Journal of Computer-Aided Molecular Design, 1 (1987) 31–51
ESCOM
J-CAMD 005

Strategic approaches to drug design. I. An integrated software framework for molecular modelling

J.G. Vinter, A. Davis and M.R. Saunders

Smith Kline & French Research, The Frythe, Welwyn, Herts AL6 9AR, U.K.

Received 12 February 1987
Revised 2 March 1987
Accepted 4 March 1987

Key words: Molecular modelling; Computational chemistry; Drug design

SUMMARY

An integrated molecular graphics and computational chemistry framework is described which has been designed primarily to handle small molecules of up to 300 atoms. The system provides a means of integrating software from any source into a single framework. It is split into two functional subsystems. The first subsystem, called COSMIC, runs on low-cost, serial-linked colour graphics terminals and allows the user to prepare and examine structural data and to submit them for extensive computational chemistry. Links also allow access to databases, other modelling systems and user-written modules. Much of the output from COSMIC cannot be examined with low level graphics. A second subsystem, called ASTRAL, has been developed for the high-resolution Evans & Sutherland PS300 colour graphics terminal and is designed to manipulate complex display structures. The COSMIC minimisers, geometry investigators, molecular orbital displays, electrostatic isopotential generators and various interfaces and utilities are described.

Part 1. COSMIC

INTRODUCTION

COSMIC (COmputation and Structure Manipulation In Chemistry) is a software framework for chemistry into which any desired computational or graphical function can be integrated. It has been designed to handle most chemical information including databases, spectral analysis and expert systems.

COSMIC has evolved over eight years primarily for use as a proliferated modelling aid for bench medicinal chemists in the pharmaceutical industry. Although some of the modules can handle fragments of up to 10 000 atoms, most are presently parameterised for 300 atoms. The aim in designing COSMIC was to combine fast interactive response with scientific reliability. Little effort has been expended on producing cosmetically attractive graphics at the expense of scientific infor-

TABLE I

* COSMIC *

COmputation and Structure Manipulation In Chemistry

A Davis M Saunders JG Vinter 1987

Chemical display, Computational Chemistry, Databases and other Graphics Packages for Sigmex, Tektronix and VT Terminals

(Top Level Menu Items are numbered. Second Level Menu Items are indented.)

	Ref./notes
** 1] DRAW- Molecule & fragment construction	
** 2] VIEW- General view & manipulation	
** 3] INPOUT- Input/output routines	
**Read a COSMIC .DAT file from Backing Store	
**Write a COSMIC .DAT file to Backing Store	
**Enter coordinates manually eg from literature	
**Build an amino-acid sequence	
**Convert the running file to .MOL (Sybyl) file	1
**Convert from .MOL file to the running file	
**Convert from .XR(Chemgraf) file to the running file	2
**Convert from the running/.DAT file to .XR file	
**Conversions for < 10000 atoms for PS300 display	
**Display or Plot a large .DAT file	
**Convert from a .XR file to .DAT file	
**Convert from a .DAT file to a .XR file	
**Convert from a .DAT file to a .DSP file	
**Convert from a .PDB file to a .DAT file	3
** 4] ATMOL- Atomic & molecular changes & states	
**Geometry changes- bonds, angles, tors, invert, chirality	
**Geometry measures- distance, angle, planes, centroids	4
**Surgery- atoms, bonds, frags H, lp-off/on	
**Output all molecular internal coordinates	4
**Output polar atoms, close contacts, > 1-5 Hbonds	
**Renumber the atoms of a molecule	
**Moments of Inertia and Principal Ellipsoids	
**Accessible Surface area with or without Solvent	5
**Overall Molecular Volumes and Dimensions	6
**Accurate Molecular Volumes and Surface Areas	7
** 5] SPIN- To probe rotation barriers	
**SPIN01-Calculate 1 2 or 3 bond Rotational Map	#
**Display and/or Plot 1 Bond Barriers	
**Display and/or Plot 2 Bond Barriers	
**Store all minima of 3 Bond spin	*
**Display Motions during Rotations	
**RINGSPIN -Rotational barriers along an axis	#*
** 6] MIN- Geometry optimisation methods	
**MIN02-Newton-Raphson atom relaxer	#
**MIN03-Simplex atom relaxer	#
**MIN04-Simplex/NR combined atom relaxer	#
**MIN05-Parabolic Torsional Minimiser	#
**MIN06-Conjugate gradient minimiser	#
** 7] MULTIMOL- Conformer Space, Fitting & Docking	
**MIN01- Conf Space by sequential bond rotation & ring shapes (.MN1)	# *

	Ref./notes
**Show/Examine/Dump all Conformers from .MN1 file	*
**Distance search- compare up to six .MN1 files	
**Compare two structures manually & by lstsqrs	
**Generate Conformers with Matching Atoms	
**Dock Two Molecules Together	#*
** 8] MO-Approx Charges & MO data processing	
**CNDO/INDO- Closed/Open Neutral/Charged	8
**AMPAC- including MNDO MINDO/3 & AM1	9
**GAUSSIAN 80 ab initio	10
**Add Charges by Liverpool Method	11
**Display Eigenvectors and Dipoles	#
**ORBIT- Eigenvector display routine	*12
Display Frontier Orbital Parameters	
** 9] EIP-Electrostatic Isopotentials	
**VSS1- Calculate a proton 2D EIP map from MO Charges	13
**VSS2- Calculate a proton 3D EIP map from MO Charges	13
**NBMAP- Calculate an H-bond 3D EIP map by MM (.V3D)	#
**FINDIPMIN- EIP Minimum locator	#
**Display a 2D EIP map	
**Display a 3D EIP map	*
**Display 3D EIP Minima & Maxima only	
**Plot a 3D EIP Contour Map	
**10] DB-Database Links and other Systems	
**SYNLIB- Organic reactions and conditions	14
**LHASA- Organic reactions and transforms	15
**ORAC- Organic reactions and conditions	16
**MEDCHEM- Pomona Physical Chemistry & LogP	17
**CAMBRIDGE- Small molecule crystal structures	18
**BROOKHAVEN- Protein Data Bank	3
**PIR- Protein & Nucleotide Sequences	19
**FCD- Fine Chemicals Suppliers Database	20
**LABCHEM- In-house chemicals directory	
**SYBYL- Tripos molecular modelling programme	1
**CHEMGRAF- CDL molecular modelling programme	2
**MACROMODEL- Amber, MM2, Charmm from Columbia	21
**11] XYPLOT-Dumping routines	
**Stick pictures labels, types or stereo	
**Ball and stick or spacefill	22
**Full Output stick and spacefill	
**2D Plots	23
**ORTEP -Crystallographic pictures	24
**Plot a 2D Isopotential Map	
**Plot a 3D Isopotential Map	
**Plot the Electron Densities	
**12] Spectroscopy Programs	
**CNDUV99 -UV Spectra by MO methods	# 25
**NMRPROG -Ring Current Shifts & NOE effects	# 26

* Output data from these modules are routinely examined with ASTRAL on the Evans & Sutherland PS330.

