1   Name

MELTING - nearest-neighbor computation of nucleic acid hybridation

2   Synopsis

melting [*options*]

3   Description

MELTING computes, for a nucleic acid duplex, the enthalpy and the entropy of the helix-coil transition, and then its melting temperature. Three types of hybridisation are possible: DNA/DNA, DNA/RNA, and RNA/RNA. The program uses the method of nearest-neighbors. The set of thermodynamic parameters can be easely changed, for instance following an experimental breakthrough. Melting is a free program in both sense of the term. It comes with no cost and it is open-source. In addition it is coded in ISO C and can be compiled on any operating system. Some perl scripts are provided to show how melting can be used as a block to construct more ambitious programs.

If you use MELTING, please quote

Le Novère. MELTING, a free tool to compute the melting temperature of nucleic acid duplex. *Bioinformatics*, 17: 1226-1227.

4   Options

The options are treated sequentially. If there is a conflict between the value of two options, the latter normally erases the former.

-A*file.nn*

Informs the program to use *file.nn* as an alternative set of nearest-neighbor parameters, rather than the default for the specified hybridisation type (option **-H**). The standard distribution of melting provides some files ready-to-use: *all97a.nn* (Allawi et al 1997), *bre86a.nn* (Breslauer et al 1986), *san96a.nn* (SantaLucia et al 1996), *sug96a.nn* (Sugimoto et al 1996) (DNA/DNA), *fre86a.nn* (Freier et al 1986), *xia98a.nn* (Xia et al 1998) (RNA/RNA) and *sug95a.nn* (Sugimoto et al 1995) (DNA/RNA). The program will look for the file in a directory specified during the installation. However, if an environment variable NN_PATH is defined, melting will search in this one first. Be careful, the option **-A** changes the default parameter set defined by the option **-H.**

-C*complementary_sequence*

Enters the complementary sequence, from 3' to 5'. This option is mandatory if there are mismatches between the two strands. If it is not used, the program will compute it as the complement of the sequence entered with the option **-S**

1

-D*dnadnade.nn*

> Informs the program to use the file *dnadnade.nn* to compute the contribution of dangling ends to the thermodynamic of helix-coil transition. The dangling ends are not taken into account by the approximative mode.

-F*factor*

> This is the a correction factor used to modulate the effect of the nucleic acid concentration in the computation of the melting temperature. See section AL-GORITHM for details.

-h

> Displays a short help and quit with EXIT_SUCCESS.

-H*hybridisation_type*

> Specifies the hybridisation type. This will set the nearest-neighbor set to use if no alternative set is provided by the option **-A** (remember the options are read sequentially). Moreover this parameter determines the equation to use if the sequence length exceeds the limit of application of the nearest-neighbor approach (arbitrarily set up by the author). Possible values are *dnadna*, *dnarna* and *rnadna* (synonymous), and *rnarna*. For reasons of compatibility the values of the previous versions of melting *A,B,C,F,R,S,T,U,W* are still available although **strongly** deprecated. Use the option **-A** to require an alternative set of thermodynamic parameters. **Important:** If the duplex is a DNA/RNA heteroduplex, the sequence of the DNA strand has to be entered with the option **-S**

-I*input_file*

> Provides the name of an input file containing the parameters of the run. The input has to contain one parameter per line, formatted as in the command line. The order is not important, as well as blank lines. example:

```
-Hdnadna
-Asug96a.nn
-SAGCTCGACTC
-CTCGAGGTGAG
-N0.2
-P0.0001
-v
-Ksan96a
```

-K*salt_correction*

> Permits to chose another correction for the concentration in sodium. Currently, one can chose between *wet91a, san96a, san98a*. See section ALGORITHM

-L

> Prints the legal informations and quit with EXIT_SUCCESS.

2

-M*dnadnamm.nn*

Informs the program to use the file *dnadnamm.nn* to compute the contribution of mismatches to the thermodynamic of helix-coil transition. Note that not all the mismatched Crick's pairs have been investigated. Therefore it could be impossible to compute the Tm of a mismatched duplex. Moreover, those mismatches are not taken into account by the approximative mode.

-N*x.xxe-xx*

Sodium concentration (between 0 and 10 M). The effect of ions on thermodynamic stability of nucleic acid duplexes is complex, and the correcting functions are at best rough approximations. Moreover, they are generally reliable only for [Na^+] belonging to [0.1,1 M]. Note that the divalent ions, notably Mg^{2+} have a drastic effect, and therefore the result of MELTING is correct only in absence of such divalent ions. Some authors showed that a mix of 0.15 M NaCl with 10 mM MgCl2 could be equivalent to 1 M NaCl. An artificial increase of the sodium concentration prior to the computation could be a way to proceed.

