# MOE

## Molecular Operating Environment



Browse: C:/ccg/work/0204/o/s.mdb

Subject: mol | 6CMG.G

✓ 3D: Atoms | Ribbon | Options | ? | ✓ 2D: Options | Keep

α6
α3
β4 β4 α3

None ▾ None ▾

Entry: Browse Only Selected | Select | Mark | None ▾

7

- ❏ **Molecular Modeling and Simulations**
- ❏ **Protein Modeling and Bioinformatics**
- ❏ **Structure/Fragment-Based Drug Design**
- ❏ **Pharmacophore Modeling**
- ❏ **Cheminformatics**
- ❏ **Development Environment**

2022/07

Comprehensive Software System for Life Science

# Molecular Operating Environment (MOE)

MOE, the Molecular Operating Environment, is a comprehensive software system for Life Science developed by Chemical Computing Group ULC. (CCG).

MOE is a combined Applications Environment and Methodology Development Platform that integrates visualization, simulation and application development in one package.

MOE strongly supports drug design through molecular simulation, protein structure analysis, data processing of small molecules, docking study of proteins and small molecules, and so on under the unified operations.

SVL, the Scientific Vector Language, is the portable high-performance programming language built-in MOE. SVL is the Vectorized command language, scripting language, and applications programming language.   One-tenth reductions in code size over C and FORTRAN are routinely realized. SVL source code to MOE applications is distributed with MOE to end-users; this allows MOE applications to be freely customized or modified.
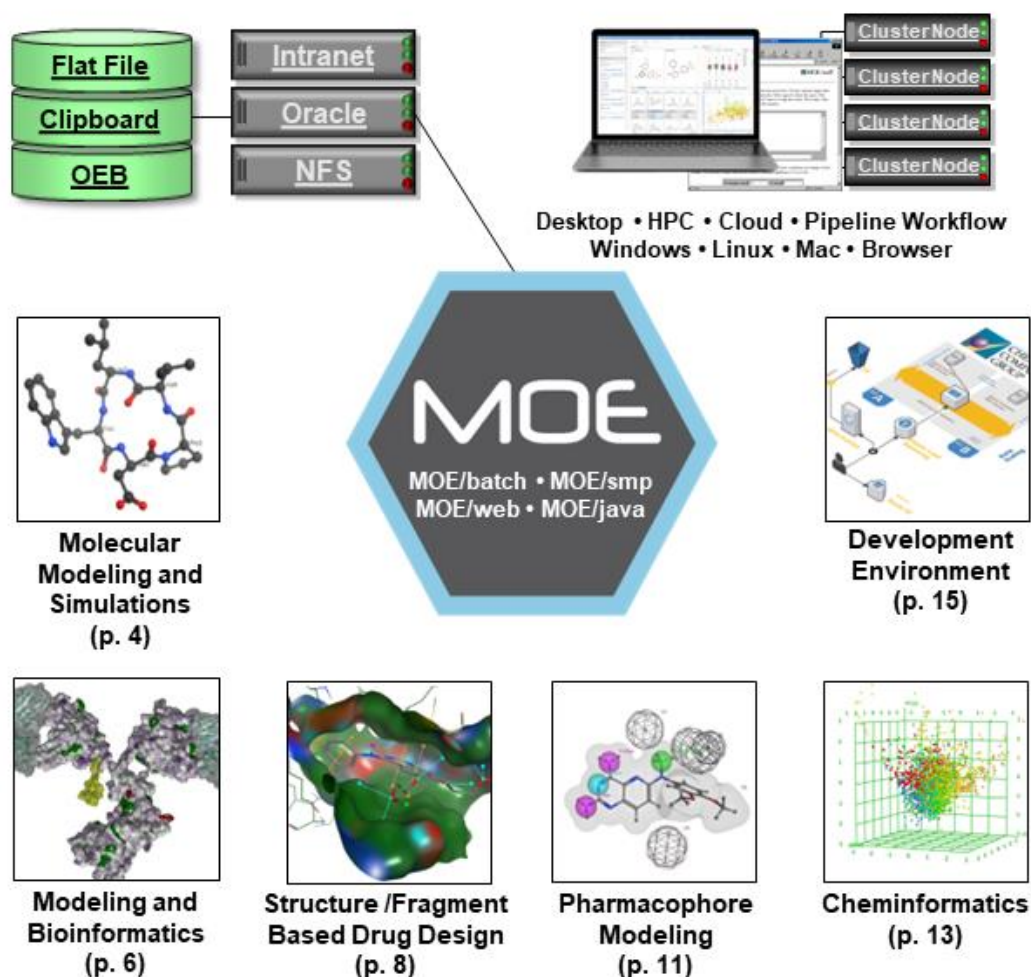
## ❑ Graphical User Interface (GUI)

The GUI of MOE provides the common working environment on various platforms (Microsoft Windows, macOS and Linux) from PCs to workstations.

Since the molecular structure data are managed hierarchically, it is easy to refer and edit the data. It also provides various display functions like molecular surfaces, protein secondary structures, hydrogen bonds and so on. MOE also supports the hardware stereo.
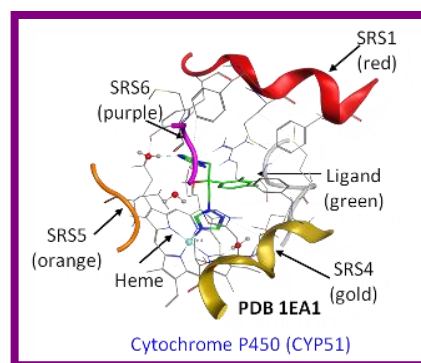


## ❑ Molecular Database Management

The database has an original format, MDB (Molecular Database), for effectively processing huge data used in computational chemistry, and can be accessed in high speed through the spreadsheet-like user interface. It is able to save any results of molecular dynamics, conformational search, construction of compound libraries, QSAR, *etc*. in a unified format and to analyze them with statistics and graphs.

It can read/write SD files and ASCII files and save SD file contents in 1/5 of its disk space with MDB format. e.g. MOE provides a database of non-redundant protein domains from the Protein Data Bank
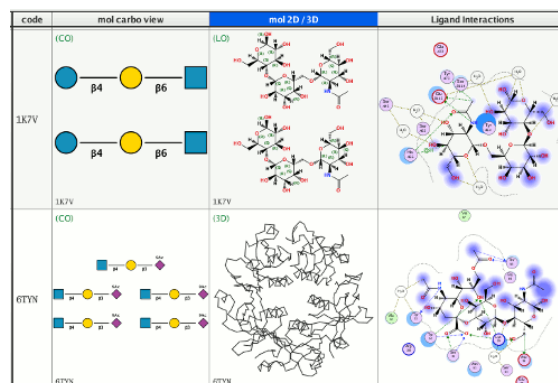


## ❑ SVL Source Code

MOE applications written in SVL are basically open-source. With the use of the development tools, *e.g.*, SVL Text Editor, users can easily refer to the contents of programs and modify the algorithms or parameters.

