targets. It can find parts of macrocyclic or acyclic structures that could serve as crosslinking units to constrain a flexible molecule. Or it can be used to uncover structures that could satisfy the hydrogen-bonding requirements of a binding site. In short, any search query that can be phrased as a 3D relationship between bonds can be tackled with CAVEAT.

The CAVEAT suite of programs also embodies sophisticated screening and clustering algorithms as the second tier of evaluation. The hits retrieved from the database search can be filtered according to a number of criteria including: structural complexity, the presence or absence of specific functionality, hybridization of the matching bonds and attachment centers, the quality of the conformational and steric fit between the fragment and the target molecule, and complementarity to an active site. CAVEAT also includes a uniquely powerful clustering routine, which classifies the acceptable hits according to additional geometric characteristics. By presenting only the best-matching structure from each cluster, CAVEAT helps the chemist survey the entire range of frameworks which satisfy the query.



C^2•CAVEAT search for linking fragment for Cyclosporin-A, identified linking unit in green; see D.G Alberg, S.L. Schreiber, "Structure-Based Design of a Cyclophilin-Calcineurin Bridging Ligand", Science, 1993 V.262 pp.248-250

What sets CAVEAT apart from other 3D database programs is its speed. A typical CAVEAT search on a desktop workstation has turn around times of 2-3 minutes, even in large (> 500,000 molecule) databases.

The Cerius² integration module for CAVEAT – C^2•CAVEAT – makes all of the capabilities of this program available within the Cerius² environment. Cerius² and CAVEAT provide an interactive solution to the design of novel bioactive compounds, with a fast database engine to reduce search time, a powerful classifier to reduce evaluation time, and a seamless interface to the rest of MSI's drug discovery capabilities for further stages in structure-based drug design. Call Dr. Scott Kahn at (408) 522-0100 for more information.

MODELER: Implementing 3D Protein Modeling

by Dr. Andrej Šali, Harvard University

Molecular biologists, structural biologists, and computational chemists frequently need an approximate 3D model for a given protein sequence. In many cases, a satisfactory model can be provided within hours by homology or comparative modeling[1] using MODELER. MODELER is a computer program that implements comparative protein modeling by satisfaction of spatial restraints[1-3]. The input to the program is an alignment of the sequence being predicted (target) with related known 3D structures (templates).

In homology modeling, a 3D model is derived from 3D structures of related proteins that have been determined by X-ray crystallography or NMR techniques. The usefulness of comparative modeling is increasing, largely because the genome projects are producing more sequences, and novel protein folds are being determined experimentally.

The output from MODELER, obtained without any user intervention, is a 3D model for the target sequence containing all mainchain and sidechain heavy atoms. First, MODELER derives distance, angle, and dihedral angle restraints on the target sequence from its alignment with template 3D structures. Second, the spatial restraints and energy terms enforcing proper stereochemistry are combined into an objective function. Third, the model is obtained by optimizing this objective function. This optimization is carried out employing methods of conjugate gradients and molecular dynamics with simulated annealing. The resulting model simultaneously satisfies the restraints and stereochemistry.

The preparation of the input alignment consists of a search for the template structures and their alignment with the target sequence. The accuracy of the different regions in the model can then be estimated from the remaining restraint violations. When the sequence identity between the target and template structures is higher than 40%, the model is equivalent in many respects to a model determined by medium resolution X-ray crystallography.

MODELER can create a model even with only one source protein. In this case, the structure for dihydrofolate reductase from Lactobacillus Casei is used to generate a model for the E. Coli protein. The model (yellow) is 2.2 A RMS deviation from the crystal structure of the E. Coli protein.

Spatial restraints derived from a number of different sources can be used together with homology-derived restraints. In this way, MODELER can improve a homology model by making it consistent with partial experimental data (e.g. NMR) on the target.

The combination of QUANTA and MODELER provides a complete system for creating a reliable homology model. QUANTA's interactive sequence alignment simplifies setup and adjustment of alignments. And its protein structural analysis tools and visualization provide assessment of the reliability of resulting models, making QUANTA an ideal complement to MODELER. MODELER is now available from MSI. Call (617) 229-9800 for details.

References

(1) *A. Sali and T.L. Blundell. J. Mol. Biol.* ***234****, 779–815, 1993.*
(2) *A. Sali and J.P. Overington. Prot. Sci.* ***3****, 1582–1596, 1994.*
(3) *A. Sali, R. Matsumoto, H.P. McNeil, M. Karplus, and R.L. Stevens. J. Biol. Chem.* ***268****, 9023–9034, 1993.*