-O*output_file*

The output is directed to this file instead of the standard output. The name of the file can be omitted. An automatic name is then generated, of the form meltingYYYYMMMDD_HHhMMm.out (of course, on POSIX compliant systems, you can emulate this with the redirection of stdout to a file constructed with the program date).

-P*x.xxe-xx*

Concentration of the nucleic acid strand in excess (between 0 and 0.1 M).

-p

Return the directory supposed to contain the sets of calorimetric parameters and quit with EXIT_SUCCESS. If the environment variable NN_PATH is set, it is returned. Otherwise, the value defined by default during the compilation is returned.

-q

Turn off the interactive correction of wrongly entered parameter. Useful for run through a server, or a batch script. Default is OFF (i.e. interactive on). The switch works in both sens. Therefore if **-q** has been set in an input file, another **-q** on the command line will switch the quiet mode OFF (same thing if two **-q** are set on the same command line).

-S*sequence*

Sequence of one strand of the nucleic acid duplex, entered 5' to 3'. **Important:** If it is a DNA/RNA heteroduplex, the sequence of the DNA strand has to be entered. Uridine and thymidine are considered as identical. The bases can be upper or lowercase.

-T*xxx*

Size threshold before approximative computation. The nearest-neighbour approach will be used only if the length of the sequence is inferior to this threshold.

-v

Control the verbose mode, issuing a lot more information about the current run (try it once to see if you can get something interesting). Default is OFF. The switch works in both sens. Therefore if **-v** has been set in an input file, another **-v** on the command line will switch the verbose mode OFF (same thing if two **-v** are set on the same command line).

-V

Displays the version number and quit with EXIT_SUCCESS

-x

Force the program to compute an approximative tm, based on G+C content. This option has to be used with caution. Note that such a calcul is increasingly incorrect when the length of the duplex decreases. Moreover, it does not take into account nucleic acid concentration, which is a strong mistake.

5 Algorithm

5.1 Thermodynamics of helix-coil transition of nucleic acid

The nearest-neighbor approach is based on the fact that the helix-coil transition works as a zipper. After an initial attachment, the hybridisation propagates laterally. Therefore, the process depends on the adjacent nucleotides on each strand (the Crick's pairs). Two duplexes with the same base pairs could have different stabilities, and on the contrary, two duplexes with different sequences but identical sets of Crick's pairs will have the same thermodynamics properties (see Sugimoto et al. 1994). This program first computes the hybridisation enthalpy and entropy from the elementary parameters of each Crick's pair.

$$
\begin{aligned}
\Delta H &= \delta h_{\text{initiation}} + \sum \delta h_{\text{Crick's pair}} \\
\Delta S &= \delta s_{\text{initiation}} + \sum \delta s_{\text{Crick's pair}}
\end{aligned}
$$

See Wetmur J.G. (1991) and SantaLucia (1998) for deep reviews on the nucleic acid hybridisation and on the different set of nearest-neighbor parameters.

