The software consists of several programs for the hybridization and folding of nucleic acid sequences. Included are programs to hybridize two sequences, fold a single sequence and create plots of base pair probabilities, as well as several utilities.

Syntax

Each program takes command line arguments in the traditional Unix style: 0 or more options followed by 1 or more file names or prefixes. An option consists of a dash, a single letter, an optional space and a value; or else of two dashes, a word, a space or `=` and a value. For example, if `tmin` is a long option with corresponding short option `t`, then `-t37`, `-t 37`, `--tmin=37` and `--tmin 37` all set `tmin` to 37.

Sequence file format

The programs read sequences from standard text files. In these files, whitespace characters (spaces, tabs and carriage returns) are ignored, and case is irrelevant. Any alphabet character that is not an A, C, G, T or U (T and U are considered equivalent) is considered to be an "unknown" character. (Numbers and symbols are ignored along with whitespace.)

Programs

The programs included are:

hybrid-min – to determine the minimum hybridization energy and corresponding structure of two sequences
hybrid-ss-min – to determine the minimum folding energy and corresponding structure of a sequence
ct-energy – to determine the free energy of a given structure

hybrid-min

hybrid-min determines the minimum energy hybridization of two nucleic acid sequences and outputs it as a `.ct` file. Two file names are required, which specify the files from which to read the first and second sequences. The `tmin` and `tmax` parameters specify the starting and ending temperature, in degrees Celsius. The `tinc` parameter specifies the temperature increment; the default is one degree. (Since temperatures are stored as double-precision floating-point values internally, the temperature parameters need not be integers.) The `NA` parameter can take two values: `NA=RNA` (the default) or `NA=DNA`, indicating which set of energy rules to use in hybridizing the sequences. The `suffix` option can be used to specify an alternate of set energy rules, in which case the `tmin`, `tinc` and `tmax` options are ignored. The `dHscale` parameter allows for scaling the enthalpies at $37\,^{\circ}C$. The `loopS` option can be used to give an extra entropy (expressed in $cal/mol/K$) to all bulge and interior loops. For example, setting `loopS` to -6.45 has the effect of destabilizing bulge/interior loops by about 2 kcal at $37\,^{\circ}C$. If the `allpairs` option is specified, hybrid-min will allow for all base pairs; otherwise any base pairs except the Watson-Crick and wobble pairs are automatically assigned a probability of 0. (However, since non-canonical base pairs typically have energies of $+\infty$, the `allpairs` parameter generally has little effect.) The `nodangle` option removes dangle energies from consideration, essentially setting each dangle energy to $+\infty$. The `maxloop` option can be used to restrict the maximum size of bulge and interior loops to be considered, in order to speed up calculation. (`maxloop` defaults to 30.) The `sodium` and `magnesium` options can be used to specify Sodium- and Magnesium-ion concentrations; the default is $[Na^{+}] = 1M$ and $[Mg^{++}] = 0M$. (Note that salt corrections have no effect on RNA sequences.) When salt corrections are used with DNA sequences, the `polymer` option causes the salt corrections for polymers to be used instead of those for oligomers. (In polymer mode, Magnesium-ion concentrations are ignored.) All salt corrections are ignored when the `suffix` option is used. Specific `i,j` base pairs can be prevented from forming using the `prohibit` option. If `j` is set to 0 or omitted, base `i` in the first sequence is prohibited from pairing with any base; likewise if `i` is 0 or not present, base `j` in the second sequence if prohibited from pairing with any base. hybrid-min also supports the `noisolate` option, which prohibits any isolated base

pair, i.e. any base pair not involved in a stack. Finally, the `energyOnly` option causes hybrid to compute only the free energy; i.e. to skip the computation of the optimal structure.

Currently only base pairs involving one base from the first sequence and one base from the second sequence are considered. Therefore, the hybridizations consist entirely of stacks, bulge/interior loops and dangling bases; there are no hairpin loops or multibranch loops.

Output

All of the output files begin with a `prefix`: the first file name (with a trailing `.seq` removed if applicable) followed by a hyphen and the second file name (again with a trailing `.seq` removed). For each temperature `temp`, hybrid-min writes a file `prefix.temp.plot`, containing for each possible (i, j) pair (where i indexes the first sequence and j the second) the probability $P(i, j)$ that i and j are paired. hybrid-min also writes `prefix.temp.ext`, which contains for each i the probability that base i is unpaired and the probability that bases i and $i + 1$ are both unpaired (and likewise for the second sequence). With hybrid-min, as opposed to the hybrid program, all probabilities are either 0 or 1. Finally, hybrid-min writes `prefix.temp.ct` which contains the minimum energy and corresponding structure; and `prefix.temp.asc`, which contains an ASCII representation of the structure.

hybrid-min also writes `prefix.dG`, which contains for each temperature the value $-RTlnZ$.

hybrid-ss-min

hybrid-ss-min determines the minimum energy folding of a nucleic acid sequence and outputs it as a `.ct` file. Because hybrid-min does not consider intrasequence base pairs, hybrid-ss-min is in fact much more complex than hybrid-min. hybrid-ss-min requires one file name, which specifies the file from which to read the sequence. The `tmin`, `tinc`, `tmax`, `NA`, `suffix`, `dGscale`, `dHscale`, `loopS`, `allpairs`, `nodangle`, `sodium`, `magnesium`, `polymer`, `filter`, `prohibit`, `noisolate` and `energyOnly` options function as in hybrid-min. The `maxloop` option functions as in hybrid-min, but also restricts the size of hairpin loops. hybrid-ss-min also supports the `simple` option, which simplifies the formula used to calculate the energy of a multibranch loop by setting `b=0` and `c=0`, so that the penalty is simply a constant.

Output

hybrid-ss-min writes `prefix.dG` and, for each temperature, `prefix.temp.plot`, `prefix.temp.ext` and `prefix.temp.ct` for each temperature `temp`, as does hybrid-min.

ct-energy

ct-energy computes the free energy of each structure in a `.ct` file and writes it to standard output. The `NA`, `temperature`, `suffix`, `sodium`, `magnesium` and `polymer` options determine which set of energy rules to use. The `verbose` option causes ct-energy to display the energy for each loop/stack, as well as the total energy; specifying `verbose` twice causes the components of each loop to be displayed as well.

Energy Rules

hybrid-min, and hybrid-ss-min read energy rules from several files. To determine the energy associated with one base pair stacked on another, they read free energies at $37\,^{\circ}C$ from `stack.DG` (or `stack.DGD` for DNA) and enthalpies from `stack.DH` (or `stack.DHD`). These files are generated automatically (by make) from human-readable versions with the same names but with lowercase suffixes.

The same suffixes are used with the prefixes `dangle`, for single bases dangling on base pairs; `sint2`, for 1x1 symmetric interior loops; `asint1x2`, for 1x2 (and 2x1) asymmetric interior loops; `sint4`, for 2x2 symmetric interior loops; `tstacki`, for terminal mismatch stacking energies for interior loops; `loop`, for hairpin, bulge and interior loops other than those already mentioned; and `miscloop`, for miscellaneous rules.

hybrid-ss-min also loads files with the prefixes `tstackh`, for terminal mismatch stacking energies for hairpin loops; `triloop`, for special hairpin triloops; and `tloop`, for special hairpin tetraloops. hybrid-ss and hybrid-ss-min also read the affine penalty for multi-branch loops as well as some miscellaneous parameters from `miscloop.*`.

Each of these files is loaded from the current directory, if it exists there, or else from the directory set in the `HYBRIDLIB` environment variable.