The programs and scripts written in SVL run on various hardware; Applications developed on Windows PC's can be run on Linux PC's, workstations or Mac's without any changes in codes, and *vice versa*. SVL codes are automatically loaded and compiled at the startup of MOE in a short time.
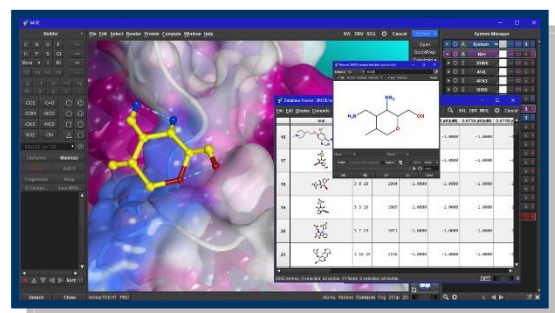


## ❑ High Performance Computing Framework

Large, parallelizable calculations can be submitted to external queueing systems directly from MOE through MOE's high performance computing (HPC) framework. Using this framework, MOE applications are able to provide a simple interface for submitting batch jobs to a job management system, and for monitoring their status. At the same time, the ability to run locally as a batch process or to save a job directory for calculation at a later time are also easily accessed.



## ❑ MOE GPU-accelerated applications

The SVL GPU API provides a unified interface to integrate CUDA® and OpenCL™ kernel source code into SVL programs. It allows to send SVL vector data to the GPU and use it for the computation for some applications.

3
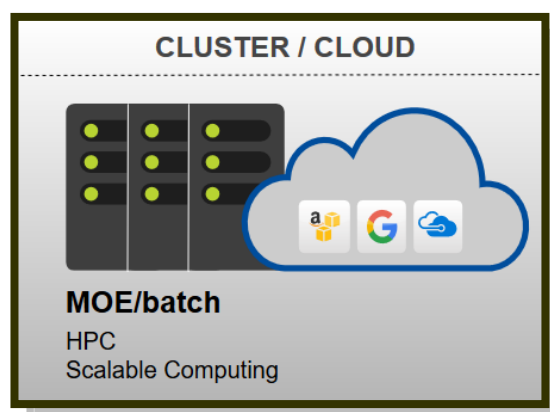
# Molecular Modeling and Simulations

**MOE's representation of organic chemical structures and flexible architecture provide a solid foundation for molecular modeling and computational chemistry. User-friendly molecule editors, wide variety of available file formats, choice of validated forcefields, powerful modeling applications, and customizability all make MOE the most flexible molecular modeling environment.**

## ❑ Molecular Builders and Data Import/Export

Create or edit small molecules, peptides, proteins, carbohydrates, DNA/RNA and crystal structures using a variety of molecular builders. Import and export molecular structure information, and other data, in many standard file formats. Use SD pipeline command-line tools for compound filtering, descriptor calculation and structure depiction directly on SD files.

## ❑ Molecular Mechanics and Dynamics

MOE contains an integrated multiprocessor forcefield engine with open parameterization. Parameters for Amber '94/'99, Amber10/14EHT*, PFROSST**, CHARMM27, MMFF94(s), OPLS-AA and Engh-Huber models are included in the distribution. Three large-scale energy minimization algorithms are available, and distance, angle and torsion angle constraints can be imposed. Atomistic NVE, NVT, NPT and NPH ensemble molecular dynamics simulations can be conducted, optionally including atomic wall restraints and tethers. To utilize HPC, NAMD and AMBER interface is also available. Free Energy can be calculated with Thermodynamic Integration(AMBER TI) using AMBER from the interface. AMBER TI interface generates the submission scripts to run AMBER TI simulations on a cluster or in the cloud directly or through a job management system.
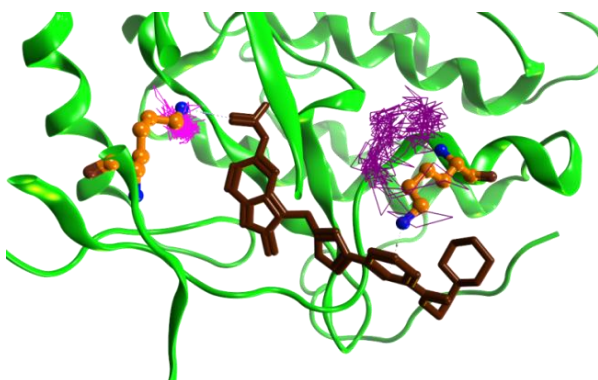
And support for Hydrogen Mass Repartitioning (HMR), which redistributes mass from heavy atoms to light atoms to slow down motion without changing the overall mass of the system, has been added to AMBER Thermodynamic integration, Dynamics and Low Mode MD.

## ❑ An all-atom forcefield combining EHT and Amber10/14EHT

MOE's forcefields have been added which combines Amber10/14 and Extended Hückel Theory (EHT) bonded parameters. The Amber upper case atom types and parameters are used for amino acids and nucleic acids as well as certain other analogous chemical motifs. For other moieties, a minimal set of lower case atom types are used to strictly select the Amber non-bonded parameters and EHT bonded parameters. Thus, the Amber non-bonded model is used for all atoms, and the Amber10/14 or EHT bonded parameters are used where appropriate. Amber10/14EHT is suitable for small molecules, macromolecules or both. AM1-BCC charges are expected for small molecules. Group II ion and Group VIII parameters from OPLS-AA.

## ❑ Macromolecular System Preparation

This application for preparing macromolecular structures for the subsequent simulation works is common structural problems in X-ray crystal structure data (lack of terminal caps, missing atoms, missing loops, *etc*.) are automatically diagnosed and presented in a list. Each diagnosed problem can be inspected and fixed from the graphical user interface.

## ❑ Spectral Analysis

The Spectral Analysis application supports computational studies of NMR and VCD spectra for small molecules and peptides. In addition, chemical shift, coupling constant and/or pairwise distance data can

be used to determine the relative populations within a set of conformers, yielding the in-solution conformer distribution. This information can be used for compound analysis as well as for understanding the energetics of binding.

## ❑ Database Viewer Display

The Database Viewer has new molecule display features and also supports the display of graphic objects. A carbohydrate display mode has been added in which carbohydrates are automatically detected and displayed using SNFG symbols. As chirality display, the option has been added to display atom stereochemistry with bond wedges and chirality labels.
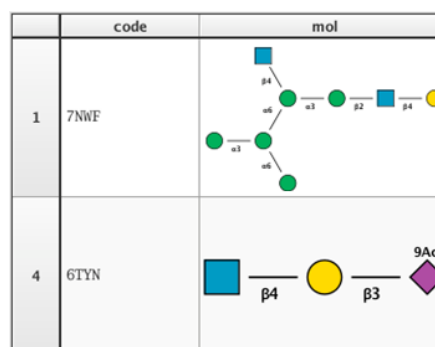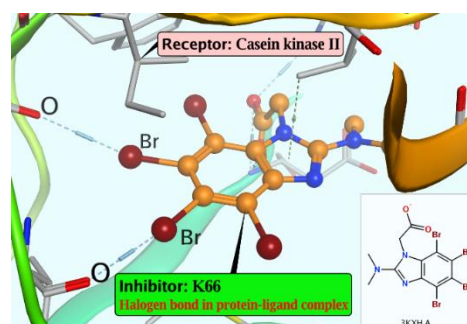


## ❑ Conformational Analysis

Perform a conformational analysis using a Systematic, Stochastic or LowModeMD search methodology (with the full choice of forcefield). Use a fragment-based high-throughput methodology for construction of large conformation databases. Visualize and analyze the resulting ensembles of conformations. LowModeMD is a conformational search method that uses implicit vibrational analysis to focus an MD trajectory along the low-mode vibrations. This has the effect of searching for minima along the valleys and troughs on the potential energy surface. LowModeMD has the best overall performance on a wide set of structures from small molecules to peptides, macrocycles, and protein loops.

