Chapter 1

CSearch: CHARMm Free Congen

1.1 Introduction

CSearch is a version of the conformational searching program, CONGEN, written by Dr. Robert E. Bruccoleri of Bristol Myers Squibb[**?**]. CONGEN was written as an extension to the CHARMm molecular mechanics package[**?**] developed by Dr. Bruccoleri and other workers at Harvard University under the direction of Professor Martin Karplus.

When Oxford Molecular, Ltd. wished to produce a commercial version (AbM) of the antibody modelling package[**?**, **?**] developed by Dr. Andrew Martin in Professor Tony Rees's group in Oxford, they needed to include CONGEN as part of the package. However, for licensing reasons (CHARMm is available as a commercial package from Polygen, Inc.) they were unable to redistribute the original CONGEN package which included the complete implementation of CHARMm. Thus CSearch, a CHARMm free version of CONGEN was developed by Dr. Martin for OML.

The development of CSearch was wholly backed by Dr. Bruccoleri who wishes to make his work freely available to as many people as possible. CSearch is essentially a CHARMm-free version of CONGEN, i.e. it takes all the source code from CONGEN which is not part of CHARMm and supplements this with additional code to replace the required functionality from CHARMm.

The conformational search part of CONGEN relies on CHARMm in the following ways:

- Data structures such as molecular topology information and coordinates,
- Energy parameters to create the CHARMm potential,
- Energy calculation and minimisation,
- Command parsing.

In CSearch, the CHARMm free version of CONGEN, these issues have been addressed as follows:

Data structures The internal data structures are maintained in the form used by CHARMm as they were in CONGEN. However, the code which fills in these data structures has been completely replaced.

Energy parameters The parameters which represent the CHARMm force field have been published in the literature and are thus freely useable. The original CHARMm code which reads these data has been completely replaced.

Energy calculation CONGEN evaluates the energy of each sidechain conformation it generates and checks for Van der Waals clashes using routines separate from those in the rest of CHARMm; these routines have thus been retained in CSearch. CONGEN also allows you to apply CHARMm energy minimisation to each conformation as it is generated; this option is not available in CSearch. The AbM package allows this option to be implemented in Eureka.

Command parsing The CHARMm command parser has been completely replaced by a less sophisticated, but much simpler keyword parser.

Normally, when running the AbM package, you will never need to worry about creating input files for CSearch. However, you may find it desirable to explore the conformational space in particular regions of the molecule without making use of the rest of the AbM package. This guide explains the format of the CSearch input file allowing you to perform these types of experiments.

1.2 Understanding CONGEN and CSearch

The conformational search procedure is essentially a simple one; a fragment of protein structure is deleted and rebuilt by spinning around torsion angles using a grid of, for example, 30°. As each new atom position is generated, this is used as a starting point for spinning the next torsion angle.

As additional torsion angles are added to the search, the number of conformations generated and the amount of computer time required rises exponentially. CONGEN implements a number of strategies to reduce the amount of time required, though searches of more than 5 amino acids will still take many hours on a machine such as a Silicon Graphics Iris workstation. Typically a 5 amino acid fragment will take between 30 and 60 minutes. The exact time depends on the nature of the amino acids present and on the conformational flexibility available to the peptide.

The conformational search is organised as a tree search. The first amino acid whose conformation is searched forms the base of the tree. Each conformation generated for this amino acid represents a branch eminating from the tree's main trunk. In turn, each of these branches (a conformation for the first residue being constructed) splits into a number of sub-branches as conformations for a second amino acid are generated using the first amino acid as a starting point. This sub-branching of the tree continues as each additional amino acid is added to the search. At the end of each branching system a leaf represents a fully built conformation for the peptide fragment being constructed.

Clearly, it is advantageous if non-productive branches (i.e. those which will never generate a leaf because they are going to be of very high energy, or where it is impossible to span the gap in the protein structure because the residues are pointing away from eachother) are pruned as early as possible. If branches which are not going to produce any leaves are identified close to the root, the program will spend less time generating futile combinations of residue conformations. CONGEN implements a number of optimisation strategies in order to introduce this pruning.

