ELSEVIER

ALTER: Eclectic management of molecular structure data

Darren R. Flower

Department of Physical and Metabolic Sciences, Astra Charnwood, Loughborough, Leicestershire, UK

ALTER is a computer program written to facilitate easy conversion between different representations of molecular structure data. The program functions as a file converter, data generation engine, and through the creation of control or input files, as an interface to other programs. The main aspects of program function—the reading and writing of files; coordinate transformation; data reorganization; structure building; data abstraction, including the generation of a wide variety of topological indices and constitutional descriptors; and display—are described in appropriate detail. © 1997 by Elsevier Science Inc.

Keywords: molecular representation, file format, data conversion, topological indices, ray tracing

INTRODUCTION

Molecular structure data comprise the principal commodity in an information economy that encompasses many distinct scientific disciplines: molecular physics, synthetic chemistry, and molecular biology. The independent development of these, and other relevant fields of scientific endeavor, has led to the incompatible representation of what are, essentially, equivalent data. Each such representation is, generally, an incomplete one, capturing some, but not all, aspects of a structure. By way of example, it is possible to express the structure of a given molecule, a small peptide, say, in a variety of ways: Figure 1 illustrates pictorially a subset of possible representations.

Each of the many possible representations, however complex, is no more than a subset of the complete description of a structure. Because it is possible to represent a molecule in so many different ways it is convenient to classify these different representations. In this context, the concept of the notional dimensionality of molecular data is a useful one because of its ubiquity.

One-dimensional (1D) data: Primarily macromolecular sequence data, particularly biopolymers such as proteins and nucleic acids, drawn from a limited alphabet of monomers

Two-dimensional (2D) data: Chemical connectivity data such as is used in chemical information technology and familiar to the synthetic organic chemist. Often, and erroneously, this term is taken as synonymous with the concept of a molecular graph. Representing a chemical structure as a graph, where vertices represent atoms and edges represent bonds, dates back almost to the founding of modern organic chemistry. However, strictly speaking, graphs do not have a dimensionality; as dimensionless descriptions of molecular structure molecular graphs might perhaps be better called 0-dimensional, or 0D, data

Three-dimensional (3D) data: Three dimensional coordinates produced by crystallography, multidimensional nuclear magnetic resonance (NMR), and molecular modeling

The problem of representation has been much compounded in the computer age by the development of competing software systems that record and express even identical data in a mutually inconsistent manner. Computer programs capable of accessing molecular data have proliferated. For example, structural databanks, such as PDB[1] or CSD[2]; chemical database search systems; molecular modeling packages; computational chemistry software, such as molecular mechanics or molecular orbital programs; and protein or nucleic acid sequence analysis packages have all generally tended to create their own proprietary, and typically incompatible, data formats tailored to meet their own specific needs. Although the primary objective of creating efficient working systems is well met by this strategy, it does not allow for the easy exchange or combination of data, nor the straightforward integration of software. This phenomenon is not restricted to molecular data: the most widely encountered examples include image files and word processor documents.

The undesirability of this situation is widely acknowledged, and has led to several attempts to define a standard file format capable of storing an arbitrary representation of molecular structure. The SMD format[3] is a good example of such an attempt. The development of CIF[4] and Chemical MIME,[5] etc., is continuing this tradition. Laudable though these undertakings are, thus far at least, not one of them has successfully

Color Plates for this article are on pages 181–184.

Address reprint requests to: Darren R. Flower, Department of Physical and Metabolic Sciences, Astra Charnwood, Bakewell Road, Loughborough, Leicester, UK LE11 5RH.

Received 4 February 1997; revised 8 May 1997; accepted 13 May 1997.

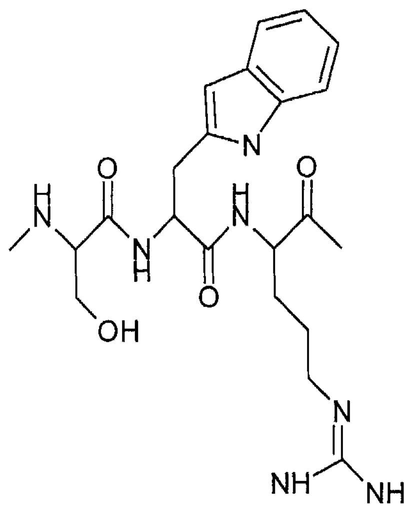
Journal of Molecular Graphics and Modelling 15, 161–169, 1997
© 1997 by Elsevier Science Inc.
655 Avenue of the Americas, New York, NY 10010

1093-3263/97/$17.00
PII S1093-3263(97)00101-0

1D -SWR- -SER-TRP-ARG-

NC(CO)C(=O)NC(Cc1cc2ccccc2n1)C(=O)NC(CCCN=C(N)N)C(=O)