5.2 Effect of mismatches and dangling ends

The mismatching pairs are also taken into account. However the thermodynamic parameters are still not available for every possible cases (notably when both positions are mismatched). In such a case, the program, unable to compute any relevant result, will quit with a warning. The two first and positions cannot be mismatched. in such a case, the result is unpredictable, and all cases are possible. for instance (see Allawi and SanLucia 1997), the duplex

```
A           T
 GTGAGCTCAT
 TACTCGAGTG
T           A
```
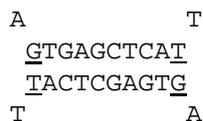
Figure 1: Comparison of experimental and computed Tm for various sets of nearest-neighbor parameters. $[\text{Na}^+] = 1$ M, [nucleic acid] $= 4 \cdot 10^{-4}$ M
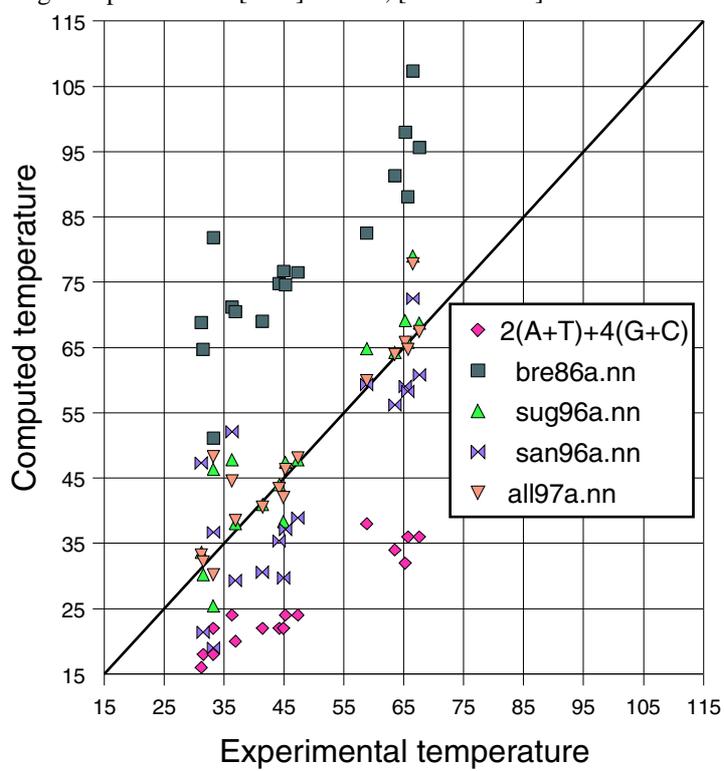
Figure 2: Comparison of experimental and computed Tm for various sets of nearest-neighbor parameters. $[Na^+] = 0.11$ M, [nucleic acid] $= 8 \cdot 10^{-6}$ M
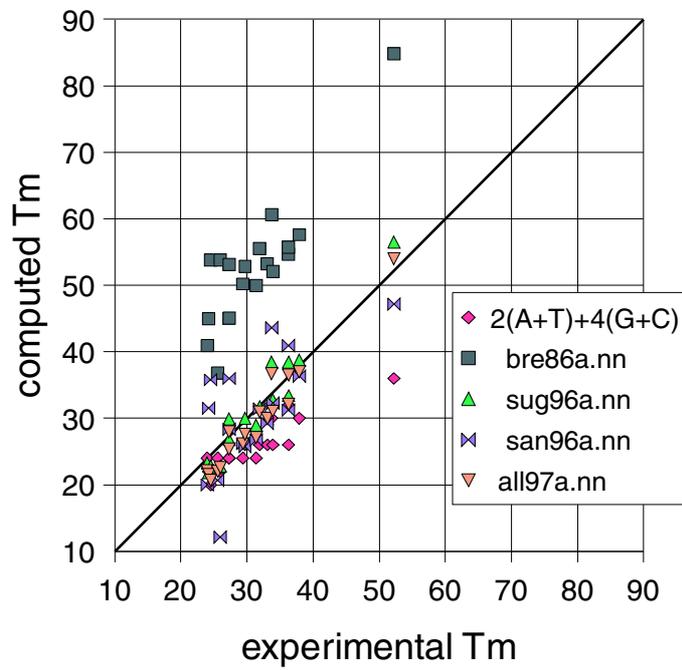
is more stable than

```
AGTGAGCTCATT
TTACTCGAGTGA
```

The dangling ends, that is the unmatched terminal nucleotides, can be taken into account.

5.3 Example

$$\Delta H \begin{pmatrix} \text{AGCGATGAA-} \\ \text{-CGCTGCTTT} \end{pmatrix} = \Delta H \begin{pmatrix} \text{AG} \\ \text{-C} \end{pmatrix} + \Delta H \begin{pmatrix} \text{A-} \\ \text{TT} \end{pmatrix}$$
$$+ \Delta H \begin{pmatrix} \text{G} \\ \text{C} \end{pmatrix}_{\text{init}} + \Delta H \begin{pmatrix} \text{A} \\ \text{T} \end{pmatrix}_{\text{init}}$$
$$+ \Delta H \begin{pmatrix} \text{GC} \\ \text{CG} \end{pmatrix} + \Delta H \begin{pmatrix} \text{CG} \\ \text{GC} \end{pmatrix} + 2x \Delta H \begin{pmatrix} \text{GA} \\ \text{CT} \end{pmatrix} + \Delta H \begin{pmatrix} \text{AA} \\ \text{TT} \end{pmatrix}$$
$$+ \Delta H \begin{pmatrix} \text{AT} \\ \text{TG} \end{pmatrix} + \Delta H \begin{pmatrix} \text{TG} \\ \text{GC} \end{pmatrix}$$

(The same computation is performed for ΔS)