Firstly, the ϕ and ψ torsion angles are combined into a single degree of freedom. Ramachandran style energy maps containing energies for each ϕ/ψ combination are stored and only those combinations with energies below a specified point on the map are allowed. Three such maps are stored; one for glycine, one for proline and one for alanine which is used as being representative of all the other amino acids.

Secondly, as each position is constructed, a distance check is made to ensure that it is possible to span the remaining gap using the number of bonds which remain unconstructed. For the purpose of this checking, a linear peptide is assumed with each bond angle stretched by 5°. If it is not possible to span the remaining gap

using such a stretched linear peptide, this ϕ/ψ angle combination is rejected and the next combination is examined. Once a ϕ/ψ combination has been rejected, it becomes unnecessary to build any of the other residues using this conformation of the current residue as a base on which to work.

Thirdly, conformations with van der Waals energies higher than a specified cutoff are rejected.

Fourthly, the last three residues are constructed using an analytical procedure developed by Gō and Scheraga[**?**, **?**]. This maps the 6 torsional degrees of freedom available for 3 residues onto the 3 translational and 3 rotational degrees of freedom required to move from one bond vector to another.

Sidechains are built as the last level in the branching procedure. CONGEN implements 5 algorithms for generating sidechains[**?**], but in virtually all cases the 'ITER' algorithm is used. This proceeds as follows:

- Nested iterations over the sidechain torsion angles are performed until the first energetically acceptable set of sidechain conformations is generated.
- All possible conformations for the first amino acid are then generated and the conformation with the lowest energy is selected and its energy is recorded.
- Using this conformation for the first amino acid, the generation of all possible conformations is repeated for each amino acid in turn until all sidechains have been searched. In each case the lowest energy conformation for each sidechain is retained.
- The process then iterates from the first sidechain and continues until the energy converges or an iteration limit is reached.

This method for constructing the sidechains produces only one energetically favourable sidechain conformation per backbone conformation. While the procedure can be biased by the starting conformation if the number of iterations required is large, this has been shown to be the most effective method and the loss of accuracy is small compared with the savings in computer time which result from not generating *every* possible set of sidechain conformations[**?**, **?**].

1.3 Running CSearch

CSearch is run simply by typing the command:

```
csearch <input file> <output file>
```

where `<input file>` is a control file as described below and `<output file>` is a listing file which gives details of the CSearch run. In the AbM system on local at UCL, CSearch is store in the `/usr/local/bin` directory.

You should be aware that CSearch runs can take many hours to complete. The amount of time required depends on the number of residues being constructed, the nature of those residues and the conformational flexibility available to the loop. Therefore, you should probably run CSearch as a background task at lower than normal priority or as a batch job on those machines which support this type of operation.

1.4 The CSearch Input File

The CSearch input file consists of keywords followed, where necessary, by one or more parameters. The keyword command parser which reads this file will accept

upper, lower or mixed case keywords. You may abbreviate keywords to the shortest unambiguous string. CSearch will tell you if the abbreviation you use is ambiguous or a keyword is meaningless. It will also tell you if the parameters you have entered with a keyword which requires them are invalid.

When specifying filenames as parameters, you should be careful to use the correct case if you are using an operating system such as UNIX which is case-sensitive.

When specifying residues by number, note that the numbering in the PDB file used for input is also used within CSearch. However, residues will be assigned to a chain (L, H, A, B, C, D, E, F) based on information from the input sequence .PIR file. Thus, if you have a PDB file containing 2 chains each of 100 amino acids, numbered 1–200, CSearch will expect you to refer to these as residues L 1–100 in the first chain and as H 101–200 in the second chain. Note that you will also be expected to provide NTER and CTER residues in the PDB file, but it is conventional in using CSearch to give these numbers outside the range of the rest of the structure. See Section 1.5.1 for an example.

The keywords which you may use are divided into two groups:

Global commands These set general parameters. Effectively they replace the commands which were shared with CHARMm in CONGEN. In general, you will only need to use global commands to specify filenames and to define the regions which are going to be rebuilt or which are to be ignored in the constructions.

Conformational search commands These are used to specify the exact residues to be constructed by conformational search, the way in which the construction is to be performed and energy parameters relating to the construction.