2D



```
33 34
  11.5166     9.9112     0.0000 C
  12.2944     9.1334     0.0000 C
  13.6384    10.4774     0.0000 C
  12.8606    11.2552     0.0000 C
  13.3510     9.4165     0.0000 C
  11.7954    10.9721     0.0000 C
  10.5048     9.5082     0.0000 C
  10.5665     8.4160     0.0000 C
  11.6254     8.1384     0.0000 N
   9.7888     6.5334     0.0000 C
  10.5666     5.7556     0.0000 C
  10.5611     4.6556     0.0000 O
  11.3444     6.5390     0.0000 NH
  12.1222     5.7612     0.0000 C
   9.0055     5.7556     0.0000 NH
   9.7833     7.6334     0.0000 C
  12.9000     6.5445     0.0000 C
  13.6777     5.7667     0.0000 C
  12.1166     4.6612     0.0000 C
  12.8944     3.8834     0.0000 C
  12.8888     2.7834     0.0000 C
  13.6666     2.0056     0.0000 N
  13.6611     0.9056     0.0000 C
  14.4388     0.1278     0.0000 NH
  12.7055     0.3612     0.0000 NH
   8.2222     6.5390     0.0000 C
  12.8944     7.6445     0.0000 O
   8.2166     7.6390     0.0000 O
   7.4388     5.7612     0.0000 C
   6.6555     6.5445     0.0000 NH
   7.4333     4.6612     0.0000 C
   8.2111     3.8834     0.0000 OH
   5.8722     5.7667     0.0000 C
 1  2  2  1
 3  4  1  1
 1  6  1  1
 2  5  1  1
 5  3  2  1
 4  6  2  1
 8  9  1  1
 7  8  2  1
 9  2  1  1
 1  7  1  1
10 11  1  1
11 12  2  1
11 13  1  1
13 14  1  1
10 15  1  1
10 16  1  1
14 17  1  1
17 18  1  1
14 19  1  1
19 20  1  1
20 21  1  1
21 22  1  1
22 23  2  1
23 24  1  1
23 25  1  1
15 26  1  1
17 27  2  1
26 28  2  1
26 29  1  1
29 30  1  1
29 31  1  1
31 32  1  1
16  8  1  1
30 33  1  1
```

3D

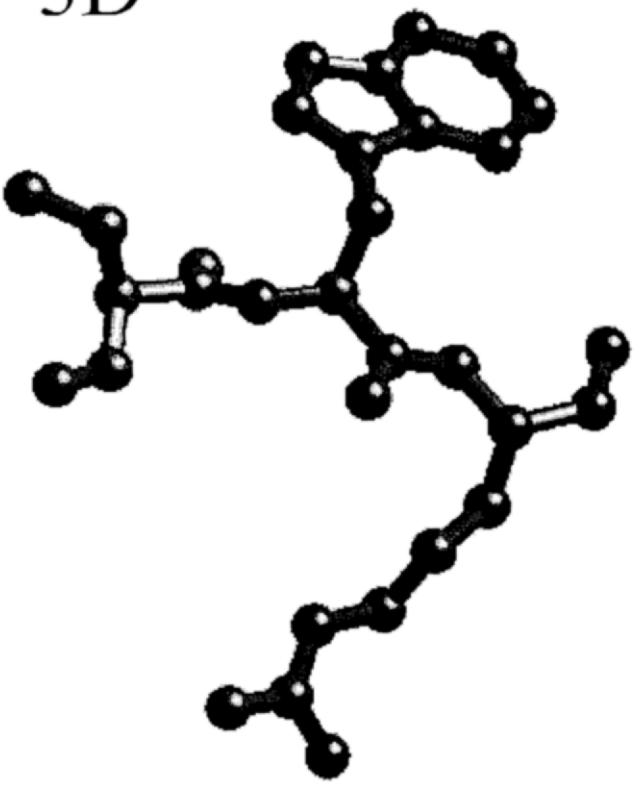


Figure 1. A pictorial illustration of a subset of possible representations of a small peptide.

established itself as a universally used standard to which all software adheres. The day when this is finally achieved cannot come too soon.

The need to convert freely between different forms of expression and different formats is clear, and is widely acknowledged. This need is encountered frequently as we seek to utilize molecular structure data deriving from many sources. For example, when one is creating databases it is often necessary to concentrate or combine data of different forms and formats into some common computer-based representation. Likewise, one may wish to analyze a data set using many different methods implemented in many different pieces of software, each requiring its own input format. We describe here our experiences of such eclectic data management, as exemplified in the program ALTER.

ALTER functions as a file converter, as an interface to other software, and as a data generation engine. A number of programs have been described that address some of the same needs. Most molecular modeling packages and database systems are, at least to some extent, able to import and export different file formats; the number supported is, however, sometimes limited. Lesk has described a program—MICROYFON—that uses a general parsing approach to the import of molecular data from different formats of data file.[6] In the area of protein structure analysis, manipulation, and display, there is some overlap between the facilities offered by ALTER and other programs such as NAOMI[7] or VMD.[8] RDSEQ is a widely used program for interconversion between different protein and nucleic acid sequence file formats.[9] However, the program BABEL[10] is perhaps the closest in conception, and in realization, to ALTER. It is able to convert molecular data between a variety of different data formats encountered in the field of molecular modeling, but lacks many of the useful features of ALTER: structure visualization, data generation, and data transformation. However, because its design strategy and user interface differ somewhat from those of ALTER, the two programs are perhaps best seen as complementary.

ALTER has developed in a somewhat haphazard manner to meet individual research needs. Although the program is not, in itself, a total or exhaustive solution to the problems of data conversion, it has a wide range of useful functionality. ALTER has not been designed as a product, or to fulfill a grand plan; nonetheless many useful insights have been gained through its development and its application. The following sections describe the main features of the program.