See main text for description.

mation. On the other hand, a number of approximations have been incorporated which have increased interactive speed to an acceptable level whilst proving over the years to yield useful and scientifically meaningful results (see molecular mechanics and electrostatic potential methods).

The COSMIC framework incorporates modules from external sources and much in-house software. The entire system is written in FORTRAN 77 (with the exception of some of the modules from external suppliers) and currently runs on a VAX 11/780 with 8 megabytes of memory. The COSMIC core which excludes the databases and ab initio molecular orbital programmes is also installed on a microVAX II. The system can be run from Sigmex 5000 series emulators or through REGIS, Tektronix 4010, 4105 and 4107 emulators. COSMIC uses executable image 'chaining' facilities via verbose menu picking to move from one module to another and passes binary file data (running file) for continuity. Every menu item is activated from a screen cursor which can be driven by a mouse, bitpad or other interactive device. Chaining allows the extensive modularity of the system to be maintained and is the heart of the framework system.

Table 1 outlines the modules currently accessible from COSMIC. The main menu items are numbered 1 to 13. Activation of items 3–13 will display a second-level menu. One third-level menu is used to convert and display protein structures in item 3, but a third-level is not otherwise utilised.

METHODS

We do not intend to describe all of the functions and features of COSMIC because of its size. The details of the drawing routines, viewing techniques such as rotation, rocking, zooming etc., and structural analysis are installation specific and of little interest despite the efforts needed to programme them. Also, external programmes which have been incorporated without change (except for linker-, converter- and driver- software) will not be described. Only the details of the in-house written or modified techniques, and those which directly affect the scientific validity of results, will be explained.

Although some very large and comprehensive external programmes appear as menu items in COSMIC, the menu usually causes chaining only to the conversion routines which are necessary to prepare and transfer data between programmes. These routines generally access the main programmes automaticaly and there is no need for separate data files or running protocols to be prepared. All external programmes are capable of being run completely independently. Since COSMIC is an accessing framework and not a package, the external systems and programmes have not been violated by an act of incorporation into another 'package'.

1. The minimising techniques

(a) Conformational space hunting

MIN01 is a programme which samples the conformational space of a molecule. Basically similar to the 'SITAR' method [27], it cycles through all the specified bonds in a molecule, finding all local minima for rotation about the single bonds, and then permutes all the local minima over all selected rotating bonds. The conformations which give rise to the lowest energies are dynamically stored. Because the local minima may be dependent upon the starting conformation, this process may be repeated after 'randomising' the starting conformation. The programme can also generate

the energetically feasible conformations of medium rings. The conformational space of rings that are detected in the structure is determined first. Each substituent is removed and stored with knowledge of relative stereochemistry. The ring is flattened (and minimised) in the x,y plane and a set offset of z component is added to each atom randomly. The ring is then allowed to relax using one of the minimisers below. This forces the ring to relax into a new conformation dependent upon the pattern of allocation of the z components among the atoms of the ring system. This process is repeated a user-specified number of times, with rejection of identical conformations. Each substituent is then re-attached and the whole molecule is minimised prior to the search for acyclic conformational space. We have found this method to be very successful for 5-10 membered rings (with or without ring fusions). Beyond this limit, the algorithm (not unexpectedly) shows a marked preference for flat rather than folded structures.

The resulting 'conformational space' containing up to 200 structures over a user-specified energy range, can be analysed through an ASTRAL file or in COSMIC.

(b) Newton-Raphson minimisation [28]

This depends on the Newton-Raphson iteration formula for finding the zero point of a function. The final working equation defines the coordinate change, delta, necessary to move towards a minimum position as

delta = − (first derivative)/(second derivative)

where derivatives are those of energy with respect to coordinate.

This simple equation is remarkably successful at finding local minima. It is easy in practice to ensure that minima are found rather than maxima or saddle points (the formula above may find any of these, depending on the starting position). One way to do this is to use the step generated above (delta) and search for any position which decreases the energy (i.e. 1.5*delta, 1.0*delta, 0.5*delta...)— if none are found, leave the atom where it is and continue with the next. Another is to ensure that the search direction is always along the direction of steepest descent (i.e. use the absolute value of the second derivative of energy w.r.t. coordinates) and, optionally, combine with the step search. Yet another is to use the step length and direction generated as input to a parabolic minimisation routine, or to determine the initial size of a simplex figure. All are in common use in COSMIC depending on the nature of the problem to be solved, the nature of the energy function and required robustness.

When minimising bond lengths or angles where a minimum is well-defined, the NR method can be expected to converge in very few iterations. Charge-charge interactions and the non-bonded potential do not converge when the starting position is well away from a minimum and in the absence of any other forces or a method to ensure convergence to a minimum. Thus hydrogen bonds already made will minimise well, but intermolecular 'docks' cannot use an un-modified NR method.

In principle it would be possible to calculate the directions and step lengths for all atoms in a molecule and then move all atoms synchronously. In practice this does not converge rapidly for the NR method. The opposite extreme of moving atoms individually in x, then y, then z is possible but expensive in CPU time (at least 9 energy calculations needed per atom). The best compromise

is to calculate derivatives in x,y and z for an atom (7 energy calculations), move the atom along steepest descent and continue with the next atom. We have found that the use of block diagonal NR techniques does not improve execution time or convergence.

Analytical first and second derivatives can be used although, in our hands, the analytical evaluation of second derivatives is no faster than numerical evaluation (and slower for some function types, e.g. torsions). The required hard-coding of energy function derivatives also limits the utility of this method. A better use of analytical derivatives is to drive a minimiser which uses first derivatives only, e.g. conjugate gradient methods [30].

All derivative methods such as this suffer from the fact that molecules that are planar in two of the three axes will not move out of plane (the atoms have zero energy derivatives towards motion perpendicular to the plane). This could be a problem in COSMIC where structures are drawn freehand on a screen (and thus have only x and y coordinates specified) but can be put to use in generating ring conformers (vide infra), or avoided by rotating the molecule about any axis by a few degrees where numerical derivatives along the newly defined axes will be non-zero.

(c) Simplex minimisation [29]

MIN03 constructs an n-dimensional figure (which for a single atom to be minimised in 3 dimensions would be a tetrahedron) at the vertices of which energy calculations are performed. The figure is then flipped about the face opposite the vertex of highest energy. A single further energy calculation gives a new face on which to flip, and so on. At suitable times the figure is reconstructed with smaller (or larger) sides. The great advantage of simplex over other minimisation techniques is that convergence to a minimum is guaranteed (even if the energy surface is discontinuous). With a suitable choice of side length for the figure (e.g. from a Newton-Raphson derivative calculation), simplex need not be as slow as some workers have found. Its guaranteed convergence greatly outweighs its slowness in a number of applications (e.g. docking). Since it need not be based on derivative calculations it is easy for simplex to surmount small energy barriers in its search for a minimum, and so will sometimes converge to a different minimum from those found by other methods. It does not suffer from the same problem as the NR method regarding planar molecules as the atoms are forced out of plane by a three- dimensional simplex figure.