## ❑ Quantum Chemical Calculations

MOE provides interfaces to popular *ab initio*, semi-empirical and density functional codes such as GAMESS, Gaussian, ADF, MOPAC7, and MOPAC2016. Uses a graphical interface to configure a calculation and view results such as electron density contours, molecular orbital plots, energy level diagrams and other properties. Use MOE/smp to conduct the calculations on collections of compounds in parallel.

## ❑ Nonbonded Interaction Visualization

The methodology used to determine, score and render hydrogen bonds, clashes and other non-bonded interactions such as Halogen bond has been replaced with a more sophisticated technique. A fast Extended Hückel Theory calculation is automatically performed on the molecular system and the resulting sigma/pi charges and fractional bond orders are used to parameterize the non-bonded energy scores.



## ❑ Torsion Scan and Analysis

A new application for performing torsion scans on small molecules has been added to MOE. The application calculates energy as a function of dihedral angle and can be used to assess conformational flexibility, estimate rotational barriers and identify preferred dihedral angles.

## ❑ ONIOM (QM/MM)

Large systems, such as ligand/pocket complexes, can be treated by QM/MM methods. In this, a (small) region of interest (the inner layer) – for example, the locus of charge transfer or bond making/breaking, i.e. "where chemistry occurs" – is treated quantum-chemically, while the rest of the system (the outer layer) is modeled by a forcefield. In complex systems, a middle layer might be included that is treated quantum-chemically using a lower level of theory. Determining the complex energy or minimized geometry using this heterogeneous approach can gain a significant reduction in computation time over treating all atoms with QM.
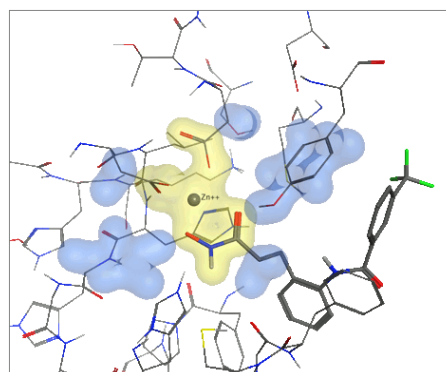
**Applications for protein structure prediction are powerful and easy to use, both for experts and occasional users. Powerful homolog search, alignment algorithm and refinement methodology make high-quality sequence-to-structure predictions possible.**

### ❑ Protein Structure and Family Databases
CCG has processed all the PDB structures, corrected common errors and produced a cleaned version of the PDB. Search the database using Code, Header, Compound, Title, HET groups, -Resolution, *etc.* The cleaned database is subjected to an exhaustive and iterative structural clustering procedure to generate the Structural Family Database. The result is a database of structural families in excellent agreement with expert hand-curated family databases.

### ❑ Remote Homology and Fold Identification
Search the Structural Family Database to identify protein families relevant to structure prediction. The search uses a FASTA-type local alignment followed by a family membership test based upon full multiple alignments and Z-score significance testing. Folds of even distantly related homologs can be reliably identified with few false positives (unlike pairwise searches).   Use the Domain Motif Search to identify structural homologs with low sequence identity.
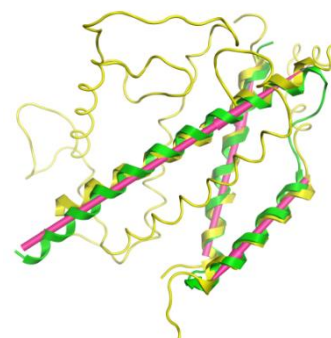
### ❑ Loop/Linker Modeler and Browser
This application in MOE is used to find and select loop conformations. Interactive searching and browsing of potential loop candidates make it possible to bring expert knowledge of a system to bear when selecting loop candidates for further optimization. Ready access to useful tools such as affinity and electron density scoring make it possible to use additional criteria when selecting a loop conformation.

### ❑ Structural Family Analysis
Understand the conserved features and differences between related protein structures (and homologous sequences). A 3D structural family analysis provides insight into conserved geometry, conserved water, salt bridges, hydrogen bonds, hydrophobic contacts and disulfide bonds that are often undetected in sequence alignments. Use structural or sequence dendrograms to eliminate outliers and improve alignments.

### ❑ Mutation and Rotamer Exploration
Discover accessible natural/non-natural amino acid side chain conformations with MOE's Rotamer Explorer. This predicts the structure of amino acid mutations in a 3D protein structure. Rank candidate rotamers using an energy-based scoring function and visually analyze them using MOE's graphical interface.
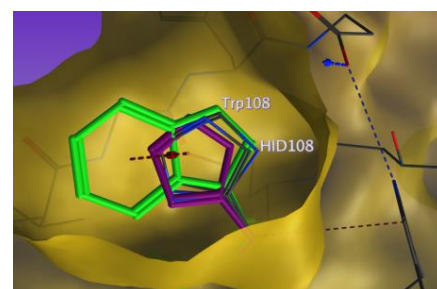
### ❑ Structural Quality Assessment
Assess the reliability of predicted structures with statistical measures of quality derived from X-ray crystallographic data. Diagnostic measures such as Ramachandran Plots and a Stereochemical Report make it easy to identify and isolate regions of predicted structures that require further treatment.

### ❑ Homology Modeling and Antibody Modeling
Build a homology model from an amino-acid sequence by statistically assembling fragments of experimentally determined backbone structures from one or more templates, selection of sidechain conformations from a rotamer library, followed by a refinement protocol based on energy minimization in the forcefields. A ranked ensemble of protein structures is produced and written to a MOE molecular database for subsequent analysis or further refinement. Include environment units such as bound ligands and conserved waters in the structural template. MOE also provides antibody specialized modeling application (Antibody Modeler) and updatable template database for antibody modeling. Antibody Modeler can also be building a single-chain Fv by joining

6

two Fv's by a linker (VL-linker-VH, VH-linker-VH), where the linker is part of the query sequence. The resulting scFv fragment can be grafted onto full or fragment antibody templates..
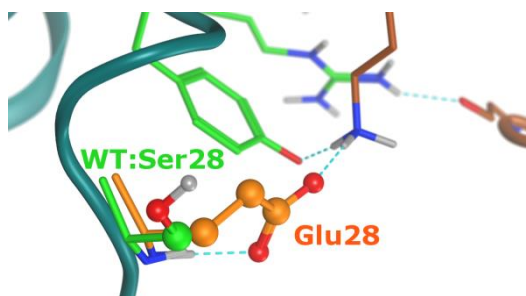
## ❑ Protein Engineering/Protein Design