OVERVIEW

ALTER is a computer program written to facilitate interconversion between different representations of molecular structure data. ALTER is not an interactive molecular modeling program as such, although it shares with such software certain of the same features. The community is already well served by many excellent examples of such programs, both commercial and public domain. Rather, ALTER is a tool for data management; it has proved particularly useful in the maintenance of chemical structure databases. It functions as a file converter, data generation engine, and through the generation of control or input files, as an interface to other programs.

The basic functioning of ALTER can be summarized as follows: A molecular structure is read from one file, optionally one or more transformations are performed on it, and is then written to another file. Such power as the program possesses comes from its ability to read and write a variety of data file formats, parse different molecular representations, perform a variety of transformations, and to concatenate these operations on multiple files or file entries.

ALTER is written in standard Fortran 77, and is supported on Silicon Graphics workstations running under UNIX and VAX running under VMS. ALTER is controlled, via a simple command line interface. ALTER was developed with a simple command line interface, through a set of keywords, and was originally envisaged to run in the same way under a set of different operating systems: This command line interface approximates to a primitive scripting language which helps facilitate automated conversions (see Figure 2 for examples of line interface). Although the emphasis of the program is on automated processing, the SG version of ALTER supports GL-based visualization of molecular structure, allowing the user to monitor program performance and to preview graphical output.

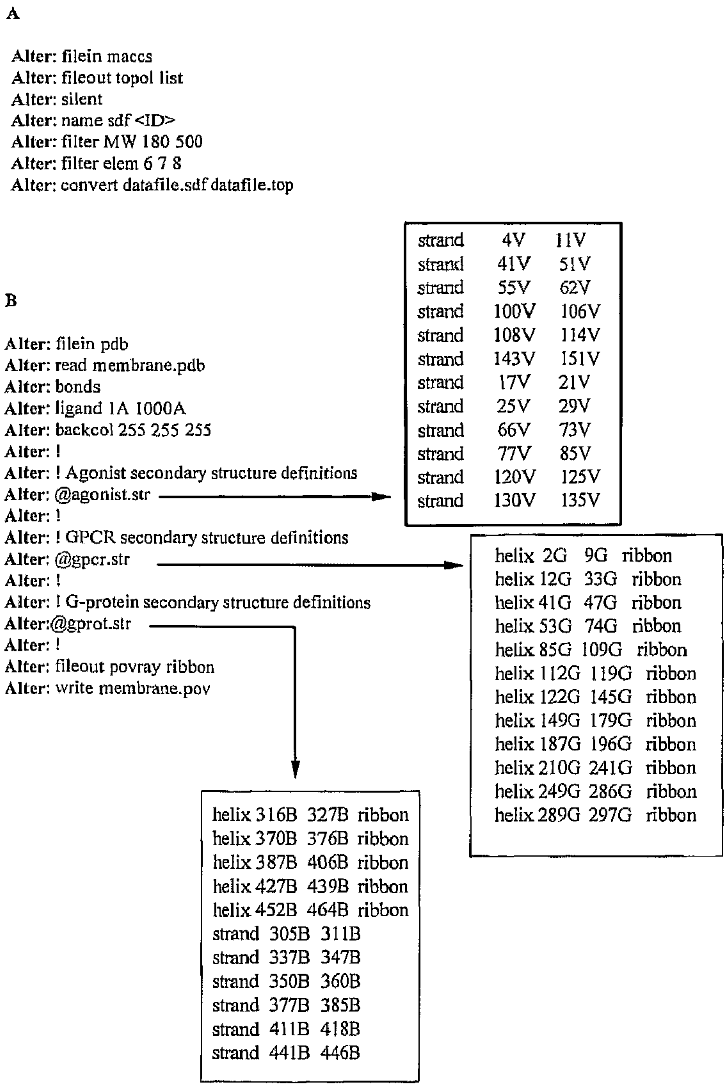
DATA CONVERSION

ALTER concerns itself with three basic operations on molecular structure data: conversion, interchange, and transformation. Interchange is the simplest to understand, it means no more than a change between file formats, and often requires no more than an alteration of layout. The data do not change form nor are they modified. Transformation might be a rigid body rotation, coordinate randomization, or the filtering out of unwanted parts of the data set. Conversion effects a change between equivalent representations. Examples might include expressing a molecular structure read as a connection table or as a SMILES string.[11]

ALTER is able to effect the interconversion of many different forms of data. All conversion and data abstraction proceeds via the internal representation by ALTER of molecular data. For the most part, transformations are effected by the program itself, except for conversion from 2D to 3D data; ALTER devolves this function to external agencies through calls to structure building programs such as CONCORD[12] or CORINA.[13] ALTER proceeds through a series of stages. Data are first imported from one of many forms. A process of data interpretation and perception of molecular features follows to yield the internal representation by ALTER of molecular structure. From here molecular data can be transformed, as necessary, prior to export or the abstraction of derived information. Proceeding through the common intermediate of an internal representation greatly simplifies construction of the program.