(d) Parabolic minimisation

The MIN05 one-dimensional minimiser can be used in conjunction with NR methods or on its own. It is most efficient when applied to functions which are close to parabolic in nature (i.e. bond stretches, angle bends and even torsional rotation). The method depends on the calculation of the value of the energy at three values of the independent variable (bond length, angle, etc.) followed by the fitting of a parabolic function and estimation of the minimum position of the function. As implemented, the subroutine can be called many times to provide a rapidly improving estimate of the minimum position. The calculation can be terminated when the energy returned by successive calls differs by some small amount, or when the minimum has been located to sufficient accuracy. Combination of a torsional angle minimiser with an atom-by-atom minimiser provides a very fast and efficient minimisation method.

(e) Conjugate gradient minimisation

Similar in concept to MIN05, the conjugate gradient minimiser (MIN06) implemented in

COSMIC uses analytical first derivatives to provide combined atom and torsional minimisation by the Fletcher-Reeves conjugate gradient method [30] and the parabolic minimiser described above. It provides an order of magnitude speed increase over, for example, the NR-based MIN02 minimiser, but is less versatile for programme development.

2. The docking techniques

(a) Single hard-sphere docking

The stationary molecule is placed at the centre of a 10 Å sphere, and the mobile molecule is placed in turn at points on the sphere about 1 Å apart, a total of 204 starting points. At each starting point, the mobile molecule is rotated about its centre in 60° steps about all three axes, and the lowest energy obtained defines the starting orientation for a subsequent simplex energy minimisation. Each molecule is treated as a rigid body, and no relaxation of bond lengths, angles or torsional angles is allowed. The simplex runs for a maximum of 500 iterations, or until the change in energy between successive iterations is less than 0.0001 kcal/mole. All 204 final positions of the mobile molecule, each at a discrete local minimum, is stored and the final picture analysed in the ASTRAL framework.

(b) Multiple docking with optional relaxation

A second procedure is used to examine build-up structures such as solvation shells and aggregates. The above process is carried out and the lowest energy dimolecular associate is stored and used in the next run as the fixed species. The following run uses the trimolecular associate and so on as many times as required. Between the completion of one run and the commencement of the next, a full relaxation of the relevant atoms helps to ensure self consistency within the growing aggregate. Furthermore, a recalculation of partial charges can be done on each new aggregate using a CNDO [8] subroutine. This allows the simulation of the charge redistribution that occurs when, for example, a hydrogen bond is made in the growing aggregate.

(c) ASTRAL interactive docking

See section on ASTRAL.

3. Investigation of barriers to rotation and ring movement

The programme SPIN01 allows the investigation of energy barriers associated with the rotation of 1,2 or 3 torsional angles consisting of four connected atoms ('defined' torsional angles) and/or the movement of rings and disconnected sets of atoms ('cyclic' or 'undefined' torsions). For 'defined' torsional angles (i.e. those where all acyclic atoms are joined) the user can perform a hard spin (involving calculation of energy only at each set torsional angle), a soft spin (allowing full atom relaxation at each torsional angle, with constraints to keep the torsional angle at its set value) or a torsional spin (allowing non-driven bonds to rotate to relieve strain). Constraints are applied in such an analysis by constraining the directions in which derivatives are taken for an NR minimisation. For 'undefined' or 'cyclic' torsions, only soft rotations are allowed, and constraints are applied in such cases by holding atoms in steep parabolic energy wells. Both soft and torsional energy minimisations are performed using a variant of the MIN05 type combined atom/torsion minimiser.

The soft method gives more accurate energy barriers than the hard but takes longer to calculate since a full energy minimisation is performed at each step. Torsional relaxation is intermediate between soft and hard methods in terms of time required and gives best results for examining co-operative rotational motions (in, for example, triphenylphosphine).

In all cases several sets of data are retained. The torsional angles which are set can be compared with the angles after minimisation to ensure that the programme has calculated correctly (when driving a torsion with a high energy barrier to rotation it is possible for a trade-off between the molecular energy and the parabolic forcing potential to disort the final angle). Energies for either hard or soft spins can be plotted to give a one-dimensional energy profile, a two-dimensional Ramachandran plot (contoured or isometric view) or a three-dimensional 'cube' suitable for display on a high resolution graphics display such as the Evans & Sutherland PS300. Molecular coordinates following soft or torsionally relaxed spins can be displayed to examine distortions introduced by the forced rotation process, or to produce animated displays which can give an insight into interconversion mechanisms. Boltzmann calculations on the energy profile can give an indication of both energetic and entropic factors in conformer interconversion free energy, and hence results may be compared directly with NMR experimental results.

Programme RINGSPIN provides a variation of a single bond rotation, whereby any axis through a group (e.g. a peptide bond, a 1,4-linked phenyl in a larger ring system) may be rotated incrementally whilst the rest of the system is allowed to relax.

4. Electrostatic isopotential (EIP) mapping

The levels of approximation used both in molecular mechanics and molecular orbital methods to investigate the relatively large structures encountered in drug design need to be carefully considered on cost- effectiveness grounds. We therefore compute the potential field using van der Waals and coulombic terms from the molecular mechanics formalism (NBMAP). The alternative wave mechanical techniques [13] were found to be too restrictive and expensive, although they are retained for comparison under the VSS heading. Extensive experiment has satisfied us that the results are qualitatively valid.

The procedure is to compute energies of interaction of the chosen probe (H^+, OH^-, K^+, etc.) on a 41 cube matrix (68 921 points) centered at the target molecule centre. The resolution can be varied and is typically 1 Å on an initial run. All points can be displayed via a .V3D file in ASTRAL but the most useful contours at ± 1 kcal/mole, ± 5 kcals/mole and all positive and negative minima, resulting from H^+ and OH^- probes, are routinely used for analysis. ASTRAL provides for an interactive display or removal of contours and minima on many molecules, so comparison of results without graphical 'clutter' is easy. Because of the resolution variability, the exact postion of a minimum and hence its magnitude is not always calculated. An extension programme has been written, called FINDIPMIN, which accurately locates minima by 'bathing' the target molecule in a 'bucket' of probes and minimising the interactions by a modified Newton-Raphson technique.

5. Display of molecular orbital (MO) results

Format conversion routines have been written to allow transparent input of the 'working struc-

ture' or a data file to be entered as a batch job into any of the three MO methods available on the menu. As the jobs finish, they are flagged to the screen where the results are picked up on the touch of a menu button as .MO files (specifically .CND, .MOM and .MOG). Each MO routine has been modified to output exactly the same data format. Graphical processing of these files results in the display of dipole moments (all components as orthogonal vectors), a total energy, the HOMO, LUMO, HOMO-1 and LUMO + 1 eigenvalues as energy levels up the screen and optional partial charge printing or displaying as scaled spheres. By activating any or all of the energy levels with the cursor, the relevant atomic orbital contributions (eigenvectors) to those levels are displayed as colour-coded, scaled and (hybridised) orbitalproportioned spheres at the atomic nodes. More accurately constructed orbitals and *d*-orbitals can be displayed using the programme ORBIT[12] on low or high resolution terminals.