1.4.1 Global Commands

Each of the global commands will be described in the following sections. Some of these commands are *essential*; your CSearch run will fail if they are not specified. These are shown with the word **Essential** as the first word in the description.

N.B. The order of the global commands is important. If you do not enter the commands in the order given, the CSearch program is likely to exit abnormally. You should use the order:

- RESTOP
- PARAMS
- SIDETOP
- GLYMAP
- ALAMAP
- PROMAP
- PROCONS
- PGP
- SEQUENCE
- COORDS

When specifying data files, unless you have them in your current directory, you will have to specify a path by which to find them. The exact nature of this path will depend on the operating system in use and the individual installation of the software. For the local machine, local, at UCL, the data files are currently in the directory `../data`.

In the descriptions which follow, the pathname will be ommited from the descriptions of files, you should insert the appropriate pathname for your system.

RESTOP *filename*

Specifies the residue topology file. This is used to store information about the topology of standard residues and the atom types they contain. You should select the standard file: `topology`

PARAMS *filename*

Specifies the CHARMm parameters file. This file contains parameters for the interactions between atoms pairs (bonded and non-bonded), for angles and both proper and improper torsion angles. You should select the standard file: `params`

SIDETOP *filename*

Specifies the sidechain topology file. This contains topology information for each of the sidechains. You should select the standard file: `sidetop`

GLYMAP *filename*

Specifies the glycine Ramachandran energy map. You may select one of four energy maps depending on the grid size you wish to use for the backbone search. You should thus select one of:

- `emapgly30`
- `emapgly15`
- `emapgly10`
- `emapgly5`

ALAMAP *filename*

Specifies the alanine Ramachandran energy map. This is used as representative of all amino acids aother than glycine and proline. You may select one of four energy maps depending on the grid size you wish to use for the backbone search. You should thus select one of:

- `emapala30`
- `emapala15`
- `emapala10`
- `emapala5`

PROMAP ***filename***

Specifies the proline Ramachandran energy map. You may select one of four energy maps depending on the grid size you wish to use for the backbone search. You should thus select one of:

- `emappro30`
- `emappro15`
- `emappro10`
- `emappro5`

PROCONS ***filename***

Specifies the proline constructor file. This contains data used in the generation of the proline ring. You should select the standard file: `procns`

PGP ***filename***

Specifies a proton generation parameter file. This file is used to specify where hydrogens need to be added to a structure and the geometry involved in those additions. You should select the standard file: `pgp`

SEQUENCE ***filename***

Specifies a sequence file containing sequence information for your protein. This information is used to build the topology information for the protein. The file is in standard PIR format. i.e. two title lines followed by the sequence in one-letter code with each chain being terminated by an asterisk (`*`). New lines and spaces in the sequence are ignored. The `RESTOP` and `PARAMS` keywords must be specified before the `SEQUENCE` keyword. The chains are given labels by the software as they are read into the program. Because the system is designed primarilly for modelling antibodies, the chains are labelled L, H, A, B, C, D, E, F. With antibodies, you should thus specify the light chain then the heavy chain and any other chains which are to be considered. CSearch only supports up to 8 chains, but this is more than adequate for the majority of purposes.

COORDS ***filename***

Specifies the input PDB coordinate file for the protein. Regions of this structure will be deleted by the `CLEAR` command before being rebuilt by conformational search. The `PGP`, `SEQUENCE` and `RESTOP` keywords must be specified before the `COORDS` keyword.

The format of this file is standard PDB format with some minor additional requirements:

1. Each chain must start with an `NTER` residue containing atom types `HT1` and `HT2`. These atoms should have 'dummy' coordinates of 9999.000; the atom and residue numbers are unimportant, but the residue number should be different from any other residue in the same chain.
2. The nitrogen atom of the first true residue of each chain should be renamed `NT` rather than `N`.
3. Each chain should end with a `CTER` residue containing a single `OT2` atom with dummy coordinates. The atom and residue number is unimportant, but the residue number should be different from any other residue in the same chain.

4. The last true residue of each chain should have its carbonyl carbon atom renamed from `O` to `OT1`.

5. The C-δ atoms in isoleucine residues must be named `CD` rather than the conventional `CD1`.

See Section 1.5.1 for an example.