READING AND WRITING FILES

ALTER is able to read and write a variety of different file formats. Some of the formats supported by the program are listed in Table 1. Within ALTER, conversion is most often performed between archive data formats, such as a PDB or MACCS[14] file. However, as an alternative, output might take the form of input or control files to some other program, rather than an archive data file. ALTER is also able to output data abstracted from the molecular structure, and pictorial output such as a stereogram. The program also supports options for the free format read of atom data. However, the development of



A

```
Alter: filein maccs
Alter: fileout topol list
Alter: silent
Alter: name sdf <ID>
Alter: filter MW 180 500
Alter: filter elem 6 7 8
Alter: convert datafile.sdf datafile.top
```

B

```
Alter: filein pdb
Alter: read membrane.pdb
Alter: bonds
Alter: ligand 1A 1000A
Alter: backcol 255 255 255
Alter: !
Alter: ! Agonist secondary structure definitions
Alter: @agonist.str
Alter: !
Alter: ! GPCR secondary structure definitions
Alter: @gpcr.str
Alter: !
Alter: ! G-protein secondary structure definitions
Alter:@gprot.str
Alter: !
Alter: fileout povray ribbon
Alter: write membrane.pov
```

```
strand   4V    11V
strand   41V   51V
strand   55V   62V
strand   100V  106V
strand   108V  114V
strand   143V  151V
strand   17V   21V
strand   25V   29V
strand   66V   73V
strand   77V   85V
strand   120V  125V
strand   130V  135V
```

```
helix 2G   9G   ribbon
helix 12G  33G  ribbon
helix 41G  47G  ribbon
helix 53G  74G  ribbon
helix 85G  109G  ribbon
helix 112G  119G  ribbon
helix 122G  145G  ribbon
helix 149G  179G  ribbon
helix 187G  196G  ribbon
helix 210G  241G  ribbon
helix 249G  286G  ribbon
helix 289G  297G  ribbon
```

```
helix 316B  327B  ribbon
helix 370B  376B  ribbon
helix 387B  406B  ribbon
helix 427B  439B  ribbon
helix 452B  464B  ribbon
strand  305B  311B
strand  337B  347B
strand  350B  360B
strand  377B  385B
strand  411B  418B
strand  441B  446B
```

Figure 2. Examples of the command line interface to ALTER. (a) Using ALTER to process a database of chemical structures. A MACCS SD file is converted to a formatted data file with about 200 different topological indices and constitutional descriptors per molecular entry. Filein *and* fileout *set the file types of input and output files.* Silent *switches off program reporting. The command* name *switches on extraction of compound names; ⟨ID⟩ is the datum in the SD file containing the compound name or identifier. The two* filter *commands select compounds with molecular weights between the limits shown and having the atomic numbers listed. The* convert *command executes the one-step database conversion. (b) Using ALTER to generate the POVRAY input file that created Color Plate 2c. Initially a PDB file containing the protein–membrane complex is read. Bonds are calculated using a distance-based algorithm. Residues 1A to 1000A are defined as ligand atoms to be rendered using a small molecule display style (in this case the default of space filling). The background color, for all graphics displays, is set to white using the command* backcol. *The next three commands read, using the @ command, separate subsidiary files containing definitions of the secondary structure elements (as generated by programs such as FOLD*[31]*) in the three separate proteins to be displayed. The contents of the three files are shown diagrammatically. Finally the output format (a POVRAY input file in protein ribbon mode) is defined and the file written using the* write *command. Note use of the pling character to allow comments in the script file.*

a specific ALTER format of molecular data storage file has been consciously avoided.

ALTER allows several ways to import or export a structure file. The first is to read or write a single structure from or to a file. Single structures can be selected from a multiple entry file, such as a MACCS SD file, using the SCAN or SCAN3D options described below. Both commands allow the contents of multientry file to be viewed sequentially. The program is also capable of the automated conversion of multiple, or multiple-entry, files. The CONVERT function will operate on the whole contents of a file or on each file listed in file of filenames; its output can be to a single file or a series of separate files. During the conversion of a data set it is possible to apply filters to the structures being converted: keeping some and rejecting others.

Table 1. Selection of file types supported by ALTER[a]

Type	Flavors	1D	2D*	3D	Other
Crystallographic					
PDB	3	+	∼	+	∼
DIAMOND	3	+	−	+	−
TNT	1	+	∼	+	∼
MERLOT	1	+	−	+	−
HENDRICKSON	1	+	−	+	−
CSD					
FDAT	1	−	−	+	∼
FCON	1	−	+	−	∼
FBIB	1	∼	−	−	∼
AMSON	1	+	−	+	−
Molecular modeling					
SYBYL					
MOL	1	−	+	+	∼
MOL2	1	+	+	+	∼
TRIBBLE	1	−	+	+	−
CSSR	2	∼	+	+	∼
CHEMLAB	1	−	+	∼	−
MACROM	1	∼	+	+	∼
BGF	1	+	∼	+	∼
Chemical database format					
MACCS	1	∼	+	∼	∼
SMD	1	∼	+	∼	∼
SMILES	1	−	+	−	−
TDT	1	∼	+	∼	∼
CLIX-PREMERGE	1	−	+	+	∼
Macromolecular sequence files					
NBRF	1	+	−	−	∼
GCG	4	+	−	−	∼
SWISS	1	+	−	−	∼
STADEN	1	+	−	−	∼
AMBER	1	+	−	−	∼
BIOSYM	1	+	−	−	∼
FASTA	1	+	−	−	∼
TREEALIGN	1	+	−	−	∼
PHYLIP	1	+	−	−	∼
SEQSEE	1	+	−	−	∼
SELEX	1	+	−	−	∼
Control files					
AMBER INPUT FILE	1	−	+	+	∼
MOPAC input file	1	−	−	+	−
POVRAY	1	−	−	−	+

[a] A selection of molecular structure file formats supported by ALTER. Hardcopy graphical output (STEREO, PSPLOT) and data export formats (STERIMOL, TOPOL, HBCOUNT) are also available. *, includes molecular graph (0D); +, always; −, never; ∼, optional.