6. Molecular orbital and charge utilities

The CNDO/INDO programme [8] has been extensively rewritten to take advantage both of the virtual memory organisation of the VAX and of faster methods of matrix diagonalisation [31]. These modifications have resulted in at least a two-fold increase in speed over the original code. There is also automatic convergence forcing via weighted interpolation of Fock matrices to prevent oscillation of charge-separated species. The software can handle 180 atoms or 450 orbitals.

AMPAC [9] and GAUSSIAN80 [10] are unmodified from the versions distributed by QCPE [33] except for some trival modifications to support the common output formats described above.

Routines which convert the running or data file to the required input for each of these, including *Z*-matrix conversion, are chained directly from the menu. Each of these drivers automatically submits jobs to a batch queue for background execution.

7. The plotting utilities

General plotting and graphics is programmed through the proprietary package SIMPLEPLOT [32]. The SIMPLEPLOT multiple device driver used in COSMIC handles Calcomp (81 and 1012) and PLXY plotters; Sigmex 5000, Tektronix (4010, 4014 and 4105) and Regis terminals; and screen and lineprinter character plots. It is also possible to produce a Metafile for other devices.

8. Spectroscopy calculations

CNDUV99 [25] is a slightly modified version of the programme available from QCPE[33]. The modifications involved the removal of the CDC-specific code and its replacement with standard FORTRAN.

NMRPROG is a programme to aid in the assignment of peaks in protein ^{1}H-NMR spectra. From the random coil shifts [34] it is possible to calculate a simulated spectrum for any desired set of protons (generally methyl and aromatic protons). From the X-ray coordinates it is possible to calculate ring-current contributions to the shifts using either the Pople dipolar model [35] or the Johnson-Bovey model [36] and adjust the simulated spectrum accordingly. It is our intention to include an implementation of the Haigh-Mallion [37] model at a later date. Constituents of the various peaks can be identified by cursor picking, and comprehensive printed output is available.

Nuclear Overhauser Enhancement can be simulated by defining a sphere around each proton making up the irradiated peak and enhancing the intensity of those signals due to protons within that sphere. Results are best displayed as simulated Nuclear Overhauser difference spectra. This can be done automatically for all peaks and the hard-copy output is very suitable for comparison with experimental traces. Variations of the user-defined sphere size simulate the effect of changing the irradiation time for the NOE experiment.

Calculated shifts can be overridden by the user if they are felt to be unreasonable or if a particular peak has been assigned. The programme also has a variety of options for averaging the shifts of possible magnetically equivalent protons (methyl protons, tyrosine and phenylaniline D and E protons or corresponding protons in crystallographically distinct chains in homopolymeric proteins such as human serum prealbumin) to give different models of the relative rates of interconversion on the NMR timescale.

Part 2. ASTRAL

INTRODUCTION

ASTRAL (Advanced System To Remove All Limitations) is the software package, written for the high-resolution Evans & Sutherland PS300 terminal, which complements the COSMIC framework. ASTRAL has been designed to display and manipulate any graphical data in eight display areas. Each area can accommodate any number of data, depending on the memory size, and each area can be manipulated and displayed individually or in combination with other areas. Data can be transferred from one area to another, written back to COSMIC or sent to a pen plotter. The PS300 network is small, flexible and general, and loading time is kept to a minimum. Our PS330 has two megabytes of memory and is serial-linked to the host VAX 11/780 via a 55K baud line. Table 2 lists the functions and utilities which make up the ASTRAL package.

ASTRAL communicates with a number of programmes in COSMIC which perform chemical calculations and output the results in a common file format (.AST or .V3D files). These files are read by the ASTRAL display routines and downloaded to the PS300 with user-specified characteristics. Once displayed, the structures can be manipulated with the usual tools (i.e. rotation, translation, clipping and zooming). The common file format means that communication between programme segments is simplified. The display routines are written to be as flexible as possible and can cope with many different display formats whilst retaining the possibility of identifying the fragments with a bitpad. Fast communication with the host computer (VAX 11/780) means that the host can, at any time, pick up the position of the fragment on the screen, update its coordinate files accordingly, and perform a new calculation. ASTRAL is written specifically for a virtual memory environment and has no hard-coded array sizes. This means that the only practical limit on the size or complexity of the structure to be displayed is the memory available on the PS300.

METHODS

1. ASTRAL PS330 network

Programme PBEGIN sends the manipulation networks to the PS300. The major function networks are: Scaling (all areas are scaled in unison) and Z-clip (also affects all areas); whole picture

TABLE 2
PS330 NETWORK FUNCTION KEY DEFINITIONS

No.	Key only	<SHIFT>key	<CTRL>key
1	Reset current area *	Stereo on/off	Reset spin function
2	Global	Quadrant on/off	Send global matrices **
3	Use area 1	Display area 1	Send area 1 matrices **
4	Use area 2	Display area 2	Send area 2 matrices **
5	Use area 3	Display area 3	Send area 3 matrices **
6	Use area 4	Display area 4	Send area 4 matrices **
7	Use area 5	Display area 5	Send area 5 matrices **
8	Use area 6	Display area 6	Send area 6 matrices **
9	Use area 7	Display area 7	Send area 7 matrices **
10	Use area 8	Display area 8	Send area 8 matrices **
11	Rot x 90 current area	Display protein id bar	Send spin matrix
12	Rot y 90 current area	Display protein prop. bar	

* Current area is current area being manipulated, or Global.

** Matrices are for rotation, translation, and spin networks. They are usually requested as needed by the host, and are connected to the function keys only for convenience of programming.

translation or translation of any given area, both with a function key for reset; whole picture or area rotation with function key for reset, and rotation in either x or y by 90°.

Areas are activated for display or no display by toggling shift/function keys. Picking and menu selection networks allow the activation of specific modules (see Table 2) and a 'spin about any axis' network allows acyclic and cyclic bonds to be rotated.

The Leeds Stereo Viewer is installed [43] and a quadrant mode view is provided with a Z-clip depth indicator. Table 3 defines the function key actions.