CLEAR ***chain startres endres***

Clears the coordinates for a range of residues. This is necessary before rebuilding residues using the CGEN command. One may also choose to clear other regions of the protein structure such that these regions do not influence the building of the current region. For example, it is usual practice to delete all the 6 CDRs from the combining site while any one loop is being constructed.

GLYEMAX ***emax***

Specifies the maximum allowed energy from the glycine Ramachandran map. CSearch determines the lowest energy in the Ramachandran map and adds this value to the lowest energy. Only combinations of ϕ/ψ angles with energies lower than this combined value are considered. The default is 100.0, but a more typical value is 2.00

PROEMAX ***emax***

Specifies the maximum allowed energy from the proline Ramachandran map. CSearch determines the lowest energy in the Ramachandran map and adds this value to the lowest energy. Only combinations of ϕ/ψ angles with energies lower than this combined value are considered. The default is 100.0, but a more typical value is 2.00

ALAEMAX ***emax***

Specifies the maximum allowed energy from the alanine Ramachandran map. CSearch determines the lowest energy in the Ramachandran map and adds this value to the lowest energy. Only combinations of ϕ/ψ angles with energies lower than this combined value are considered. The default is 100.0, but a more typical value is 2.00

ERINGPRO ***emax***

Specifies the maximum allowed ring energy when constructing prolines. The default is the largest single precision real number (1.7×10^{38}), but a more typical value is 50.00

RESTART ***restart***

Specifies a restart string which should be enclosed in double inverted commas. If a CSearch run has been interrupted (for example, it has run out of disk space or the computer has been switched off during the run) and a `STATUS` degree of freedom has been specified, the file written by the `STATUS` option may be examined. The values specified in this file are given as a parameter to the restart command (enclosed in double inverted commas) to cause the conformational search to continue from the point at which the last status file was written. It is then necessary to merge the conformations files which are generated from the separate parts of the run.

DEBUG ***system level***

Switches on debugging information. *system* may be one of the keywords: `PRIN`, `ALLST`, `ALLOC`, `ALLHP`, `CLSC`, or `CGEN`. Some of these options are now redundant and others are only of interest to persons modifying CSearch. The only options of any real interest are `CLSC`, and `CGEN`. The *level* is a value bewteen 0 and 10; higher values generate more copious volumes of output. Specifying a value of 10 for `CGEN` will write to the disk continuously and will fill it as quickly as possible!

EPS ***dielectric***

Specifies the dielectric constant to be used in the evaluation of the electrostatic component of sidechain energies. (Default: 50.00)

CUTNB ***distance***

Specifies the non-bonded cutoff distance in Å. Non-bonded interactions between atoms separated by more than this amount will not be considered. (Default: 5.00)

CUTHB ***distance***

Specifies the hydrogen bond cutoff distance in Å. Hydrogen bond interactions between atoms separated by more than this amount will not be considered. (Default: 4.5)

CUTHA ***angle***

Specifies the hydrogen bond cutoff angle in degrees. Hydrogen bond interactions between atoms forming an angle of greater than this value will not be considered. (Default: 90.0)

ECHO

When a PDB file is read using the `COORDS` keyword, certain modifications such as filling in coordinates for the `NTER` and `CTER` residues of each chain and the addition of hydrogens are normally made. This option causes the modified coordinates to be written to the file `reference.pdb`.

NOHADD

When a PDB file is read using the `COORDS` keyword, certain modifications such as filling in coordinates for the `NTER` and `CTER` residues of each chain and the addition of hydrogens are normally made. This option stops these modifications from being made, so the input coordinate file must already contain these data.

1.4.2 Conformational Search Commands

The conformational search process is represented by a set of 'degrees of freedom' (See Section 1.2 for more information). The term is used very loosely and does not have its conventional meaning of spatial freedom. Rather it is used to describe levels in the conformational search tree. Some levels may not be involved with the generation of new conformations, but may, instead relate to the writing of coordinates, reading of previously generated coordinates, or writing of status information.

Each degree of freedom is specified by one of the keywords described below falling between the `CGEN` and `END` commands which bracket the conformational search procedure.