Explore and compare mutant series against a wild type with a unified protein engineering application. Conduct Alanine Scanning to systematically explore affinity. Assess protein stability and optimize unstable regions by identifying disulfide bridging opportunities through Cysteine Scanning. Rationalize and perform single point or multiple mutations via Residue Scanning to assess and advance lead candidates. Use Sequence Design to search for all possible multiple mutations to determine an optimal sequence. Easily identify residues prone to natural mutation, based on single nucleotide polymorphism, using Resistance Scanning. Automatically generate ensembles using molecular dynamics or LowModeMD to estimate ensemble-averaged properties.
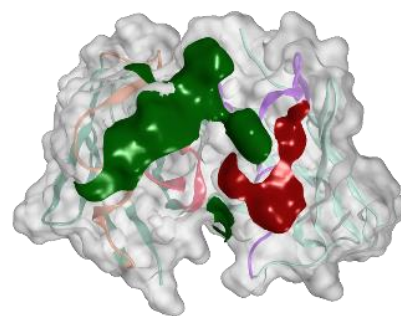
## ❑ Sidechain packing

The massively parallelizable parts of the Unary Quadratic Optimization (UQO) techniques used to efficiently search for best combinations of rotamers have been reimplemented for GPU. This results in performance enhancements in some applications (Homology Modeling, Protein-Protein Docking etc.)
Sidechain placement can see a 90x speedup, and Protein Modeling, an overall 40% speedup.

## ❑ Protein Properties

Calculating molecular properties and descriptors presents some particular challenges when dealing with large biomolecules such as proteins. The structural flexibility means that many properties of interest cannot be adequately described by a single molecular conformation and require averaging calculations across a representative ensemble of structures. In addition, there can be a large number of accessible ionization states (i.e. protonation of titratable residues) so that the observable properties do not correspond to any single state and vary as a function of the ambient pH value. It is also of interest to break down the observable properties into per-residue contributions in addition to the total values.
The Ensemble Protein Properties application in MOE provides a common interface for all calculations of properties and descriptors specialized for proteins with ensemble sampling and pH-dependence built-in as a central feature of the calculation. a comprehensive set of sequence- and structure-based physical properties such as pI, zeta potential, mobility, dipole moment, *etc.* for building QSPR model or helping rationalize protein solubility, solution viscosity, stability, and aggregation. Use the predicted properties in conjunction with preliminary experimental data to rationalize stability and aggregation at a given pH. Calculate properties for an ensemble of mutants to identify and predict physical property trends on a relative scale.

## ❑ Protein-Protein Docking

In Protein-Protein Docking, a coarse-grained bead model is used to reduce the computational search space. In the coarse-grained minimization step, the calculation of pairwise bead interaction energies, which represents a significant amount of the processing time, has been reimplemented for GPU. This improvement have resulted in a 6–10x speedup in cases dominated by bead refinement, and a 3x speedup in structures that are faster to minimize.

# Structure/Fragment Based Drug Design
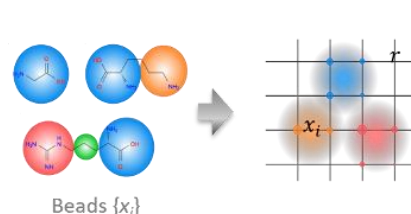
**Macromolecular crystallographic data, when available, can be a valuable source of information for discovering active ligands. MOE provides various applications for visualizing and understanding details of receptor active sites and receptor-ligand interactions. These applications can be used to suggest improvements to the known ligands or screen ligand databases for ligand candidates..**

### ❑ Active Site Detection

Detect candidate protein-ligand and protein-protein binding sites using a fast geometric algorithm based on Edelsbrunner's Alpha Shapes. Each site on a macromolecular structure is ranked according to its PLB index*. Visualize individual sites or populate them with "dummy atoms" for docking calculations or starting points for *de novo* ligand design efforts.

*PLB index: Soga, S., Shirai, H., Kobori, M., Hirayama, N.; Use of Amino Acid Composition to Predict Ligand-Binding Sites; *J. Chem. Inf. Model.* **47** (2007) 400-406.
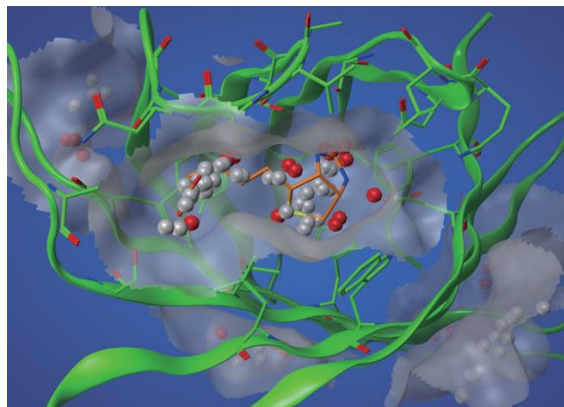


### ❑ Electrostatic Maps

Determine favorable locations of neutral, positive and negative features in an active site. The electrostatic maps are calculated by solving the non-linear Poisson-Boltzmann equation for receptor atoms and pseudo-ionic species. The isocontour levels of the generated predicted map are expressed in kcal/mol. The advantage of the present method is that fully screened electrostatic potentials are used and the domination of the potential by ionic groups is avoided.

### ❑ Probabilistic Contact Potentials

Visualize and understand directional details of hydrophobic or hydrophilic contact preferences of a receptor or ligand using probabilistic contact maps. CCG has developed a suite of analytical probability distributions that correspond very well to inter-atomic distance, angle and out-of-plane angle histograms derived from a large collection of crystallographic structures. These distributions are then used to form a composite "preference map" for a given macromolecular structure.

### ❑ Docking simulation

MOE's Dock application searches for favorable binding modes between small- to medium-sized ligands and a not-too-flexible macromolecular target, which is usually a protein. For each ligand, a number of placements called poses are generated and scored. If desired, the poses can be covalently bound to a receptor side-chain atom. The score can be calculated as either free energy of binding including, among other contributions, solvation, entropy, and enthalpy terms based on polar interaction energies (including metal ligation) or qualitative shaped-based numerical value. Poses can be optionally constrained to fit a pharmacophore query, and final filtering by pharmacophore query is also possible. The atoms of the active site can be permitted to move: sidechains can be tethered, but backbone atoms are always fixed. The final highest-scoring poses, along with their scores and conformation energies, are written to a database where they are ready for further analysis.
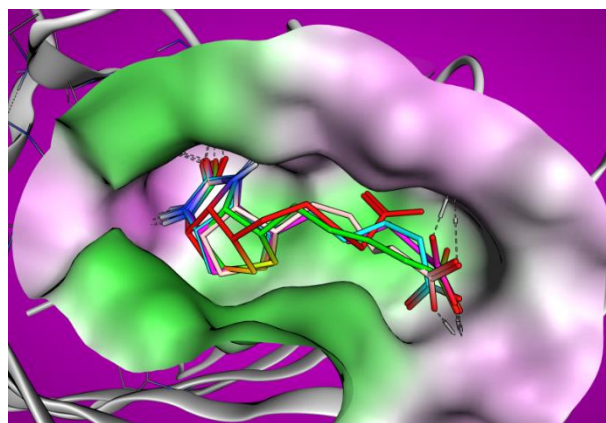


### ❑ GBVI/WSA dG scoring function

The GBVI/WSA dG scoring function is based on the forcefield, GB/VI implicit solvent model and a weighted surface area term. There are only three fit parameters in the scoring function each of which is related to

Linear Interaction Energy principles and were fit to experimental binding free energies.