2. ASTRAL PS330 routines

Programme SUPERMOL provides facilities for reading and writing molecular data files, display of area contents, removal of area contents, and transfer of fragments between areas. In future it will also contain routines for manipulation of residues within proteins. It allows display of small molecules in single colour or split-bond formalisms, with or without hydrogens. Proteins may be displayed in single colour, split bond or residue property/type formalisms, with or without hydrogens, with or without backbone or alphacarbons. Multi-fragment areas are displayed either individually according to atom type or colour coded according to energy, with or without hydrogens. Multifragment areas can be manipulated by transperring picked fragments between areas, and the animation feature simplifies the analysis of, for example, dynamics runs where each fragment of a multi-fragment file represents a time-step conformation. Single fragment areas can be manipulated by extracting sets of atoms for transfer to a new area. The extraction options include:

(1) All atoms within a given radius of a picked atom;
(2) All atoms within a given radius of a line drawn between two picked atoms;
(3) All atoms connected to a picked atom;
(4) All atoms connected to and extending in one direction from two picked bonded (acyclic) atoms;

TABLE 3

* ASTRAL *

A-dvanced S-ystem T-o R-emove A-ll L-imitations

A. Davis J.G. Vinter 1987

Chemical display package for Evans & Sutherland Picture System 300

*Top Level Menu:

NANDMENU	Links to ASTRAL facilities by a menu
STOP	-returns to DCL
SUPERMOL	-chains to molecular manipulation routines
INTERMOL	-chains to docking calculation routines
SURFACES	-chains to contouring routines
MEASURE	-chains to measuring routines
MO DISPLAY	-chains to molecular orbital display

*Second Level Menu for SUPERMOL -Molecular data file manipulations:

EXIT	returns to NANDMENU
READ.DAT	-reads COSMIC.DAT file -chains to DISPLAY
READ.AST	-reads ASTRAL.AST file -chains to DISPLAY
WRITE.DAT	writes a COSMIC.DAT file
DISPLAY	-displays the contents of area n
COLOUR	-for protein structures (in ASTRAL.DOC)
REMOVE ALL	-clear a specified area of all information
REMOVE FRG	-removes frag(s) from display for area n
SPIN	-activates the spin network
TRANSFER	-transfers frags from areas to a new area
EXTRACT	-extracts atoms from frags in an area
ANIMATE	-animates an astral file for display of dynamics etc.
HARD COPY	-writes screen to a hard copy device

*Second Level Menu for INTERMOL -To calculate intermol potential energy

EXIT	-Returns to the NANDMENU
MONITOR	-Manual 'dock' facility
SIMPLEX	-Automatic 'dock' facility

*Second Level Menu for SURFACES -Display isopot maps and VDW surfaces

EXIT	-returns to the NANDMENU
READ.V3D	-reads and displays
DISPLAY	-on/off switch display parts
ADD.V3D	-adds or subracts two isopotential maps
REMOVE	removes the fragment(s) from display
VDW MAP	-contour one or more molecules

*Second Level Menu for MEASURE -Bond lengths, angles and torsional angles

EXIT	-returns to the NANDMENU
LABEL	-display molecule with atom labels
DIST	-calculate distance between 2 atoms
ANGLE	-calculate angle between 3 atoms
TORSION	-calculate torsional angle between 4 atoms

*Second Level Menu for MO DISPLAY via ORBIT -Results of CNDO, MOPAC & GAUSSIAN

EXIT	-returns to the NANDMENU
READ	-reads in a data set
DISPLAY	-toggles display of any orbitals in any/all of 8 areas

(5) All connected atoms between two picked atoms.

Programme INTERMOL allows interactive energy monitoring between molecules in up to eight different areas, and includes a simplex rigid-body docking routine.

Isopotential surfaces and molecular orbitals are displayed by the SURFACES programmes which allow the display of properties of molecules in up to eight areas simultaneously. Picking a molecule on the screen, followed by picking an appropriate menu item, will switch on or off the property associated with that menu item. Van der Waals surfaces are generated on a 41*41*41 grid which allows users to contour at a given distance from the atoms and then perform logical operations on the surfaces for the combined sets of molecules as oriented on the screen. The resulting maps can be displayed in any colour.

CONLUSION

The COSMIC software framework has been designed as a molecular computational interface for chemical scientists. It eliminates the need for extensive knowledge of computational techniques and is expandable to include any computational methodology that may be available.

The complementary ASTRAL framework has been developed for complex visual investigations of COSMIC derived data and large molecule crystal structures. At present, the graphics for ASTRAL are processed on the Evans & Sutherland PS330 terminal.

The driving force behind the COSMIC/ASTRAL design has been the need for accessible, comprehensive computational chemistry. The low resolution graphics have been developed only to activate computational processes, the results of which are fed to high resolution graphics devices for visual investigation. In this way, the low level front end graphics devices, which may be proliferated at moderate cost, remove the workload from the much more expensive high resolution terminals.

The system is designed to investigate small molecule interactions. For drug design, the investigation of conformational space, electrostatic interaction, docking, solvation, etc. have been shown to provide a valuable input to drug design. Extensions to incorporate metals and experts systems are underway. Proteins and other large molecules will be included when we are satisfied that the required computational methods are adequate, and when general computing power increases.

REFERENCES

1 Tripos Associates Inc., 6548 Clayton Road, St. Louis, Missouri 63117, U.S.A.

2 Davies, E.K. and Prout, C.K., CHEMGRAF User Manual, Chemical Crystallography Laboratories, Oxford University, U.K., 1982.
COSMIC only supports CHEMGRAF which is an old version of the current offering called CHEM-X from Chemical Design Ltd., Oxford, U.K.

3 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535–542.