Note that the degrees of freedom become nodes in the search tree in the exact order in which they are specified in these commands. Clearly you cannot construct sidechains on a stretch of residues before the backbone has been built and you cannot, therefore, place a `SIDECHAIN` command before the `FORWARD`, `REVERSE` and `CHAIN` commands used to build the backbone. (Though you could, of course, first build sidechains on some other region of the structure which already has its backbone defined.) Similarly if you specify the `WRITE` command before you have specified any construction commands, you will get no coordinates in your conformation file.

The commands are described in the order in which they would be used in a typical construction.

None of the commands below is actually essential, but if a basic set of commands is not given, either no conformation generation will occur, or the coordinates will not be written to a file. Thus some of commands below are marked as **Essential**, even though their ommision will not cause CSearch to exit abnormally.

Note that you will need to specify a `CLEAR` command before using the `LOOPS` or backbone construction commands.

CGEN

This command is used to indicate the start of the specification of conformational search degrees of freedom.

LOOPS *filename*

Specifies that a previously written set of conformations should be read. This may either be a conformation file created by a previous run of CSearch or a set of database conformations extracted using the AbM package.

FORWARD *maxevdw chain startres lastres closeres closeatm*

Specifies a range of residues to be constructed in the 'forward' direction (i.e. from the N-terminus towards the C-terminus). The region must first have been cleared using the `CLEAR` keyword. The *maxevdw* parameter specifies the maximum allowed backbone van der Waals interaction energy (typically 20.0). The *chain*, *startres* and *lastres* parameters specify the range of residues which are to be built, while the *closeres* and *closeatm* parameters are used to determine which atom should be used for distance checks when building the residues. This 'closing' atom is effectively where the chain is 'aiming' for and should normally be the Cα atom of the last residue in the constructed region.

REVERSE *maxevdw chain startres lastres closeres closeatm*

Specifies a range of residues to be constructed in the 'reverse' direction (i.e. from the C-terminus towards the N-terminus). The region must first have been cleared using the `CLEAR` keyword. The *maxevdw* parameter specifies the maximum allowed backbone van der Waals interaction energy (typically 20.0). The *chain*, *startres* and *lastres* parameters specify the range of residues which are to be built, while the *closeres* and *closeatm* parameters are used to determine which atom should be used for distance checks when building the residues. This 'closing' atom is effectively where the chain is 'aiming' for and should normally be the N atom of the second residue in the constructed region.

CHAIN ***maxevdw chain startres***

Specifies the start of a three-residue section which is to be built by the chain closure algorithm of Gō and Scheraga[**?**]. The region must first have been cleared using the `CLEAR` keyword. `FORWARD` and `REVERSE` commands must be given such that only these three residues are left to be built. The *maxevdw* parameter specifies the maximum allowed backbone van der Waals interaction energy (typically 100.0) while the *chain* and *startres* parameters specify the beginning of the three residue section to be built.

1.4.3 STATUS *filename*

Specifies that a status file should be written containing information on the progress through the conformational search tree. The file will only contain information about degrees of freedom specified up to this point. The information in this file is used in conjunction with the `RESTART` command to recover data after CSearch has exited abnormally (for example after a system crash or when the user's disk space has filled up). Generally this command is given before the `SIDECHAIN` command.

SIDECHAIN ***maxevdw grid chain startres lastres method***

Specifies the range of sidechains to be built. The backbone for this region must exist whether already present in the structure or built by conformational search using the `FORWARD`, `REVERSE` and `CHAIN` keywords. It is *not* necessary to clear the region first; if the region *is* cleared, the backbone must be reconstructed before sidechains may be built. The *maxevdw* parameter specifies the maximum allowed van der Waals interaction energy (typically 20.0), the *grid* parameter may either be the word `MIN` or a value in degrees which is the step size for conformational search around sidechain torsion angles. Selecting `MIN` causes only the lowest energy torsion angles to be examined (generally this is the best option). The *chain*, *startres* and *lastres* parameters specify the range of sidechains to be built, while the *method* parameter specifies the sidechain construction method to be used. This must be one of `ITER`, `ALL`, `FIRST`, `INDEPENDENT`, or `COMBINATION`. The `ITER` method, which is generally the method of choice, is described in Section 1.2. For descriptions of the other methods, you are referred to Bruccoleri and Karplus[**?**] or Martin[**?**].