## ❑ Multi-Fragment Search

Multi-Fragment Search is an ensemble-based methodology for mapping the preferred locations of specific chemical groups in a receptor structure. An active site of a macromolecular structure is populated with a large number of chemical fragments, which are subjected to an energy minimization protocol. The resulting group locations are clustered, scored (including solvation effects) and output to a database for subsequent visualization and analysis.



## ❑ Active Site Depiction

Use the protein-ligand interaction diagrams to easily identify polar, hydrophobic, acidic and basic residues. Visualize solvent-exposed ligand atoms and residues in close contact with ligand atoms as well as sidechain and backbone acceptor and donor interactions.

## ❑ Protein-Ligand Interaction Fingerprints (PLIF)

This tool is a method for summarizing the interactions between ligands and proteins using a fingerprint scheme. There are two categories of interaction in which a residue may participate: potential (energy-based) contacts and surface (patch) contacts. For potential contacts, the value is that of the strongest interaction between any pair of atoms in the residue and ligand, whereas for surface contacts the value is the total contact area of each type between a residue and the other molecule.
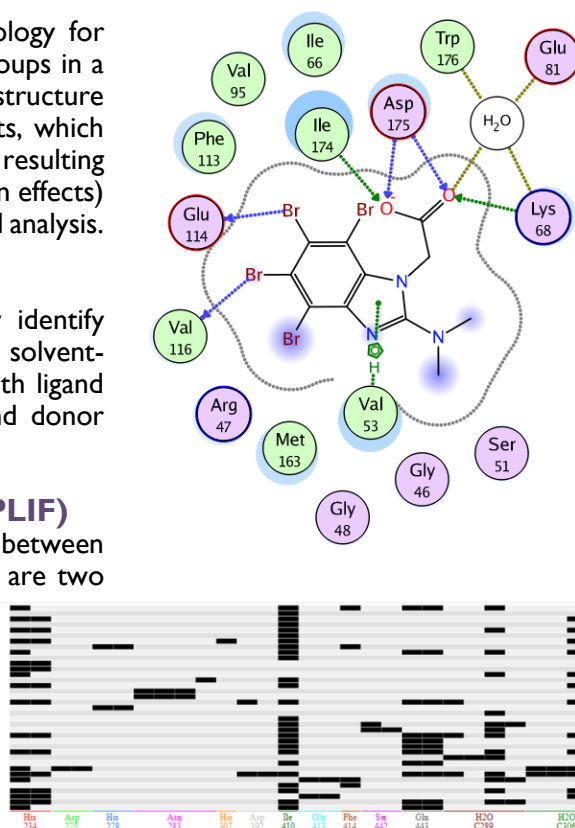


## ❑ QuickPrep

QuickPrep is a streamlined interface for interactive 3D ligand optimization in the active site. A dedicated button bar provides commands for structure preparation, active site analysis, molecular property/binding affinity calculations and ligand modification and optimization in the active site. QuickPrep can be used by both computational and medicinal chemists.

## ❑ Molecular Surface and Maps

Molecular Surface and Maps is an integrated application for active site analysis. Create molecular surfaces, predict contact preferences and calculate electrostatic maps. Color molecular surfaces by choosing from a variety of schemes such as temperature factor, pocket, lipophilicity, and electrostatic potential.



## ❑ Scaffold Replacement

Link annotations denote substitution points on a candidate scaffold molecule and the locations of potential R-group substituents. Search standard 3D conformational databases or special scaffold and linker databases to find novel chemical scaffolds that preserve substituent's geometry. Add additional pharmacophore features to preserve known scaffold interactions or volume constraints to satisfy shape requirements.

## ❑ Combinatorial Builder

Multiple attachment points and 3D R-group libraries can be systematically enumerated around a scaffold, refined and filtered in a receptor pocket.



## ❑ BREED: Ligand Hybridization

BREED is a method for the generation of novel inhibitors from
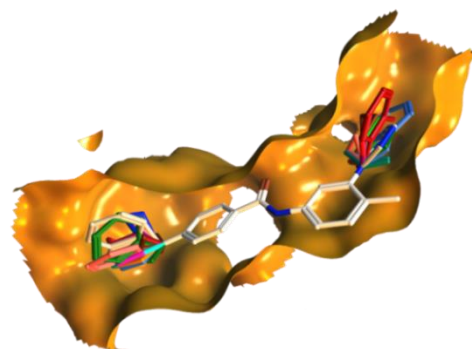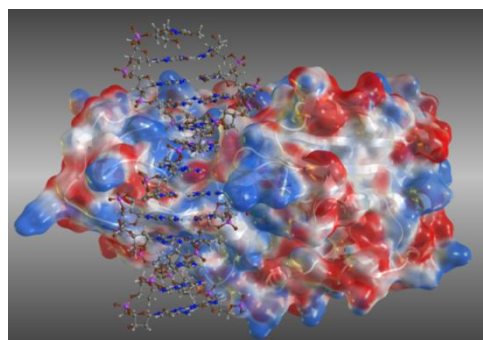
structures of known ligands bound to a common target. New ligands are generated by combining parts of existing ligands. The user can specify QuaSAR descriptor, Model file and/or Pharmacophore query filters to limit the results. If a receptor is present, structures can be refined in the binding cleft as with docking calculations. Thus, BREED can be used as part of a ligand-based or structure-based discovery methodology.

## ❑ MedChem Transformations

MedChem Transformations is an approach used to discover novel chemical structures by applying a set of transformation rules to existing ligands. Typical transformations might exchange functional groups, or alter all or part of individual rings while preserving the rest of ligand. These can be applied iteratively, resulting in cumulative changes to the starting structures. If a receptor is available, Forcefield Refinement can be used to generate bound poses along with various docking scores.
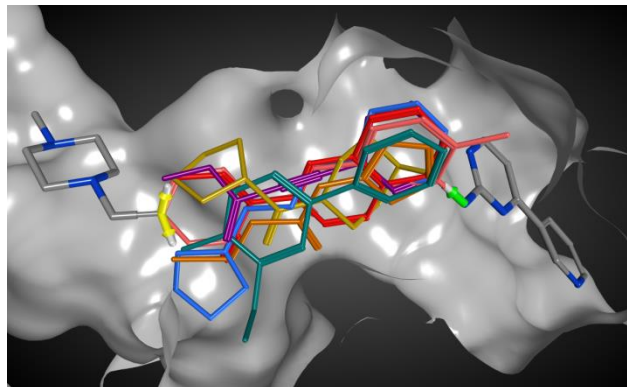