4 The help of Dr. Nicole Van Opdenbosch is acknowledged in the construction of this module.

5 Pearlman, R.S., QCPE Bull., 1 (1981) 16. (SAREA, QCPE 413)

6 Edward, J.T., J. Chem. Ed., 47 (1970) 261–270.

7 Smith, G.H., QCPE Bull., 6 (1986) 13. (MOLSV, QCPE 509)

8 Pople, J.A. and Beveridge, D.L., Approximate Molecular Orbital Theory, McGraw-Hill Inc., New York, 1970.

9 Dewar Research Group, QCPE Bull., 6 (1986) 4. (AMPAC, QCPE 506)

10 Singh, U.C. and Kollman, P., QCPE Bull., 2 (1982) 117. (GAUSSIAN 80, QCPE 446)
11 *a*. Abraham, R.J., Griffiths, L. and Loftus, P., J. Comp. Chem., 3 (1982) 407–416.
b. Abraham, R.J. and Hudson, B.D., J. Comp. Chem., 5 (1984) 562–570.
c. Abraham, R.J. and Hudson, B.D., J. Comp. Chem., 6 (1985) 173–181.
12 Dickens, T.K., Prout, C.K. and Saunders, M.R., J. Mol. Graph. (submitted)
13 Giessner-Prettre, C. and Pullman, A., Theoretica Chimica Acta (Berlin), 25 (1972) 83–88. (Modified from VSS, QCPE 249.)
14 Chodosh, D.F. and Mendelson, W.L., Pharm. Technol., 7 (1983) 90–92.
15 *a*. Corey, E.J. and Wipke, W.T., Science, 166 (1969) 178–192.
b. Corey, E.J., Wipke, W.T., Cramer, R.D. and Howe, W.J., J. Amer. Chem. Soc., 94 (1972) 421–430.
16 Wolfson CADOS Unit, ORAC 6.0/VMS 4.2, University of Leeds, U.K., 1986.
17 Medicinal Chemistry Project, Pomona College, Claremont, CA, USA; MEDCHEM, Release 3.41, 1986.
18 Allen, F.H., Bellard, S.H., Brice, M.D., Cartwright, B.A., Doubleway, A., Higgs H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.A. and Watson, D.G., Acta Cryst. B35 (1979) 2331–2339.
19 Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC 20007, U.S.A.
20 Fraser Williams (Scientific Systems) Ltd., London House, London Road South, Poynton, Cheshire SK12 1YP, U.K., 1985.
21 Still, C., MacroModel Version 1.1, Columbia University, Department of Chemistry, 507 Havemeyer Hall, New York, NY 10027, U.S.A., 1986.
22 Smith, G.M. and Gund, P., J. Chem. Inf. & Comp. Sci., 18 (1978) 207–210.
23 Leach, A.R., Software written for SK&F whilst working with Bradford Software Services Ltd. and with C.K. Prout, Chemical Crystallography, University of Oxford, U.K., 1985–1986.
24 Johnson, C.K., ORTREP II, Oak Ridge National Laboratory, Oak Ridge, Tennessee, Rep. ORNL-3704.
25 Baumann, H., QCPE, 11 (1977) 333. (CNDUV99)
26 Saunders, M.R., Software written at SK&F, 1986. (See main text.)
27 White, D.N.J. and Pearson, J.E., J. Mol. Graph., 4 (1986) 134.
28 White, D.N.J., Computers & Chemistry, 1 (1977) 225–233.
29 Nelder, J.A. and Mead, R., Comput. J., 7 (1965) 308–313.
30 Fletcher, R. and Reeves, C.M., Comput. J., 7 (1964) 149. (NAGFLIB: 1287/209: Mk10:1982)
31 *a*. Stewart, J.J.P., Csaszar, P. and Pulay, P., J. Comp. Chem., 3 (1982) 227–228.
b. Beppu, Y. and Ninomiya, I., Computers & Chemistry, 6 (1982) 87–88.
32 Butland, J., SIMPLEPLOT Mark 2, Bradford University Software Services Ltd., 16 Campus Road, Bradford, West Yorkshire BD7 1HR, U.K.
33 Quantum Chemistry Program Exchange, Chemistry Department, Indiana University, Bloomington, Indiana 47405, U.S.A.
34 Perkin, S.J., In Berliner, L.J. and Reuben, J. (Eds) Biological Magnetic Resonance, Vol. 4, Plenum Press, New York, 1982, Chapter 4.
35 Pople, J.A., J. Chem. Phys, 24 (1956) 1111.
36 Johnson, C.E. and Bovey, F.A., J. Chem. Phys, 29 (1958) 1012–1014.
37 Haigh, C.W. and Mallion, R.B., Org. Magn. Reson., 4 (1972) 203–208.
38 White, D.N.J. and Bovill, M.J., J. Chem. Soc. Perkin II, 12 (1977) 1610–1623.
39 Hill, T.L., J. Chem. Phys., 16 (1948) 399–404.
40 *a*. Allinger, N.L., In Gold, V. and Bethell, D. (Eds) Advances In Physical Organic Chemistry, Vol. 13, Academic Press, 1976, p. 17–93.
b. Allinger, N.L., Miller, M.A., Van Catledge, F.A. and Hirsch, J.A., J. Am. Chem. Soc., 89 (1967) 4345–4357.
c. Allinger, N.L., Hirsch. J.A., Miller, M.A. and Tyminski, I.J., J. Am. Chem. Soc., 91 (1969) 337–343.
d. Allinger, N.L., Tribble, M.T., Miller, J.A. and Wertz, D.H., J. Am. Chem. Soc., 93 (1971) 1637–1648.
41 Bondi, A., J. Phys. Chem., 68 (1964) 441–451.
42 Lifson, S., Hagler, A.T. and Dauber, P., J. Am. Chem. Soc., 101 (1979) 5111 and J. Am. Chem. Soc., 96 (1974) 5319–5327.
43 Leeds Liquid Crystal 3-D Viewer supplied by Millennium, Boulton Road, Stevenage, Herts SG1 4QX, U.K.

APPENDIX I

COSMIC MOLECULAR MECHANICS POTENTIALS AND THEIR CURRENT FORCE FIELDS

Introduction

The potential functions were chosen and put together to enable a general molecular mechanics utility to be developed. This combination has been tested over 10 years and has proved to be of value within both COSMIC and other commercially available packages. The essential criterion for design and development has been the need to handle a wide variety of chemical entities. To this end, the force fields have been simplified as far as possible in order to allow users to modify and extend them to suit their problems. Atom types have been kept to a minimum and some care and experience is needed to allocate these correctly (Appendix 3). Because of this generality, some structural features may not be well represented without specific changes to the force field parameters. Absolute energy values are, as usual, not to be taken seriously but energy differences have been shown to compare well with physical results and molecular orbital methods. The present force fields have been built to satisfy the needs so far encountered and are in no way considered to be exhaustive.

a) Bond stretching potential

$E_b = {}^{k}/_{2}\,(\Delta l)^2$ for each bond

k = stretch force constant (kcal mole $^{-1}$ angstrom^{-2})
Δl = difference between actual and equilibrium bond length

COSMIC Bond Force Field (modified and extended from White [38]) (73 entries to date)

At.Ty.	k	l	At.Ty.	k	l	At.Ty.	k	l
1 1	316.8	1.520	1 2	319.5	1.501	1 3	320.0	1.525
1 4	320.0	1.458	1 5	380.0	1.470	1 6	380.0	1.470
1 7	338.0	1.445	1 8	380.1	1.440	1 9	338.8	1.450
1 10	309.4	1.470	1 12	190.8	1.810	1 13	190.8	1.800
1 14	203.8	1.790	1 15	331.2	1.100	1 16	300.0	1.360
1 17	300.0	1.767	1 18	300.0	1.938	1 19	300.0	2.139
2 2	670.0	1.335	2 3	670.0	1.510	2 4	670.0	1.440
2 7	600.0	1.350	2 8	653.0	1.330	2 9	435.1	1.345
2 10	349.9	1.310	2 11	777.6	1.220	2 12	200.0	1.740
2 13	180.0	1.710	2 15	346.0	1.089	2 16	600.0	1.330
2 17	260.0	1.750	3 3	700.0	1.395	3 4	670.0	1.420
3 7	653.0	1.420	3 8	653.0	1.346	3 9	545.0	1.416
3 10	350.0	1.300	3 12	180.0	1.770	3 15	346.0	1.084
3 16	250.0	1.330	3 17	256.7	1.750	3 18	250.0	1.850
3 19	245.0	2.050	4 8	650.0	1.330	4 23	800.0	1.158

At.Ty.	k	l	At.Ty.	k	l	At.Ty.	k	l
5 5	300.0	1.520	5 6	320.0	1.480	5 15	331.2	1.000
6 6	366.0	1.470	6 12	180.0	1.610	6 15	346.0	1.040
7 8	400.0	1.380	7 14	400.0	1.710	7 15	346.0	1.030
7 19	400.0	2.170	7 23	600.0	1.230	8 8	653.0	1.346
8 9	333.8	1.440	8 12	200.0	1.680	8 15	346.0	1.000
9 9	372.2	1.450	9 11	777.6	1.280	9 12	180.0	1.590
9 15	350.0	1.000	9 22	560.0	1.240	10 10	586.1	1.480
10 14	400.0	1.600	10 15	503.8	0.950	10 12	200.0	1.500
11 12	800.0	1.450	12 12	300.0	2.030	14 24	700.0	1.430
23 23	600.0	1.130						

b) Bond angle deformation potential

$E_a = {}^{k}/_{2}\,(\Delta\theta)^2$ for each angle

k = bending constant (kcal mole^{-1} degree^{-2})
$\Delta\theta$ = difference between actual and equilibrium bond angle