5.4 The melting temperature

Then the melting temperature is computed by the following formula:

$$\text{Tm} \quad = \quad \frac{\Delta H}{\Delta S + R \ln(C_T/F)} \quad + \quad \mathcal{F}([\text{Na}^+]) - 273.15$$

Tm in K (for $[\text{Na}^+]$ = 1 M)   *correction* for the salt concentration and to get the temperature in degree Celsius. (In fact some corrections are directly included in the ΔS see that of SanLucia 1998)

5.5 Correction for the concentration of nucleic acid

F is 1 in the case of self-complementarity oligonucleotides. If the ODNs are not self-complementary, F is 4 if both strands are present in equivalent amount and F is 2 if one strand is in excess (for instance in PCR experiments). Actually in the latter case, the formula would have to use the difference of concentrations rather than the total concentration. But if the excess is sufficient, the total concentration can be assumed to be identical to the concentration of the strand in excess. That is, if one strand is in excess, the actual formula is effectively $(C_{\text{max}} - C_{\text{min}})/2$ but if $C_{\text{max}} \gg C_{\text{min}}$, $C_{\text{max}} - C_{\text{min}}$ is close to the total concentration C_T. If C_{max} is close to C_{min}, $(C_{\text{max}} - C_{\text{min}})/2$ is equivalent to $C_T/4$, which is the default correction.

Note however that MELTING makes the assumption of no self-assembly, *i.e.* the computation does not take any entropic term to correct for self-complementarity.

Figure 3: Comparison of experimental and computed Tm for various correction of salt concentration.
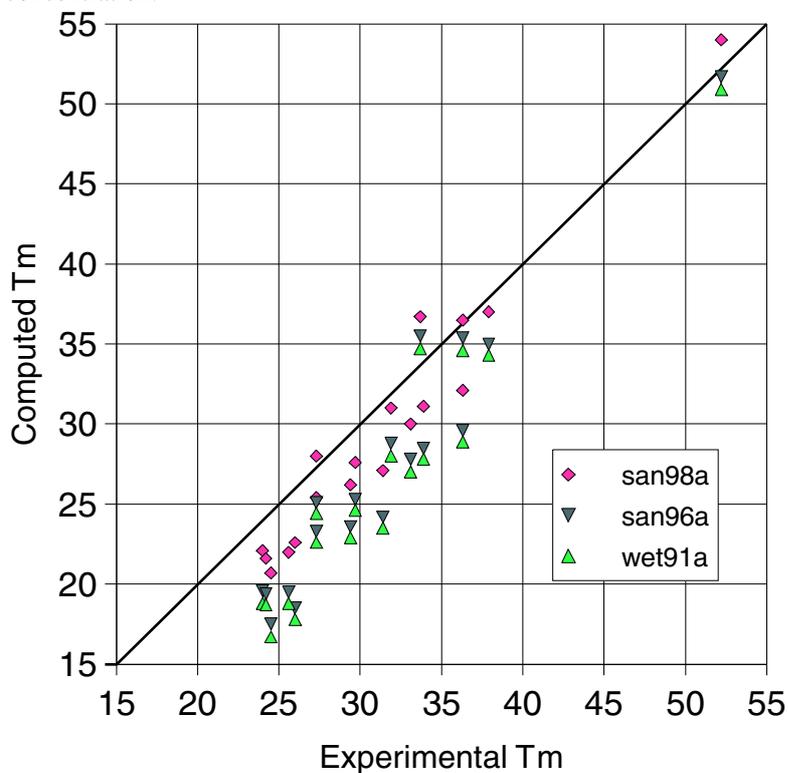
5.6 Correction for the concentration of salt

Currently, the only available correction is for sodium (i.e. monovalent cations). the correction can be chosen between *wet91a,* presented in Wetmur 1991 *i.e.*

$$16.6 \log \frac{[\text{Na}^+]}{1 + 0.7[\text{Na}^+]} + 3.85$$

san96a presented in SantaLucia et al. 1996 *i.e.*

$$12.5 \log[\text{Na}^+]$$

and *san98a* presented in SantaLucia 1998 *i.e.* a correction of the entropic term without modification of enthalpy

$$\Delta S = \Delta S_{[\text{Na}^+]=1 \text{ M}} + 0.368(N - 1) \ln[\text{Na}^+]$$

Where *N* is the length of the duplex.