WRITE ***filename***

Specifies that atom positions generated at the current depth in the conformational search tree should be written to the specified file.

END

Specifies the end of the conformational search degrees of freedom.

1.5 Examples

1.5.1 PDB File format

The following example shows the form required for a PDB file to be read by CSearch. The example shows two chains of an antibody, but in order to keep the example short, shows only the first and last two true residues from each chain.

Note the addition of `NTER` and `CTER` residues and the modifications to the first and last true residues in each chain. See Section 1.4.1 for details.

```
ATOM   9998  HT1 NTER    0    9999.0009999.0009999.000
ATOM   9998  HT2 NTER    0    9999.0009999.0009999.000
ATOM      1  NT  ASP     1      14.916  12.012   0.335  0.00  0.00
ATOM      2  CA  ASP     1      13.738  11.236  -0.105  0.00  0.00
ATOM      3  CB  ASP     1      12.455  11.967   0.314  0.00  0.00
ATOM      4  CG  ASP     1      12.186  11.734   1.803  0.00  0.00
ATOM      5  OD1 ASP     1      11.458  10.795   2.183  0.00  0.00
ATOM      6  OD2 ASP     1      12.746  12.526   2.598  0.00  0.00
ATOM      7  C   ASP     1      13.826  10.889  -1.584  0.00  0.00
ATOM      8  O   ASP     1      14.486  11.580  -2.389  0.00  0.00
ATOM      9  N   ILE     2      13.143   9.816  -1.924  0.00  0.00
ATOM     10  CA  ILE     2      13.114   9.316  -3.365  0.00  0.00
ATOM     11  CB  ILE     2      13.797   7.940  -3.263  0.00  0.00
ATOM     12  CG1 ILE     2      13.855   7.088  -4.541  0.00  0.00
ATOM     13  CG2 ILE     2      13.206   7.107  -2.077  0.00  0.00
ATOM     14  CD  ILE     2      15.021   6.021  -4.377  0.00  0.00
ATOM     15  C   ILE     2      11.668   9.438  -3.782  0.00  0.00
ATOM     16  O   ILE     2      10.785   8.897  -3.086  0.00  0.00
...
ATOM    887  N   ARG   114     -18.208   1.673 -20.442  0.00  0.00
ATOM    888  CA  ARG   114     -19.206   0.637 -20.838  0.00  0.00
ATOM    889  CB  ARG   114     -19.679  -0.148 -19.648  0.00  0.00
ATOM    890  CG  ARG   114     -19.976   0.639 -18.386  0.00  0.00
ATOM    891  CD  ARG   114     -21.157   0.102 -17.659  0.00  0.00
ATOM    892  NE  ARG   114     -22.291  -0.056 -18.553  0.00  0.00
ATOM    893  CZ  ARG   114     -23.115  -1.109 -18.540  0.00  0.00
ATOM    894  NH1 ARG   114     -22.987  -2.107 -17.676  0.00  0.00
ATOM    895  NH2 ARG   114     -24.100  -1.168 -19.452  0.00  0.00
ATOM    896  C   ARG   114     -20.315   1.331 -21.603  0.00  0.00
ATOM    897  O   ARG   114     -20.270   2.570 -21.788  0.00  0.00
ATOM    898  N   ALA   115     -21.278   0.541 -22.055  0.00  0.00
ATOM    899  CA  ALA   115     -22.411   1.141 -22.809  0.00  0.00
ATOM    900  CB  ALA   115     -23.042   0.181 -23.757  0.00  0.00
ATOM    901  C   ALA   115     -23.370   1.709 -21.753  0.00  0.00
ATOM    902  OT1 ALA   115     -23.582   1.117 -20.689  0.00  0.00
ATOM   6666  OT2 CTER  666    9999.0009999.0009999.000
ATOM   7777  HT1 NTER  777    9999.0009999.0009999.000
ATOM   7778  HT2 NTER  777    9999.0009999.0009999.000
ATOM    903  NT  GLN   116      -2.989 -19.567   8.673  0.00  0.00
ATOM    904  CA  GLN   116      -3.213 -18.168   8.472  0.00  0.00
ATOM    905  CB  GLN   116      -4.406 -17.712   9.381  0.00  0.00
ATOM    906  CG  GLN   116      -5.450 -16.887   8.651  0.00  0.00
ATOM    907  CD  GLN   116      -4.848 -15.846   7.740  0.00  0.00
ATOM    908  OE1 GLN   116      -4.317 -14.828   8.175  0.00  0.00
ATOM    909  NE2 GLN   116      -4.949 -16.107   6.439  0.00  0.00
ATOM    910  C   GLN   116      -1.861 -17.592   8.923  0.00  0.00
ATOM    911  O   GLN   116      -1.125 -18.277   9.659  0.00  0.00
ATOM    912  N   VAL   117      -1.603 -16.346   8.537  0.00  0.00
ATOM    913  CA  VAL   117      -0.425 -15.599   8.851  0.00  0.00
ATOM    914  CB  VAL   117       0.391 -15.576   7.547  0.00  0.00
ATOM    915  CG1 VAL   117       1.135 -14.301   7.225  0.00  0.00
ATOM    916  CG2 VAL   117       1.386 -16.674   7.758  0.00  0.00
ATOM    917  C   VAL   117      -1.191 -14.336   9.180  0.00  0.00
ATOM    918  O   VAL   117      -1.510 -13.561   8.280  0.00  0.00
```

```
...
ATOM   1805  N   SER   231      -8.439  12.161  14.381  0.00  0.00
ATOM   1806  CA  SER   231      -7.945  13.106  15.323  0.00  0.00
ATOM   1807  CB  SER   231      -6.497  13.495  15.127  0.00  0.00
ATOM   1808  OG  SER   231      -5.994  14.291  16.184  0.00  0.00
ATOM   1809  C   SER   231      -8.878  14.307  15.082  0.00  0.00
ATOM   1810  O   SER   231      -8.932  14.973  14.032  0.00  0.00
ATOM   1811  N   SER   232      -9.714  14.508  16.091  0.00  0.00
ATOM   1812  CA  SER   232     -10.663  15.579  16.104  0.00  0.00
ATOM   1813  CB  SER   232     -11.826  15.111  15.208  0.00  0.00
ATOM   1814  OG  SER   232     -12.998  15.935  15.141  0.00  0.00
ATOM   1815  C   SER   232     -11.129  16.014  17.502  0.00  0.00
ATOM   1816  OT1 SER   232     -11.245  15.324  18.507  0.00  0.00
ATOM   8888  OT2 CTER  888    9999.0009999.0009999.000
```

