ECalc V1.4
A protein energy calculation program.

(c) 1994, Dr. Andrew C. R. Martin,
Biomolecular Structure and Modelling Unit,
Department of Biochemistry and Molecular Biology,
University College London,
Gower Street,
London WC1E 6BT

22nd May, 1995

Abstract

ECalc is a program designed to calculate the energy of a protein. It may also be used to calculate the energies of a number of conformations of a region within a protein. The calculations may be restricted to the region of interest (though non-bonded contacts with other parts of the protein will also be considered). In addition, other parts of the protein may be specified which should be excluded from consideration in the non-bonded contact calculations. Parts of the potential may be switched on or off and scaled with respect to one another. This allows, for example, the simple calculation of just the van der Waals repulsion or just the bond energy.

The program may be run using a control file (as described below), or using an X-windows interface (**EC**) on machines equipped with the Tcl/Tk package. This interface builds a control file and spawns the normal **ECalc** program to perform the calculations.

1 Installation

1.1 Unpacking

ECalc is supplied as a gzipped tar file. To unpack, go to the directory in which you want the **ECalc** directory to be installed and issue the command:

```
zcat ecalc.tar.gz | tar -xvf -
```

This assumes that `ecalc.tar.gz` is in your current directory (if not, you must supply the full path) and that `zcat` is the GNU/gzip version. If not, you may have to include the full path to the GNU `zcat` program.

1.2 Compiling

To compile **ECalc**, you must already have installed the **BiopLib** library. Assuming this has been done, change to the `src` subdirectory of the new `ecalc` directory and edit the supplied `makefile`. Modify `INCFLAGS` and `LIBFLAGS` to reflect the directory in which you have installed **BiopLib**. Modify `CC` and `LINK` for your C compiler. You only need to alter `PROTO` if you wish to regenerate prototype files. This will only be necessary if you have modified any of the routines.

Having modified the `makefile` as required, type `make` to compile the program. Move the resulting `ecalc` executable to somewhere in your path (e.g. `/usr/local/bin`).

Finally, type `make clean` to remove the object files.

1.3 Environment variables

One environment variable should be set to specify the directory containing the residue topology and parameter data files. Therefore, if you are using a csh-like shell, you should add the following to your `.cshrc` file:

```
setenv ECALCDATA '/usr/local/ecalc/data'
```

where `/usr/local/ecalc/data` is replaced by the name of the directory in which your **ECalc** datafiles are stored. If you are using an sh-like shell, add the following to your `.profile` file:

```
export ECALCDATA='/usr/local/ecalc/data'
```

1.4 Customising for your local setup

You must now customise the file `local.tcl` for your local system. Change to the data directory using the command:

```
cd $ECALCDATA
```

Now edit the file `local.tcl`. Find the line which starts `set ecalc` and change the path to reflect the location of you **ECalc** executable. For example:

```
set ecalc "/usr/local/bin/ecalc"
```

Find the line which starts `set auto_path` and modify the directory specification to reflect your main **ECalc** directory which contains `ecalc.tcl` and the other Tcl files.

```
set auto_path "/usr/local/ecalc $auto_path"
```

You may also modify the colours to be used for the buttons and the maximum number of disulphides, zones and ignores if required.

Now move to the main **ECalc** directory. Depending on the version of Tcl/Tk you have installed, you will require either the `tclIndex.old` or `tclIndex.new` file. Copy the appropriate version to `tclIndex`. If you have the wrong version, the ecalc interface will exit with a message like: *format error in library index*.

Finally, make a symbolic link from somewhere in your path to the main **ECalc** Tcl file. For example:

```
ln -s /usr/local/ecalc/ecalc.tcl /usr/local/bin/ec
```

2 Utilities

Two utilities are provided to help convert XPLOR (CHARMM) parameter and topology files to **ECalc** format. These programs are written in Perl and require that you have the Perl interpreter installed on your system.

Unfortunately, while the programs will produce valid data files for ECalc, the conversion is not complete; some hand editing will also be necessary. This is because

1. **ECalc** requires all atom parameter data to be provided in the parameter file while XPLOR has the atom masses and radii in the topology file,
2. The XPLOR topology file does not contain atom exclusion data.

Both conversion utilities are run in the same way:

```
xpar.perl params.xplor   >params.ecalc
xtop.perl topology.xplor >topology.ecalc
```

You should substitute appropriate filenames for the `.xplor` and `.ecalc` files sepcified here.

In addition the atom ordering in the residue topology file is critical if the program is to be used for conformation screening. The ordering must match that used in CSearch. i.e. N, H, CA, C, O, sidechain.