Options including filtering on molecular weight, number of atoms or bonds, on acceptable element, or length of sequence. Filtering also allows for the deletion of all small fragments from a data set prior to other filtering and output.

PROCESSING

As a preliminary to any conversion or manipulation within ALTER, imported data undergo a process of data interpretation to yield the internal representation by ALTER of molecular structure. This processing step involves the identification or recognition of various molecular features, for example the perception of multiple bonds and atom hybridization.

By default, the program works with a delocalized, rather than Kekulé, representation of aromaticity. ALTER automatically converts an imported Kekulé structure to an aromatic representation. The algorithm that performs this makes use of ring perception.

The automatic identification and quantification of cycles in

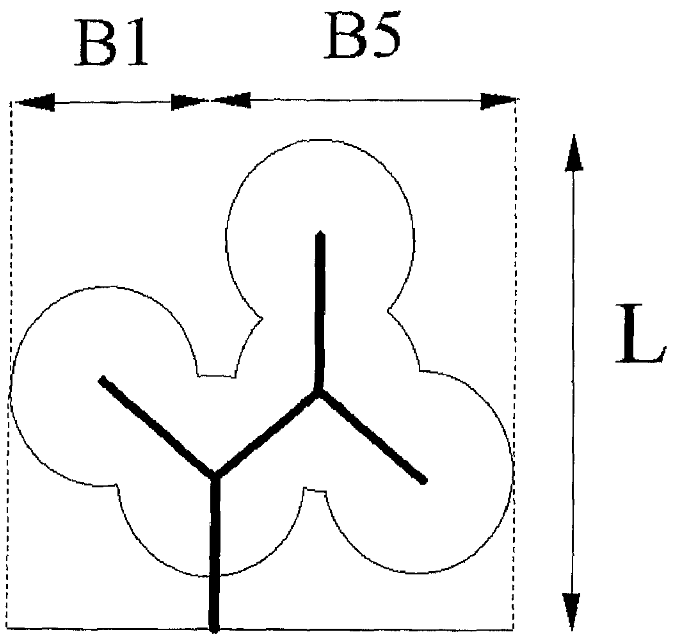


Figure 3. Diagrammatic representation of the three nonredundant Sterimol parameters: B1, B5, and L.

connectivity graphs has attracted a considerable literature. The theoretical basis of ring perception, and many of the algorithms used to perform this function in practice, have been reviewed in detail by Downs et al.[15] ALTER finds a smallest set of smallest rings using the method of Paton,[16] which has proved to be both fast and reliable. A depth-first path is traced through the atom adjacency matrix using a push-down stack of unused edges. When an edge is found to link two vertices already in the path this edge is flagged as ring closing. When the edge list is exhausted then the trace is terminated. The set of vertices forming the cycle associated with each ring closing edge is found by backtracking through the adjacency matrix using the path trace ordering to determine the shortest path, other than their common edge, between ring closing vertices.

ALTER is also able to effect the conversion from an aromatic to a Kekulé form. In algorithmic terms, this is a somewhat more challenging task, which is addressed within ALTER using a relatively simple method based on graph trimming. This problem has been discussed by Kearsley,[17] who presents an elegant algorithm addressing this problem.

DATA REORGANIZATION

ALTER supports a variety of functions for manipulating molecular data. These include different options for renumbering atoms and residues, for changing chain identifiers, and for manually or automatically renaming atoms and residues. ALTER also affords control over the amount and ordering of output data, selecting atoms by atom or residue ranges or by atom name or atom type. ALTER also allows the user to specify, increment, scale, clamp, or randomize temperature factors, occupancies, charges, and other atomic properties. Other functions allow for various sorts of coordinate randomization. These different functions can be combined to perform extensive reorganization of data with considerable ease and celerity. ALTER can also operate on the multiple fragments in a molecular data set. For example, ALTER can delete all but the largest fragment in a data set, fusing the first two fragments, or disjoint subgraphs, in the molecular ensemble by joining at tagged atoms, which are deleted and replaced by a bond.

COORDINATE TRANSFORMATIONS

ALTER is able to perform a number of different types of coordinate transformation. These operate on the entire molecular data set and change atomic positions in three-dimensional space. They include both isometric transformations (translation, including automatic centering at the origin, and rotation) and nonisometric transformations (isotropic and anisotropic scaling, various crystallographic transformations, and arbitrary deformation using a nonorthogonal transformation matrix). These different operations can also be combined.

ALTER supports various different parameterizations of rotation: an arbitrary Eulerian rotation matrix can be specified by giving the three successive rotation axes as a sequence of arguments together with corresponding rotation angles; a polar angle representation; a quaternion-based representation, the four elements of the quaternion being given as arguments; the Cayley–Klein representation; the three Cayley–Klein parameters should be given as arguments; and rotation by a given angle about a given vector.