1.5.2 Rebuilding 5 Residues

This example illustrates the technique for rebuilding a section of 5 amino acids numbered 24–28 in the light chain. The first residue is built in the forward direction, the last in the reverse direction, with the three in the middle being built by the chain-closure method.

```
! Data files
RESTOP   /usr/people/AbM/ABM/DAT/topology
PARAMS   /usr/people/AbM/ABM/DAT/params
SIDETOP  /usr/people/AbM/ABM/DAT/sidetop
ALAMAP   /usr/people/AbM/ABM/DAT/emapala30
GLYMAP   /usr/people/AbM/ABM/DAT/emapgly30
PROMAP   /usr/people/AbM/ABM/DAT/emappro30
PROCONS  /usr/people/AbM/ABM/DAT/procns
PGP      /usr/people/AbM/ABM/DAT/pgp
!
SEQUENCE example1.pir   ! Sequence for this run
COORDS   example1.pdb   ! Input structure file for this run
! Parameters for this run
GLYEMAX  2.00
ALAEMAX  2.00
PROEMAX  2.00
ERINGPRO 50.00
EPS      50.0
CUTNB    5.0
CUTHB    4.5
CUTHA    90.0
!
ECHO                    ! Write coordinate file after input
CLEAR L   24   28       ! Clear the range for reconstruction
! Start the actual conformational search specification
CGEN
   FORWARD    20.0 L 24 24 28 CA
   REVERSE    20.0 L 28 28 25 N
   CHAIN     100.0 L 25
   SIDECHAIN  20.0 MIN L 24 28 ITER
   WRITE      L24_28.cg
END
```

1.5.3 Rebuilding Onto Previous Conformations

This example illustrates the method used in AbM to read a set of database conformations for the base of a loop and then to reconstruct the mid-section of the loop by conformational search. This is done in the same way as the 'real space renormalisation' (RSR) technique described by Bruccoleri[**?**, **?**].