3 Preparing input data

Generally, **ECalc** will be run to follow on from a **CSearch** (CHARMM-free CONGEN) run, in which case all the necessary files will have been prepared already. The basic input files are a PDB file with hydrogens and CHARMM-style NTER and CTER residues[1] and, optionally, a conformations file in **CSearch** format.

[1] All future references to a PDB file will imply a PDB file with hydrogens, NTER and CTER residues added.

A suitable PDB file may be prepared using the following Unix command line:

```
pdbhadd -c xxxx.pdb | pdbcter -c | renumpdb | pdborder -c > xxxx.pdh
```

where `xxxx.pdb` is a normal PDB file and `xxxx.pdh` has had the hydrogens, NTER and CTER residues added. The `pdbhadd -c` command adds hydrogens and an NTER residue, the `pdbcter -c` command adds C-terminal oxygens and a CTER residue while `renumpdb` renumbers the chains sequentially.

4 Running the programs

4.1 EC

EC is the X-windows graphical interface for **ECalc**. It is run by typing `ec` at the command line prompt.

4.2 ECalc

The command syntax for **ECalc** is:

```
ecalc [-p (xxxx.pdh | -)] [(control.dat | -)]
```

where `xxxx.pdh` is a PDB file (see Section 3) and `control.dat` is a control file (see Section 5). In either case, the filename may be replaced by a `-`, in which case, the appropriate input will be expected on standard input.

If **ECalc** is run with a PDB file, but no control file, the full-potential energy of the whole structure will be calculated and displayed. Normally, **ECalc** will be run with a control file and the PDB filename will be specified within the control file.

5 The control file

5.1 PDBFILE *filename*

Specifies the PDB file to be used for energy calculation. If not specified in the file, the PDB file may be specified on the command line using the `-p` option.

5.2 CONFFILE *filename*

Specifies an optional conformation file. If specified, the energy of each conformation in the context of the PDB file will be calculated.

5.3 OUTFILE *filename*

Specifies an optional output file. If not specified, output goes to the standrad output stream.

5.4 POTENTIAL

If the **POTENTIAL** keyword is not specified, the standard full potential will be calculated. If the keyword is given, it is followed by one or more of the commands given in Table 1 *on separate lines*. After all potential options have been specified, the **END** keyword must be given.

BONDS [*scale*]	Calculate bond energy
ANGLES [*scale*]	Calculate angle energy
TORSIONS [*scale*]	Calculate torsion angle energy
IMPROPERS [*scale*]	Calculate improper torsion angle energy
HBONDS [*scale*]	Calculate hydrogen bond energy
VDWA [*scale*]	Calculate van der Waals attraction
VDWR [*scale*]	Calculate van der Waals repulsion
ELECTROSTATICS [*scale*]	Calculate electrostatic energy
RESIDUE [*scale*]	Calculate an empirical residue energy

Table 1: Keywords specifying parts of the potential to be included. In each case, the scale parameter is optional and, if not specified, will default to 1.0

5.5 DISPLAY

The **DISPLAY** command controls which components of the energy are displayed in the output. By default, only the total energy is displayed. The command is followed by keywords from Table 1, each of which must be given on a separate line and without the *scale* parameter. As with the **POTENTIAL** command, the options are ended with the **END** command.

5.6 CACHE n

At the end of the run, the conformation numbers and energies of the lowest energy conformations are displayed. The **CACHE** command is used to control how many conformations should be displayed. The default is 1 conformation. A value of 0 is also legal.

5.7 GRIDCUT n

This is a very important optimisation parameter. In order to speed the calculation of non-bonded interactions, a 'contact grid' is constructed at the beginning of the run using the coordinates in the parent structure. The grid is not recalculated at any time. This means that non-bonded atom contacts are only considered between atoms in the grid, thus speeding the calculations for each conformation. However, the atoms in each conformation are in different positions, so the grid must be larger than the non-bonded cutoff distance by an amount at least equal to the greatest movement of any one atom.

The utility `spancg` reports the greatest movement from the reference position for any atom in a CG file. The value reported by this program should be added to the non-bonded cutoff distance and used for the **GRIDCUT** parameter

See also the REGRID command (Section 5.23).

5.8 NONBONDCUT n

Cutoff to be used when calculating van der Waals and electrostatic energies. Any atom pairs separated by more than this distance will not be considered in calculating the energy. Defaults to 8.0Å.

5.9 CONSTDIELECTRIC

Use a constant dielectric when calculating the electrostatic energy. The value of the dielectric constant is specified using the **ETA** keyword. i.e.

$$E_{el} = C_{el}\frac{q_1 q_2}{d\eta}$$

5.10 DISTDIELECTRIC

Use a distance dependent dielectric when calculating the electrostatic energy. i.e.

$$E_{el} = C_{el}\frac{q_1 q_2}{d^2}$$

5.11 ETA n

Specifies the dielectric constant to be used when a constant dielectric is selected. Default = 50.0

5.12 CUTOFFHB n

Specifies the distance above which hydrogen bond energy should not be calculated. Atom pairs separated by more than this distance will be ignored. Default: 5.0Å.