ALTER can apply several crystallographic transformations to coordinate sets, given user-supplied unit cell parameters. A coordinate set can be reorthogonalized between different Ncodes, converted to and from fractional coordinates, or converted to and from axis-skewed real-space coordinates.

BUILDING

Although ALTER makes use of dedicated, external small molecule structure building programs, such as CONCORD or CORINA, to effect 2D to 3D conversion, at least for small organic molecules, it does support other forms of structure building. One of the simplest is the capacity to generate, adding or replacing as appropriate, the position of hydrogen atoms, where the coordinates of the heavy atoms are known. ALTER can also try to generate a set of reasonable 2D, rather than 3D, coordinates of a small molecule, given its molecular topology graph.

ALTER also supports 1D–3D conversion with various options for the simple building and homology modeling of proteins. The 3D structure of protein can be created from an amino acid sequence read from a sequence file. Optionally, a file of angles can be used to set the torsion angles of the protein chain as it is built—rather than use a default fully extended conformation. ALTER can also automatically build the full complement of amino acid sides onto a protein backbone. It can be used to swap between different types of side chains, effectively mutating, on an individual basis, one amino acid residue for another. Extending this feature, ALTER is capable of a simple type of homology modeling: the mutation of the sequence of a protein structure to that read from a sequence file that contains a pairwise or multiple sequence alignment. This process can be further automated to generate models for each sequence in a multiple alignment, by working through the alignment building a sequence at a time onto the structure of the initial coordinate set.

DATA ABSTRACTION

In addition to its other features, ALTER is able to calculate a number of different properties or characteristics of a molecule. Output information generated by the various options described below can be written to the screen or to a file, or, alternatively, it can be written to specially formatted output data files via WRITE or CONVERT commands.

The simplest is to list the structure of the current molecule: atoms, residues, their properties, connectivity, and topology. Likewise, the program can provide such related data as listing discontinuous fragments; the number, sizes, and composition of all rings; and, for proteins, secondary structure ranges. As well as the ability to calculate bond lengths, bond angles, and torsion angles, ALTER can calculate the geometric limits of the coordinate set in three dimensions.

ALTER is also able to generate a wide range of molecular characteristics that one might loosely label descriptors. These are quantities that might be useful in, say, formulating a QSAR or in studies of molecular similarity. For example, ALTER can count the number of hydrogen bond donors and acceptors present in a molecule. It is also able to calculate Sterimol parameters for molecular fragments. The three least redundant Sterimol[18] parameters—B1, B5, and L—can be calculated by the program given a molecular structure and tagged atom (see Figure 3).

In addition, ALTER is able to calculate numerous so-called topological indices. These quantities, also known as structural invariants, are derived from properties of the graph corresponding to a molecular structure. The list of generated indices[19-23] includes the Weiner index, Altenburg index, Balaban index, centric index, Zagreb M_1 and M_2 index, Gordon–Scantlebury index, the Platt number, Randic indices, PetitJohn R^2 and D^2 indices, the Harary number, the Schultz index, mean distance index, Balaban RMSD index, the graph distance index, information Weiner index, Burden molecular identification numbers, Kier and Hall shape indices κ_0, κ_1, κ_2, and κ_3; heteroatomic κ_0, κ_1, κ_2, and κ_3; and Kier's flexibility index φ. The program is also able to generate generalized Kier and Hall path indices of lengths 1 to 10, and Kier and Hall cluster and cluster/path indices, and count thereof. ALTER also generates several molecular symmetry indices, as well as calculating the electropological state of all heavy atoms in the molecule.

ALTER is also able to generate a set of related quantities that, for want of a better term, we shall call constitutional descriptors. They are related to, but are simpler and more easily interpreted than, topological indices. Values calculated include: number of rings, number and percentage of rotatable bonds, and the percentage of different element types within the organic subset (C, O, N, P, S, F, Cl, Br, I). The program also generates count-based quantities derived from partitioning atoms according to different rules and from the classification of bond types.

The number, range, and variety of topological indices and constitutional descriptors generated by ALTER compares favorably with those produced by MOLCONN-X[24] and POLLY[25]: the only other generally available programs of this type.

SMALL MOLECULE DISPLAY

ALTER supports two kinds of graphical output: the interactive display of a molecular structure and different sorts of output graphics file. Interactive visualization is supported only by the Silicon Graphics version of ALTER.

The VIEW command activates the interactive graphical options. ALTER has different visualization modes for small and macromolecules. The command VIEW issued without arguments enters an all-atom display function, intended to allow inspection and error checking of small molecule structures. The atomic display style shows a detail color-coded wire, or line-bond, all-atom representation of the molecule with the option to display different atom labels. Making use of the speed and power of the GL graphics library ALTER also supports, within VIEW, space-filling "licorice" bond, and ball-and-stick rendered molecular representations. Within all of these various different graphical modes, ALTER affords complete control over the specification of atom, residue, or region coloring. Color Plate 1 gives an example of a molecule variously displayed using VIEW. Two commands related to VIEW can be used to view sequentially the contents of a multiple entry database file, such as an MACCS SD file, one entry after the other. SCAN works through a database of 2D structures, and SCAN3D through a file of 3D structures, displaying them and allowing limited interactive manipulation.