In RSR, one builds the outside part of a long loop (maybe the first 2 residues and the last 2 residues) in one run of the program. A specifed number of the lowest energy conformations (maybe 10) are then read back into a second run of the program and used as a base on which to build the next residues.

CSearch does not evaluate the energy of the loops, so no energy is written into the conformation file. It is thus not possible to read back only the lowest energy conformations. Should you wish to attempt RSR, you would need an external program capable of reading the conformations file generated in the first run and writing a new conformations file containing only the low energy conformations.

When used with database loops as is the case in AbM, one always wishes to scan *all* the loop conformations.

The following example rebuilds the middle 5 residues (27–31) of a loop spanning residues 24–33 in the light chain. The example assumes that a file exists containing database conformations for the base regions of the loop (24–26 and 32–33).

```
! Data files
RESTOP   /usr/people/AbM/ABM/DAT/topology
PARAMS   /usr/people/AbM/ABM/DAT/params
SIDETOP  /usr/people/AbM/ABM/DAT/sidetop
ALAMAP   /usr/people/AbM/ABM/DAT/emapala30
GLYMAP   /usr/people/AbM/ABM/DAT/emapgly30
PROMAP   /usr/people/AbM/ABM/DAT/emappro30
PROCONS  /usr/people/AbM/ABM/DAT/procns
PGP      /usr/people/AbM/ABM/DAT/pgp
!
SEQUENCE example2.pir   ! Sequence for this run
COORDS   example2.pdb   ! Input structure file for this run
! Parameters for this run
GLYEMAX  2.00
ALAEMAX  2.00
PROEMAX  2.00
ERINGPRO 50.00
EPS      50.0
CUTNB    5.0
CUTHB    4.5
CUTHA    90.0
!
ECHO                    ! Write coordinate file after input
CLEAR L   24   33       ! Clear the range for reconstruction
! Start the actual conformational search specification
CGEN
   LOOPS      example2.cgin           ! Database loops
   FORWARD    20.0 L 27 27 31 CA
   REVERSE    20.0 L 31 31 28 N
   CHAIN     100.0 L 28
   SIDECHAIN  20.0 MIN L 24 33 ITER
   WRITE      L24_33.cg
END
```

1.5.4 Rebuilding Sidechains

This example illustrates the rebuilding of a set of sidechains without any backbone reconstruction. Sidechains for residues 33–39 in the heavy chain are rebuilt and the positions of residues 50–60 are not considered during the construction.

Note that we do not clear the coordinates for a region in which only the sidechains are being built.

```
! Data files
RESTOP   /usr/people/AbM/ABM/DAT/topology
PARAMS   /usr/people/AbM/ABM/DAT/params
SIDETOP  /usr/people/AbM/ABM/DAT/sidetop
ALAMAP   /usr/people/AbM/ABM/DAT/emapala30
GLYMAP   /usr/people/AbM/ABM/DAT/emapgly30
PROMAP   /usr/people/AbM/ABM/DAT/emappro30
PROCONS  /usr/people/AbM/ABM/DAT/procns
PGP      /usr/people/AbM/ABM/DAT/pgp
!
SEQUENCE example3.pir   ! Sequence for this run
COORDS   example3.pdb   ! Input structure file for this run
! Parameters for this run
GLYEMAX  2.00
ALAEMAX  2.00
PROEMAX  2.00
ERINGPRO 50.00
EPS      50.0
CUTNB    5.0
CUTHB    4.5
CUTHA    90.0
!
ECHO                    ! Write coordinate file after input
CLEAR H   50   60       ! Clear the region to be ignored
! Start the actual conformational search specification
CGEN
   SIDECHAIN  20.0 MIN H 33 39 ITER
   WRITE H33_39SC.cg
END
```