## ❑ Solvent Analysis with 3D-RISM

This application uses the 3D-RISM (three-dimensional reference interaction site model) method to analyze the role of solvent in macromolecular systems. The application computes a time-averaged distribution of water H and O densities, along with free-energy maps for analyzing solvent stability and solvation contributions to binding free-energy. A unique feature of 3D-RISM is its ability to include various concentrations of salt or hydrophobic molecules as part of the solvent, thereby producing density maps for ions and hydrophobic groups as well as for water. The Solvent Analysis tool can be applied to various water-related tasks such as water placement in low-resolution X-ray protein structures, estimating solvent effects in protein-ligand binding, and designing ligands to exploit solvation effects such as water displacement.
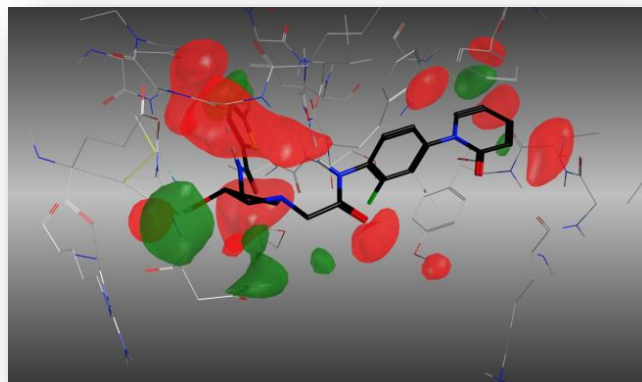


## ❑ MOE project applications

The MOE Project is a highly configurable meta-protocol for data organization. It takes a set of standard format (PDB, mmCIF, SDF, MTZ, etc.) data files as input, processes this data, and outputs a single .mdb file. Structures can be aligned and superposed, optionally with electron density around the active site and/or ligand topology correction. Various forms of Structure Preparation can also be carried out, and associated data assembled or calculated, possibly using SOAP, HTTP or a number of other common access methods.

The MOE Project application is used to manage this information and keep it up-to-date by assembling and organizing the data into a single MOE database (MDB) file. The MOE Project framework supports aligning and superposing all structures into a common reference frame, performing structure preparation work, and more. By centralizing data into a single repository, MOE Project helps to ensure data integrity and consistency. The use of a single integrated structure also facilitates data dissemination and maintenance..

# Pharmacophore Modeling

**MOE's pharmacophore modeling is a powerful means to generate and use 3D geometric information to search for novel active compounds, particularly when no receptor geometry is available. Pharmacophore methods use a generalized ligand representation and geometric constraints to bypass the structural or chemical class bias of 2D methods.**

### ❏ Pharmacophore Annotation Scheme

A pharmacophore annotation scheme, or more simply a scheme, is an automatic procedure to assign pharmacophore annotation points (such as H-bond donor, H-bond acceptor, hydrophobe, etc.) to a 3D conformation of a molecule. The pharmacophore annotation points are not part of the molecule proper, even though some of the annotation points may be located directly on top of a particular atom of the molecule. MOE has a variety of built-in schemes which provide different types of annotations and use different policies to locate the annotations about a particular conformation.
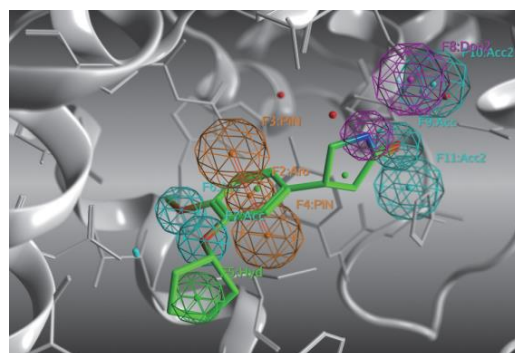


### ❏ Extended Hückel Theory for Pharmacophore Discovery (EHT scheme)

The EHT scheme provides a more nuanced means of investigating non-bonded interactions than the Unified or other traditional alternatives. These older methods use a SMARTS-pattern based atom/environment mapping approach similar to that used by classical force fields. The EHT scheme replaces this with annotation points derived from 2D EHT calculation which provides both the location and interaction type along with a direct measure of the H-bond interaction energies (strengths) for each interacting atom pair. This approach provides a better representation of the specific chemical environment of each ligand and can represent non-standard interactions not amenable to rule-based methods.
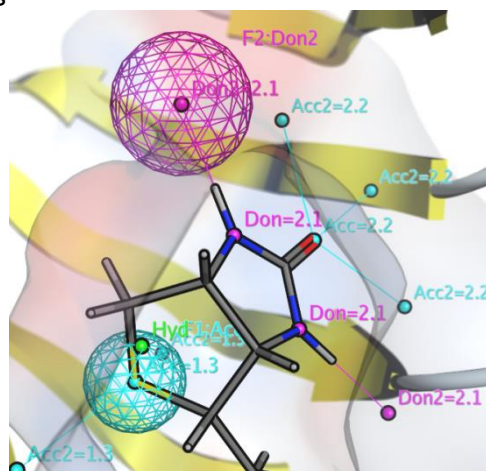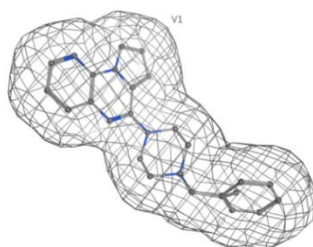


### ❏ Volume and Shape Filters

Volume constraints are spatial constraints imposed on particular atoms. These constraints serve both to confine hits to spatial regions in which steric clashes with the receptor are avoided as well as to ensure that favorable contact regions will be populated with appropriate atoms.
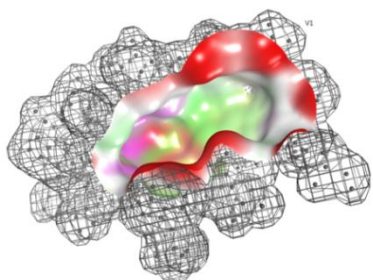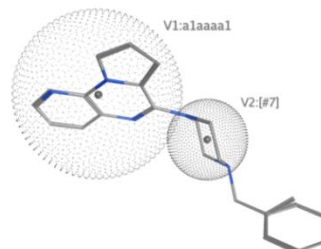


Excluded Volume:          Ligand Shape:          Occupied Volume:

## Pharmacophore Query Editor

Use an interactive editor to construct a 3D query from a molecular alignment or macromolecular structure. Use the query to filter a conformational database in an effort to determine candidate active compounds that satisfy the pharmacophore model. Refine the query to contain locations of Boolean expressions of pharmacophore features as well as restrictions on shape by using union-of-spheres for included, excluded and exterior volumes each containing an optional SMARTS chemical pattern.

### ❑ Pharmacophore Search

Rapidly search a conformational database for compound conformations that satisfy a pharmacophore query. Search multiple databases, a sub-range of molecules or a database of docked compounds. Output data consists of molecules that satisfy the 3D pharmacophore query (either all conformations or just the conformations that satisfy the query). Partial matches, the output of all symmetric matches and the specification of essential features are supported.
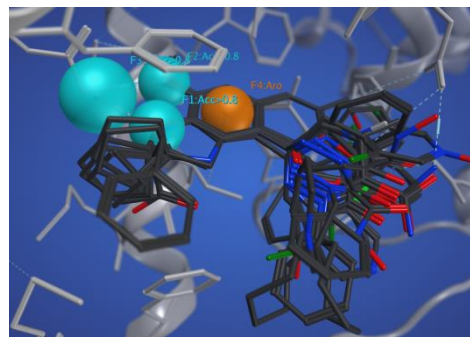


### ❑ High Throughput Conformational Analysis

Use MOE's High Throughput Conformational Search methodology to construct conformation databases for virtual screening. Conformational databases are constructed using a parallelized fragment-based approach. Molecules are subdivided into overlapping fragments each of which is subjected to a rigorous stochastic search. The fragment conformations are rapidly assembled by superposing the overlapping atoms. A database of fragments is maintained (and augmented as the search proceeds) making conformation generation of combinatorial libraries very fast.