COSMIC Angle Force field (modified and extended from White [38]) (195 entries to date)

Atom Type	k × 100	θ	Atom Type	k × 100	θ	Atom Type	k × 100	θ
1 1 1	1.2	109.5	1 1 2	0.9	109.5	1 1 3	1.2	109.5
1 1 4	1.2	109.5	1 1 5	1.2	109.5	1 1 6	0.9	109.5
1 1 7	1.0	109.5	1 1 8	0.9	109.5	1 1 9	0.9	109.5
1 1 10	1.1	109.5	1 1 12	0.9	107.8	1 1 14	0.7	112.0
1 1 16	1.0	109.5	1 1 17	1.0	109.5	1 1 18	1.0	109.5
1 1 19	1.0	109.5	1 2 1	2.3	116.4	1 2 2	1.2	121.0
1 2 3	2.0	120.0	1 2 8	1.0	118.0	1 2 9	1.0	117.0
1 2 10	1.0	120.0	1 2 10	3.6	120.0	1 2 11	1.3	120.0
1 2 12	1.3	120.0	1 2 13	3.0	122.0	1 3 3	1.2	120.0
1 3 7	2.0	120.0	1 3 8	2.0	120.0	1 3 10	2.0	120.0
1 4 4	1.2	109.5	1 4 23	2.0	180.0	1 5 1	1.2	109.5
1 6 1	0.9	109.5	1 6 12	2.0	109.5	1 7 23	2.0	114.3
1 8 2	4.1	110.0	1 8 3	4.1	110.0	1 8 8	2.0	118.0
1 9 1	2.0	122.0	1 9 2	2.2	118.0	1 9 3	2.2	118.0
1 9 8	1.2	120.0	1 9 9	1.2	120.0	1 9 22	1.0	120.0
1 10 1	2.2	115.0	1 10 2	2.2	115.0	1 10 3	1.0	116.0
1 10 10	4.7	104.0	1 10 14	0.5	120.0	1 12 1	1.0	98.0
1 12 2	1.1	94.5	1 12 3	1.0	105.0	1 12 6	2.0	111.0
1 12 10	2.0	108.0	1 12 11	2.0	108.0	1 14 24	0.7	110.0
2 1 2	0.9	109.5	2 1 3	0.9	109.5	2 1 5	1.2	109.5
2 1 6	0.9	109.5	2 1 7	0.9	109.5	2 1 9	1.1	109.5

Atom Type			k × 100	θ	Atom Type			k × 100	θ	Atom Type			k × 100	θ
2	1	10	1.1	109.5	2	1	12	0.9	107.8	2	2	2	0.9	122.0
2	2	3	1.3	120.0	2	2	4	1.2	120.0	2	2	5	1.2	120.0
2	2	6	1.2	120.0	2	2	7	1.2	120.0	2	2	8	1.2	122.0
2	2	9	1.2	114.5	2	2	10	3.6	120.0	2	2	11	1.3	125.0
2	2	12	2.5	92.0	2	2	13	3.0	120.0	2	2	17	1.8	120.0
2	3	8	2.0	120.0	2	4	23	2.0	180.0	2	7	23	2.0	114.0
2	7	8	3.0	120.0	2	8	2	4.1	120.0	2	8	3	4.0	123.0
2	8	4	2.0	120.0	2	8	7	3.0	120.0	2	8	8	2.2	112.0
2	8	9	2.2	112.0	2	8	12	2.0	115.0	2	9	2	0.9	120.0
2	9	3	2.6	120.0	2	9	8	1.5	127.0	2	9	9	0.9	120.0
2	9	22	1.2	120.0	2	10	2	1.0	116.0	2	10	3	1.0	116.0
2	12	11	3.0	108.0	2	12	2	3.0	91.0	3	1	3	0.9	109.5
3	1	6	0.9	109.5	3	1	7	1.0	109.5	3	1	9	1.0	109.5
3	1	10	0.9	109.5	3	1	12	0.9	108.0	3	2	3	1.2	120.0
3	2	4	1.2	120.0	3	2	8	2.0	120.0	3	2	9	2.0	120.0
3	2	10	1.5	120.0	3	2	11	1.3	120.0	3	2	17	1.8	120.0
3	3	3	1.2	120.0	3	3	2	1.2	120.0	3	3	4	1.3	120.0
3	3	7	2.0	120.0	3	3	8	2.0	120.0	3	3	9	3.1	120.0
3	3	10	3.1	120.0	3	3	12	3.1	118.0	3	3	16	1.8	120.0
3	3	17	1.8	120.0	3	3	18	1.8	120.0	3	3	19	1.8	120.0
3	4	23	2.0	180.0	3	7	23	2.0	114.0	3	7	3	2.0	120.0
3	8	3	2.0	120.0	3	8	7	2.0	118.0	3	8	8	2.0	118.0
3	8	9	2.0	118.0	3	9	3	2.2	120.0	3	9	8	1.0	120.0
3	9	11	2.2	120.0	3	9	22	1.2	120.0	3	9	9	2.6	120.0
3	9	12	2.0	115.0	3	10	3	1.0	116.0	3	12	3	3.1	97.0
3	12	6	2.0	111.0	3	12	9	2.0	111.0	3	12	11	2.0	108.0
3	12	12	1.5	103.0	4	2	8	3.5	123.0	6	2	8	3.5	123.0
6	6	1	2.0	106.5	6	12	9	2.0	112.0	6	12	11	2.0	107.0
7	2	8	3.0	120.0	7	2	11	1.5	121.0	7	3	7	2.0	120.0
7	3	8	2.0	120.0	7	3	9	2.0	120.0	7	23	23	3.0	172.0
8	2	8	2.0	120.0	8	2	9	3.5	123.0	8	2	11	1.3	120.0
8	2	12	1.4	125.5	8	3	8	2.0	120.0	8	3	9	2.0	118.0
8	3	10	2.0	118.0	8	3	12	2.0	126.0	8	3	23	2.0	180.0
8	12	1	1.0	108.0	8	12	8	1.0	108.0	8	12	11	1.0	108.0
9	1	9	2.0	110.0	9	1	6	2.0	110.0	9	1	12	1.0	109.0
9	2	9	1.5	120.0	9	2	10	0.7	110.5	9	2	11	1.5	123.0
9	2	12	1.5	111.5	9	2	13	2.5	120.0	9	3	9	1.5	120.0
9	12	11	2.0	108.0	10	1	9	1.0	109.5	10	1	10	1.0	109.5
10	1	12	1.0	108.0	10	2	11	1.5	120.0	10	12	11	1.5	120.0
10	14	10	0.9	110.0	10	14	24	0.9	124.0	11	2	12	0.8	125.0
11	12	11	3.0	118.0	11	2	11	3.0	120.0	12	9	12	2.0	115.0
14	7	23	2.0	114.0	14	10	14	2.0	130.0	16	1	16	2.0	109.5
17	1	17	1.0	109.5	17	2	17	1.5	122.0	19	7	23	2.0	114.0
22	9	22	1.0	120.0	24	14	24	1.0	120.0	24	14	24	0.7	114.5

c) Periodic torsional barrier potential

$E_t = {}^{V}/_{2}\,(1 + \cos \# \omega)$ for each torsional angle

V = barrier to rotation (kcal mole^{-1})
ω = torsional angle (radians)
= rotation degeneracy, sign indicates eclipsed/staggered preference