PROTEIN STRUCTURE DISPLAY

ALTER has two interactive modes for visualizing simplified representations of protein structures. The simpler of the two options, the VIEW TRACE command, allows the interactive display of a protein C_α trace. Various different representations, including a smoothed backbone representation and smoothed helicoidal axis representation, with the option of line-based display of ligands, are supported.

A number of programs have been developed that produce schematic illustrations of proteins, directly from atomic coordinates, in an essentially device-independent manner. The most popular of these programs have been RIBBON,[26] and its derivatives,[27] and MOLSCRIPT.[28] An alternative approach, exemplified by the programs RIBBONS,[29] SETOR,[30] and FOLD,[31] makes use of hardware rendering to produce high-quality images. More recently, several programs have appeared that combine the intrinsic esthetic appeal of ribbon drawings with the photorealism of ray-tracing techniques. Examples of such software include MOLMOL[32] and the widely used Raster3D.[33] ALTER supports the schematic display of protein structures both through interactive hardware rendering and noninteractive ray tracing: the first display allowing one, in effect, to preview the second (see below).

The command VIEW RIBBON allows the interactive display of various rendered representations of a protein structure, principally variants on the Richardson-style protein ribbon,[34] although rendered versions of representations similar to those of the VIEW TRACE option are also supported. In this schematic mode, ALTER seeks to represent the overall structure of a protein in a simplified, but esthetically appealing, way. In common with all so-called ribbon drawings β strands are depicted as arrows, α helices as spiral ribbons or cylinders, and other structures as a coiling rope or line. Definitions of secondary structure are either imported with the structure or defined separately by the user. At the same time, the display style used for the helix can be defined as either a Richardson-type smoothly curving ribbon or as a cylindrical arrow. Likewise, the display style used for the strand can be defined as a Richardson-type smoothly curving arrow, as a kinked arrow style, or as a flat arrow style. Ligands can also be displayed within VIEW RIBBON, using either a space-filling, a "licorice" bond, or a ball-and-stick rendered all-atom representation.

GRAPHICAL OUTPUT

Graphical output files take one of two forms. The first consists of directly printable PostScript files of different molecular representations. These include stereo pictures of protein C_α traces or PLUTO-style all-atom pictures. The second consists of output files that are themselves input control files for some other graphics program. Primary among these is POVRAY, a public-domain ray-tracing program of considerable scope and power. Dion has adumbrated the use of POVRAY to visualize the structure of small molecules[35]; independently of that report, we have developed ALTER as an interface to POVRAY. Molecules can be displayed in a space-filling, a "licorice" bond, or a ball-and-stick representation. Likewise, proteins can be displayed using a Richardson-type representation that mirrors that offered by the VIEW RIBBON option. POVRAY, like most similar ray-tracing packages, offers a wealth of powerful tools for generating images of outstanding realism and three-dimensionality. Beyond simple coloring POVRAY supports control of a rich variety of textures including the incorporation of images; it allows for object transparency, complex lighting models, and atmospheric effects; and the program facilitates easy animation of images. The generality and flexibility of systems such as POVRAY, supplemented by interfaces such as ALTER, make them good alternatives to dedicated, purpose-built systems such as Raster3D. Color Plate 2, and figures prepared elsewhere,[36] give examples of POVRAY-based molecular visualization created using ALTER. Figure 2 shows the relatively brief set of ALTER commands necessary to generate the input file used to create Color Plate 2c.

DISCUSSION

ALTER is able to function as a file converter, data generation engine, and as an interface to other programs, for example previewing output to POVRAY. Although useful in itself, the program is, like similar software, only a partial solution to the general dilemma of data interconversion. Beyond its immediate utility for some specific task, the importance of a particular program lies in the ideas—the methods and algorithms that are embodied by it. Technological change, whether these are advances in hardware, changes in operating system design, or the creeping obsolescence of computer languages, will ultimately render any piece of software redundant. Our experience with the development of ALTER suggests that an approach based on a program combining a common data structure with a library of functions for input, output, and data manipulation is a sound one, but a more flexible, more general implementation of this design strategy is required to provide a sufficiently complete solution to this problem.

ALTER constitutes a powerful and flexible data-handling tool. In trying to develop a computer-based representation of molecular structure, and a means to archive this, there is typically a conflict between generality and the need for conciseness of expression. For many applications, the need for optimized speed and minimal storage requirements outweighs the desirability of an exhaustive representation. Thus the propagation of different file formats geared to different applications is likely to continue. In the absence of a universal standard for representing and storing molecular structure data, there will probably always be a need for programs such as ALTER.

ALTER can be obtained from the author, or by anonymous ftp from the following: guitar.rockeller.edu/pub/jpo/alter.tar.gz (login: FTP).

ACKNOWLEDGMENTS

I should like to thank the following, who have, in various ways and at various times, shown a helpful interest in ALTER: Dr. Sandy Geddes, Dr. Nigel Gensmantel, Dr. Colin Groom, Richard Ince, Dr. Johanna Jansen, Carol Manners, Dr. Will Somers, Dave Marriott, Dr. Lillian Nandi, and Dr. Nick Tomkinson—whom I should like to thank, in particular, for help with POVRAY.