5.7 Long sequences

It is important to realise that the nearest-neighbor approach has been established on small oligonucleotides. Therefore the use of MELTING in the non-approximative mode is really accurate only for relatively short sequences (Although if the sequences are two short, let's say < 6 bp, the influence of extremities becomes too important and the reliability decreases a lot). For long sequences an approximative mode has been designed. This mode is launched if the sequence length is higher than the value given by the option -T (the default threshold is 60 bp).

The melting temperature is computed by the following formulas:

ADN/ADN:

$$Tm = 81.5 + 16.6 \log \frac{[\text{Na}^+]}{1 + 0.7[\text{Na}^+]} + 0.41\% GC - \frac{500}{size}$$

ADN/ARN:

$$Tm = 67 + 16.6 \log \frac{[\text{Na}^+]}{1 + 0.7[\text{Na}^+]} + 0.8\% GC - \frac{500}{size}$$

ARN/ARN:

$$Tm = 78 + 16.6 \log \frac{[\text{Na}^+]}{1 + 0.7[\text{Na}^+]} + 0.7\% GC - \frac{500}{size}$$

The usage of this mode is nevertheless **strongly disencouraged.**

5.8 Miscellaneous comments

MELTING is currently accurate only when the hybridisation is performed at pH 7 ± 1. The computation is valid only for the hybridisations performed in aqueous medium. Therefore the use of denaturing agents such as formamide completely invalidates the results.

6 References

Allawi H.T., SantaLucia J. (1997). Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry* 36: 10581-10594

Bommarito S., Peyret N., SantaLucia J. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res* 28: 1929-1934

Breslauer K.J., Frank R., Blöcker H., Marky L.A. (1986). Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* 83: 3746-3750

Freier S.M., Kierzek R., Jaeger J.A., Sugimoto N., Caruthers M.H., Neilson T., Turner D.H. (1986). *Biochemistry* 83:9373-9377

SantaLucia J. Jr, Allawi H.T., Seneviratne P.A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35: 3555-3562

Sugimoto N., Katoh M., Nakano S., Ohmichi T., Sasaki M. (1994). RNA/DNA hybrid duplexes with identical nearest-neighbor base-pairs hve identical stability. *FEBS Letters* 354: 74-78

Sugimoto N., Nakano S., Katoh M., Matsumura A., Nakamuta H., Ohmichi T., Yoneyama M., Sasaki M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* 34: 11211-11216

Sugimoto N., Nakano S., Yoneyama M., Honda K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nuc Acids Res* 24: 4501-4505

Xia T., SantaLucia J., Burkard M.E., Kierzek R., Schroeder S.J., Jiao X., Cox C., Turner D.H. (1998). Thermodynamics parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37: 14719-14735

For review see:

SantaLucia J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* 95: 1460-1465

Wetmur J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol* 26: 227-259

7   Files

*.nn Files containing the nearest-neighbor parameters, enthalpy and entropy, for each Crick's pair. They have to be placed in a directory defined during the compilation or targeted by the environment variable NN_PATH.

tkmelting.pl A Graphical User Interface written in perl/tk is available for users who prefer the 'button and menu' approach.

*.pl Scripts are available to use MELTING iteratively. For instance, the script multi.pl permits to predict the Tm of several duplexes in one shot. The script profil.pl allow an interactive computation along a sequence, by sliding a window of specified width.

8   See Also

New versions and related material can be found at `http://www.pasteur.fr/recherche/unites/neubiomol/meltinghome.html`

You can use MELTING through a web server at `http://bioweb.pasteur.fr/seqanal/interfaces/melting.html`

9   Known Bugs

The infiles have to be ended by a blank line because otherwise the last line is not decoded.

If an infile is called, containing the address of another input file, it does not care of this latter. If it is its own address, the program quit (is it a bug or a feature?).

In interactive mode, a sequence can be entered on several lines with a backslash

```
AGCGACGAGCTAGCCTA\
AGGACCTATACGAC
```

If by mistake it is entered as

```
AGCGACGAGCTAGCCTA\AGGACCTATACGAC
```

The backslash will be considered as an illegal character. Here again, I do not think it is actually a bug (even if it is unlikely, there is a small probability that the backslash could actually be a mistyped base).

10 Copyright

11 Acknowledgements

12 Author

Nicolas Le Novère,
EMBL-EBI
Wellcome-Trust Genome Campus
Hinxton Cambridge
CB10 1SD United-Kingdom
lenov@ebi.ac.uk

13 History

The version 4 has been entirely rewritten from zero. Therefore an history of previous versions would be of little interest.

See the file ChangeLog for the changes of the versions 4 and more recent.