5.13 CUTONHB n

Specifies a distance at which smoothing of the hydrogen bond energy calculations should begin. This prevents discontinuities as a result of the hydrogen bond distance cutoff. **CUTONHB** must be $\leq$ **CUTOFFHB**. Default: 4.0Å.

5.14 CUTOFFANG n

Specifies an angle below which hydrogen bond energy should not be calculated. If the angle subtended at the hydrogen by the acceptor and the hydrogen's antecedent is less than this value, the hydrogen bond will be ignored. Default: 90°.

5.15 CUTONANG n

Specifies an angle at which smoothing of the hydrogen bond energy calculations should begin. This prevents discontinuities as a result of the hydrogen bond angle cutoff. **CUTONANG** must be $\geq$ **CUTOFFANG**. Default: 90°.

5.16 SHOWPARAMS

Causes parameters read from the parameter file to be displayed.

5.17 SHOWRTOP

Causes the residue topology read from the residue topology file to be displayed.

5.18 SHOWTIMINGS

Causes timings to be displayed after calculation of the close contacts grid and each conformation.

5.19 PARAMFILE *filename*

Specifies the parameter file. Defaults to `params.dat`.

5.20 RTOPFILE *filename*

Specifies the residue topology file. Defaults to `topology.dat`.

5.21 DEBUG

Switches on certain debugging options. Currently only the routine name display with each error message is switched on. This gives a simulated trace-back of where the error occurred.

5.22 DISULPHIDES (OFF | *firstres lastres*)

By default, disulphides are found by a distance search; cysteine Sγ atoms separated by less than 2.236Å are considered to be bonded. If the **DISULPHIDES** keyword is followed by the word **OFF**, no search for disulphides will be performed. Alternatively, the **DISULPHIDE** command may be followed by residue pair specifications for disulphides to be specified manually. This also causes the automatic search for disulphides to be switched off. Residues are specified using the chain name (if present) followed immediately by the residue number (e.g. `L24`). Insertions in the numbering are not allowed; numbering is consecutive starting with the NTER residue being residue number 1.

5.23 REGRID n

By default, a close contacts grid is only calculated once, before any energies are calculated. If **REGRID** is specified with a parameter of 1, the grid will be recalculated before each energy is evaluated. Other values may also be used, but this must be done with caution; for example, you must know that, for example, every 10 conformations fall into a cluster in space. See the GRIDCUT command (Section 5.7) for further details on the close contacts grid.

5.24 ZONE *firstres lastres*

Specifies a zone over which calculations should be performed. More than one **ZONE** command may be specified. Bonded topology (bonds, angles, torsions and impropers) which does not include at least one atom in a zone is discarded. Non-bonded contacts (van der Waals, electrostatics and hydrogen bonds) are considered between atoms in the specified zones and atoms outside the zones.

Residues are specified using the chain name (if present) followed immediately by the residue number (e.g. `L24`). Insertions in the numbering are not allowed; numbering is consecutive starting with the NTER residue being residue number 1.

When applying zones to regions included in a conformation file, you should make the zones 1 residue wider on either side than the zone covered by the conformations. This is because of the way in which splicing between conformation and framework is performed by CSearch.

5.25 IGNORE *firstres lastres* [SIDECHAIN]

Specifies a zone to be ignored when calculating non-bond contacts. More thans one **IGNORE** command may be specified. Any non-bonded interactions (van der Waals, electrostatics and hydrogen bonds) which involve atoms in the specified range are ignored.

If the optional **SIDECHAIN** qualifier is specified, only sidechain atoms (from the γ atom out) are ignored.

Residues are specified using the chain name (if present) followed immediately by the residue number (e.g. `L24`). Insertions in the numbering are not allowed; numbering is consecutive starting with the NTER residue being residue number 1.

5.26 RUN

This command is provided only for interactive use and causes the actual calculations to begin. And end-of-file also has the same effect.

5.27 RELAX

Performs a relaxation of the worst van der Waals contacts followed by the SHAKE algorithm to regularise bond lengths.

5.28 TOL [SHAKE|VDW] *value*

Specifies the tolerence for the SHAKE algorithm (default: 0.001) or the allowed fraction VDW clash (default: 0.5). Higher values for VDW allow more atoms to clash without taking any action. The value is the fraction of the optimum separation squared by which the atoms clash.

6 Future ideas

Modify the grid recalculation calls to store the conformation used at grid calculation time then, for each conformation, check the atom positions to see if any atom has moved outside the current grid. If so recalculate the grid and store the new conformation.