REFERENCES

1 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Databank: A Computer archive for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542

2 Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R., and Watson, D.G. The Cambridge crystal data centre: Computer-based search, retrieval, analysis, and display of information. *Acta. Crystallogr.* 1979, **B35,** 2331–2339

3 Bebak, H., Buse, C., Donner, W.T., Hoever, P., Jacob, H., Klaus, H., Pesch, J., Roemelt, J., Schilling, P., Woost, B., and Zirz, C. The Standard Molecular Data format (SMD format) as an integration tool in computer chemistry. *J. Chem. Inf. Comput. Sci.* 1989, **29,** 1–5

4 Hall, S.R., Allen, F.H., and Brown, I.D. The crystallographic information file (CIF): A new standard archive file for crystallography. *Acta Crystallogr.* 1991, **A47,** 655–673

5 Davies, A.N. Internet Chemical MIME. *Spectroscopy 1996,* **8,** 42–46

6 Lesk, A.M. A toolkit for computational molecular biology. 3. MICROYFON: A (fairly) general program for input of protein co-ordinate files. *J. Appl. Crystallogr.* 1987, **20,** 488–490

7 Brocklehurst, S.M. and Perham, R.N. Prediction of the 3-dimensional structures of the biotinylated domain of the yeast pyruvate-carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system—a new automatic method for the prediction of protein tertiary structure. *Protein Sci.* 1993, **2,** 626–639

8 Humphrey, W., Dalke, A., and Schulten, A. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 1996, **14,** 33–38

9 Gilbert, D.G. *RDSEQ.* Department of Biology, University of Indiana, Bloomington, Indiana, 1989

10 Walters, P. and Stahl, M. *BABEL 1.1.* University of Arizona, Tucson, Arizona, 1992

11 Anderson, E., Veith, G.D., and Weininger, D. SMILES: A chemical language and information system. *J. Chem. Inf. Comput. Sci.* 1988, **28,** 31–36

12 Rusinko, A., Skell, J.M., Balducci, R., McGarity, C.M., and Pearlman, R. *CONCORD: A Program for the Rapid Generation of High Quality Approximate 3-Dimensional Molecular Structures.* TRIPOS Associates, St. Louis, Missouri, 1988

13 Gasteiger, J., Rudolph, C., and Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methods* 1990, **3,** 537–547

14 Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A., and Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* 1992, **32,** 244–255

15 Downs, G.M., Gillet, V.J., Holliday, J.D., and Lynch, M.F. Review of ring perception algorithms for chemical graphs. *J. Chem. Inf. Comput. Sci.* 1989, **29,** 172–187

16 Paton, K. An algorithm for finding a fundamental set of cycles of a graph. *Commun. Am. Chem. Soc.* 1969, **12,** 514–518

17 Kearsley, S.K. A quick robust method for assigning a Kekulé structure. *Comput. Chem.* 1993, **17,** 1–10

18 Verloop, A., Hoogenstraaten, W., and Tipker, J. Development and application of new steric substituent parameters in drug design. *Drug Design* 1976, **7,** 165–207

19 Basak, S.C., Niemi, G.J., and Veith, G.D. Predicting properties of molecules using graph invariants. *J. Math. Chem.* 1991, **7,** 243–272

20 Kier, L.B. and Hall, L.H. *Molecular Connectivity in Chemistry and Drug Research.* Academic Press, New York, 1976

21 Kier, L.B. and Hall, L.H. *Molecular Connectivity in Structure Activity Analysis.* Research Studies Press, Letchworth, U.K., 1986

22 Katritzky, A.R. and Gordeeva, E.V. Traditional topological indices vs electronic, geometrical and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* 1993, **33,** 835–857

23 Bonchev, D. *Information Theoretic Indices for Characterisation of Chemical Structures.* Research Studies Press, Letchworth, U.K., 1983

24 Hall, L.H. *MOLCONN-X.* Hall Associates Consulting, Quincy, Massachusetts, 1987

25 Basak, S.C., Harriss, D.K., and Magnuson, V.R. *POLLY 2.3.* Copyright of the University of Minnesota, 1988

26 Priestle, J.P. RIBBON—a stereo cartoon drawing program for protein structures. *J. Appl. Crystallogr.* 1988, **20,** 572–576

27 Flower, D.R. Improved ribbon drawing programs. *J. Mol. Graphics* 1991, **9,** 257–258

28 Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 1991, **24,** 946–950

29 Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* 1987, **5,** 103–106

30 Evans, S.V. SETOR: Hardware lighted three-dimensional solid modelling representation of macromolecules. *J. Mol. Graphics* 1993, **11,** 134–138

31 Flower, D.R. FOLD: Integrated analysis and display of protein secondary structure. *J. Mol. Graphics* 1995, **13,** 377–384

32 Koradi, R., Billeter, M., and Wuthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* 1996, **14,** 51–55

33 Merritt, E.A. and Murphy, M.E.P. Raster3D version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr.* 1994, **D50,** 869–873

34 Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 1981, **34,** 167–339

35 Dion, A.S. Personal computer-based visualization of molecular models by available ray-tracing software. *J. Mol. Graphics* 1994, **12,** 41–44

36 Flower, D.R. The lipocalin protein family: Structure and function. *Biochem. J.* 1996, **318,** 1–14