### ❑ Pharmacophore Elucidation

Generate pharmacophore queries from a collection of input compounds (possibly with activity data) by considering all possible discrete geometries and all possible combinations of feature query expressions. A build-up strategy is used to avoid the combinatorial explosion. Enforce limits on feature counts and add custom query expressions. Each pharmacophore query is scored based on known active compound coverage, statistical activity enrichment and atomic overlap of matching conformations. The resulting scores and induced molecular alignments of high-scoring pharmacophore queries are written to a MOE molecular database for further analysis.
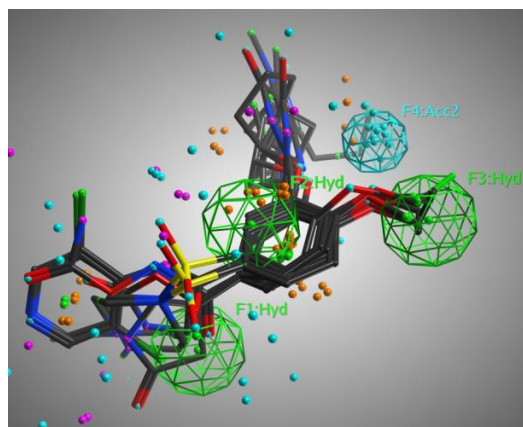


### ❑ Pharmacophore Consensus

This application suggests possible pharmacophore queries based on a set of aligned active compounds. A consensus calculation requires a set of aligned input molecules, a tolerance radius, the consensus score threshold, and the consensus score mode. Pharmacophore consensus is particularly useful when starting from a few highly active compounds.
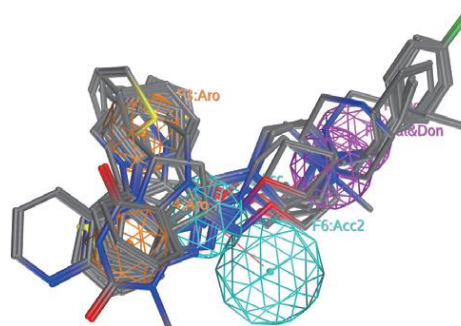
**MOE provides a suite of applications for manipulating and analyzing large collections of compounds, building property models, consensus models, and SD pipeline command-line tools**.

### ❑ MOEsaic

MOEsaic is a web-based application for analyzing SAR data, visualizing trends, exploring new virtual leads, and documenting results. Its streamlined interface features interactive MMP analysis and R-group profiling for fast assessment of property cliffs and SAR transferability of fragments, an integrated sketcher for designing novel structures, and models for property prediction.

**Matched Molecular Pair Search**

A Matched Molecular Pair (MMP) search algorithm finds chemical structures that differ in some substructures. MMP Search is a special kind of substructure search that defines the query by the portion of the molecule that is changing, as opposed to the constant portion query of the traditional substructure search. MMPs are important in SAR/SPR analysis because they can detect activity cliffs: small changes in structure that lead to large changes in activity or selectivity. Alternatively, changes in structure that preserve activity or selectivity are potential bioisosteres.

**Properties Calculation**

Perform automatic MMP fragmentation and calculation of descriptors HBA, HBD, tPSA, sLogP, MW. Real-time filtering/formatting and plotting of simple rule-of-thumb properties (Oral drugs / Sweet Spot / Golden Triangle) with conditional formatting. MOEsaic is a web-based application for analyzing SAR data, visualizing trends, exploring new virtual leads, and documenting results. Its streamlined interface features interactive MMP analysis and R-group profiling for fast assessment of property cliffs and SAR transferability of fragments, an integrated sketcher for designing novel structures, and models for property prediction.

### ❑ Molecular Descriptors

Calculate over 400 molecular descriptors including topological indices, structural keys, E-state indices, physical properties (such as *logP*, molecular weight, and molar refractivity), topological polar surface area and CCG's VSA descriptors with wide applicability to both biological activity and ADME property prediction. Use descriptors for classification, clustering, filtering, and predictive model construction. Add custom descriptors using MOE's built-in Scientific Vector Language.

### ❑ HTS-Binary QSAR

MOE's patented Binary QSAR methodology is ideal for building pass/fail models from high error content data and standard molecular descriptors. Use the resulting probabilistic models (based on Bayesian statistical inference) as a biasing agent in the design of focused combinatorial libraries.

### ❑ The Structure-Activity Report (SAReport)

SAReport contains sophisticated analysis methods to help scientists identify important groups and make more effective choices for synthesis. As experimental data builds up on discovery projects, it is often difficult for scientists and managers to keep track of the information. SAReport puts the information into a consistent context for better decision making.
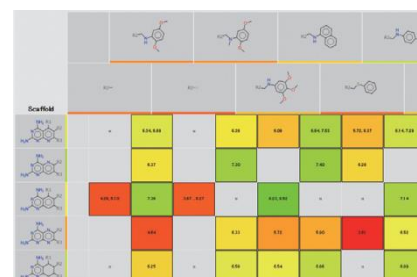
## ❏ QSAR/QSPR Predictive Modeling

Build QSAR/QSPR models using linear, probabilistic and decision-tree methodologies. CCG's unique Binary QSAR methodology is ideal for building pass/fail models from high error content data. Linear models include PCR and PLS methodologies and can support biological activity or ADME assessments.

## ❏ Similarity, Diversity, and Fingerprints

Perform similarity searching and diverse subset selection using Descriptor, Conformation, and Molecular Fingerprint methodologies. Choose between a number of fingerprint systems including 2-, 3-, and 4-point pharmacophore fingerprints in 2D or 3D and MACCS key fingerprints.

## ❏ Combinatorial Library Design

Enumerate both reaction-based or R-group-based virtual libraries either in 2D or 3D. Filter the virtual compounds by properties or pharmacophores. Build combinatorial libraries by combining scaffold and R-group databases or with MOE's reaction-based library generation method. Reactions can be chosen from a list or sketched in standard sketchers and applied to enumerate libraries of compounds. The default reagent database contains over 3,000 reagents curated from commercial vendors, however, custom reagent databases can be specified.

## ❏ RECAP Analysis and Synthesis

Analyze large collections of compounds to produce fragments resulting from retrosynthetic rules. Use the resulting fragments in a *de novo* synthesis methodology to produce novel chemical structures that have an increased likelihood of synthetic accessibility. Specify heavy atom mean and variance to control the size distribution on the randomly generated structures. Apply leadlike/druglike, QSAR/QSPR predictive models or 3D pharmacophore filters for *de novo* virtual screening applications.

## ❏ rsynth : Synthetic Score

The synthetic feasibility score, which is the fraction of the atoms of each new structure that ultimately appear in a retrosynthetic fragment found in the starting materials database. A value of one indicates that the molecule is very likely synthesizable.