COSMIC Torsional Force Field (modified and extended from White [38]) (53 Entries to date)

At.Ty.	V	#	At.Ty.	V	#	At.Ty.	V	#
1 1	0.10	3.0	1 2	0.06	-3.0	1 3	0.06	-3.0
1 4	0.00	1.0	1 5	0.10	3.0	1 6	0.10	3.0
1 7	0.20	-3.0	1 8	0.20	-3.0	1 9	0.10	3.0
1 10	0.60	3.0	1 12	0.20	3.0	1 14	0.20	3.0
2 2	6.00	-2.0	2 3	3.00	-2.0	2 4	0.00	1.0
2 5	0.06	-3.0	2 6	0.06	-3.0	2 7	6.00	-2.0
2 8	6.00	-2.0	2 9	3.23	-2.0	2 10	2.90	-2.0
2 12	0.50	-2.0	2 14	0.50	-2.0	3 3	3.00	-2.0
3 4	0.00	1.0	3 5	0.06	-3.0	3 6	0.06	-3.0
3 7	2.00	-2.0	3 8	0.80	-2.0	3 9	0.80	-2.0
3 10	0.60	-2.0	3 12	0.50	3.0	3 14	0.50	3.0
4 4	0.00	1.0	5 5	0.10	3.0	5 6	0.10	3.0
5 9	0.06	-3.0	5 10	0.10	3.0	6 6	0.13	3.0
6 9	0.06	-3.0	6 10	0.10	3.0	6 12	0.20	3.0
7 7	0.80	-2.0	7 8	0.80	-2.0	8 4	0.00	1.0
8 8	0.80	-2.0	8 9	0.80	-2.0	8 12	0.50	-2.0
9 9	0.80	-2.0	9 12	0.50	-2.0	10 10	0.37	3.0
10 14	0.20	3.0	12 12	0.10	3.0			

d) Non-bonding interactions for all 1–4 or greater non-bonded atom pairs

i) Van der Waals potential [39]

$E_w = f[-2.25/rr^6 + 8.28 \times 10^5 . e^{(-rr/0.0736)}]$

r = distance between atoms
rr = r/(sum of van der Waals radii)

Hill *f* parameters for van der Waals potential (extended from Hill [39] and Allinger [40])

Atom	H	C	N	O	F	S	Cl	Br	I	H(protonic)
vdW Rad(41)	1.20	1.70	1.55	1.52	1.47	1.80	1.75	1.85	1.98	0.2
H	0.042	0.067	0.063	0.069	0.068	0.115	0.115	0.136	0.162	0.0
C		0.107	0.100	0.111	0.108	0.183	0.183	0.215	0.258	0.0
N			0.095	0.105	0.102	0.172	0.172	0.203	0.243	0.0
O				0.116	0.112	0.190	0.190	0.224	0.268	0.0
F					0.109	0.185	0.185	0.217	0.259	0.0
S						0.314	0.314	0.369	0.442	0.0
Cl							0.314	0.369	0.442	0.0
Br								0.434	0.522	0.0
I									0.623	0.0
H(protonic)										0.0

ii) Coulombic potential

$E_c = q_1.q_2/D.r$

q = partial change (from CNDO, INDO, MNDO, GAUSSIAN80 or Liverpool Method)
r = distance between charged atoms
D = local dielectric constant set to unity in most cases

iii) Hydrogen bonding

A 'protonic hydrogen' attached to an electronegative atom is given a small or zero van der Waals contribution. A hydrogen bond need not be treated explicitly when the van der Waals potential and the coulombic potential (using reasonable partial charges) are combination [42].

APPENDIX 2

ATOM TYPES USED IN COSMIC & ASTRAL (Extended from White [38])

1	C sp3		
2	C sp2		
3	C aromatic		
4	C sp		
5	N sp3+	tetrahedral	4 valent
6	N sp3	tetrahedral	3 valent
7	N(N2)	C-N azide	2 valent
8	N sp2	trigonal	2 valent
9	N sp3	trigonal	3 valent
10	O sp3	single	2 valent
11	O sp2	double	2 valent
12	S sp3	tetrahedral	4–6 valent
13	S sp2	trigonal	2 valent
14	P	phosphate	
15	H		
16	F		
17	Cl		
18	Br		
19	I		
20	General (metals, untyped, etc.)		
22	O sp2/3 (NO2, NO3-, etc.)		
23	N sp (CN/NN)		
24	O sp2 (O(PO)) uncharged		
28	lp: lone pair		
29	Ass? Unassigned		

APPENDIX 3

THE COSMIC/ASTRAL DATA FILES

1) The COSMIC data file (.DAT) format is as follows: Title line of up to 80 characters; Integer number of bonds (nnobs); Half-connectivity vectors as bonded atom pairs (20I4 to nnobs); Integer number of atoms (natoms); Atomic number, X coordinate, Y, Z, Atom Type (Appendix 2), Partial Charge, Residue name, Residue number, Backbone atom flag; (I6, 3F10.6,2X,I6,F8.4,X,A3,I4,I2 to natoms -last three only if peptide derivative).

   Associated with every .DAT file, there is a .CMT file with the same name. This contains titles, creation and update times, creator identification and an unrestricted space for alphanumeric comments.

   It is possible to log all events in a session through a .PRJ project file. Information on times, modules entered, batch jobs initiated and essential results is stored. This editable file is used as a notebook substitute.

2) VAX file extensions used in COSMIC; .DAT = COSMIC data;.XR = CHEMGRAF data;-.MOL = SYBYL data;.DSP = SYBYL display;.PDB = Brookhaven Protein data;.MN-1 = MIN01 multiple conformers;.SPN = SPIN01 multiple rotamers/conformers;.CND & .MOS = CNDO output;.MOM & .MOP = AMPAC;.MOG & G80 = GAUSSIAN80;.AS-T = ASTRAL multiple structures (see below);.V3D = 3D EIP contours and minima.

3) The ASTRAL files are as follows: .AST(local .MOL) used to communicate between COSMIC and ASTRAL (sequential access, unformatted, each record contains information on one molecule or fragment);.ATI files containing information on the atoms displayed (one per area, unformatted);.INF files one per area, containing summarised information on the area display segments (direct access, unformatted, recl = 20).