## ❏ SD Pipeline Command Line Tools

Operate directly on SD files for structure depiction, acid/base protonation state, database filtering, sorting and descriptor calculations. Remove records that do not satisfy a series of filters (*e.g.* lead-like, reactive groups, drug-like, *etc.*), sort records and remove duplicate entries from SD files. Calculate descriptors and write the output to SD or ASCII formats.

## ❏ Combinatorial Library Enumeration.

The Combinatorial Library Enumeration (CLE) web application is a reaction-based library enumerator. It uses a novel probabilities-based virtual compound enumeration algorithm that allows for browsing and tuning the fully-enumerated library interactively without the library actually having to be fully-enumerated.

**In addition to the suite of graphical applications, MOE contains a tool for making changes to the existing applications or creating new applications. With MOE, expert modelers, application developers, and occasional users can benefit from sharing the same software system. Methodology written by application developers can be validated by expert modelers and then deployed to occasional users using either the MOE graphical interface or a Web interface.**

❑ **Scientific Vector Language (SVL)**

SVL is the built-in command language, scripting language and application development language of MOE. SVL is a "chemistry aware" vectorized computer programming language with over 1,000 specific functions for analyzing and manipulating chemical structures and related molecular objects. SVL is a concise, high-level language and SVL programs are typically 10 times smaller than equivalent programs written in C or FORTRAN. SVL source code is compiled to a "byte code" representation, which is then interpreted by the base run-time environment making SVL programs inherently portable across different computer hardware and operating systems.
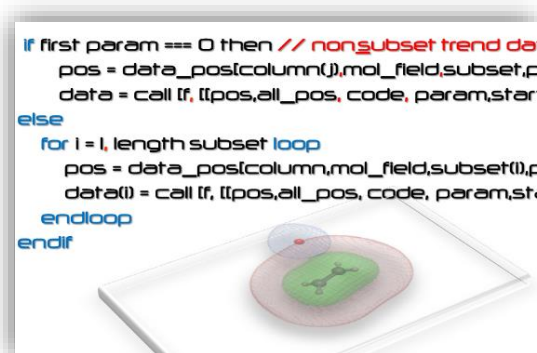


❑ **Background Computing**

MOE/batch is an adaptation of the MOE run-time environment intended for batch or background calculations that do not require a graphical interface. All non-graphical MOE functionality is accessible. MOE/web technology distributed with MOE is used to create Web Browser interfaces to MOE applications. SVL scripts and programs can be used to automate workflow, and are convenient for performing repetitive tasks or for modifying options.

❑ **Cloud Computing**

Cloud service providers such as Amazon Web Services (AWS) provide rapidly scalable on-demand compute and storage capacity allowing large scale, parallelizable computations to be carried out quickly and in real-time. In addition to AWS, there are numerous alternative cloud computing providers. You can select a provider which meets the needs of your organization.

❑ **HPC Job Management Framework**

Large, parallelizable calculations can be submitted to external queueing systems directly from MOE through MOE's high performance computing (HPC) framework. Using this framework, MOE applications are able to provide a simple interface for submitting batch jobs to a job management system, and for monitoring their status. At the same time, the ability to run locally as a batch process or to save a job directory for calculation at a later time are also easily accessed. Furthermore, calculation using GPU is also possible. The new gpu resource specification keyword is used to specify GPU usage on jobs submitted to job management systems via MOE's HPC framework.

## ❑ Cluster Computing

A standard part of MOE is the MOE/smp distributed computing technology. With MOE/smp, multiple cooperating computers can be used to perform large-scale calculations.

A heterogeneous collection of computers including laptops, workstations and multi-processor clusters, all running different operating systems, can be easily harnessed together in a single MOE session. The MOE/smp programming model in SVL makes it easy to parallelize SVL applications.
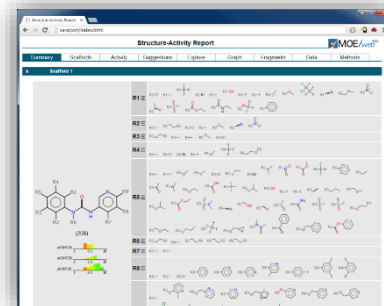


## ❑ MOE/web: Web Browser and SOAP Application Framework

MOE/web is an application environment that allows for large scale deployment of MOE applications or custom SVL programs to occasional users such as medicinal chemists via a simple web interface and/or to SOAP applications. MOE/web is distributed with MOE and customized applications can be incorporated.

## ❑ URL (HTTP/FTP) & TCP-IP Sockets

Access data from the internet via FTP, download files from web servers via HTTP and directly access third party database servers such as Oracle and DB2 via JDBC. MOE contains built-in applications to download new entries from the RCSB Protein Data Bank and a relational database browser. It is also possible to write SVL programs that interface with existing applications and libraries.

## ❑ MOE Extensions for KNIME

KNIME is a free data analytics and reporting platform. Through the MOE Extensions for KNIME, you can extend state-of-the-art Life Sciences Modeling into the KNIME platform. The largest and most complete plugin containing more than 180 components is based on MOE.
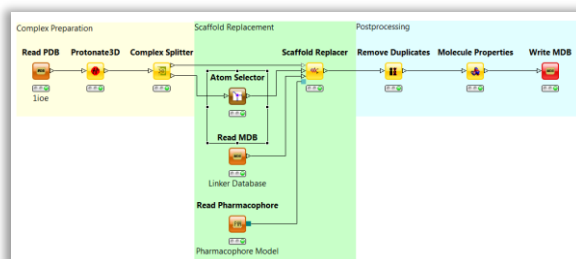
# License Style

MOE is offered under a floating license. It is possible to be used freely on any coexisting computers and workstations under the network environment with which the license of MOE is common to all the hardware supported by MOE.

The license of MOE is managed with network floating license with the token as a unit, and can be used from any machine within the limitation of the license. The number of tokens MOE uses at a time is different depending on the mode:

| Program (Mode of usage) | Number of necessary tokens |
|---|---|
| MOE (Graphics mode) | 3 tokens |
| MOE/batch (Non-graphics mode) | 1 token |
| MOE/web (Webservice mode) | 1 token and more |
| MOE/smp (Parallel computation mode) | 1 token x number of processes |
| | (When a non-graphics mode is used) |

MOE has no limit of the number of install machines.

# Support Platform

| Architecture | OS | Graphics |
|---|---|---|
| Intel | Windows 7/8/10/11 64bit | OpenGL* |
| | Linux　64bit (glibc version 2.7 or later) | OpenGL* / X11 |
| | macOS 10.12 and newer | OpenGL* / X11 |
| Apple Silicon (M1) | macOS 11　and newer, FlexNet v11.18 required | OpenGL * / X11 |

As for the supported platform, please inquire of MOLSIS Inc.

*To use full graphics effects and anaglyph stereo, OpenGL 2.1 or later with GLSL shader is needed.

## MOLSIS Inc.

3-19-9, Hatchobori, Chuo-ku, Tokyo 104-0032, Japan

URL: https://www.molsis.co.jp/en/

support@molsis.co.jp

## Chemical Computing Group

1010 Sherbrooke St. W, Suite 910 Montreal, Quebec, Canada H3A 2R7

http://www.chemcomp.com/