

STtructure Analysis of **R**NA

REFERENCE
guide for
STAR

by
F.H.D. van Batenburg et.al

© Copyright 1993 F.H.D.van Batenburg. This manual may be cited at will and parts of it may be copied for non-commercial purposes after a previous permission in writing by the copyright-holder.
Group Bio-informatika of the Institute of Theoretical Biology
E.E.W. of Leiden University
Kaiserstraat 63, 2311GP Leiden,
The Netherlands
e-mail: SBQBEB@RULSFB.LeidenUniv.NL

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Batenburg, F.H.D. van

STructured Analysis of Rna / door F.H.D. van Batenburg ;

[ill. van de auteur]. - Leiden : groep Bio-informatika van het

Instituut voor Theoretische Biologie ;

ISBN 90-6385-162-6

SISO 573.2 UDC 681.3.06.577.21

Trefw.: programmeren (computer) biologie

Typewerk: Mw.H.Regeer & Mw.Y.Zitman

REFERENCE

guide

for

STructure

Analysis of

R_{na}

by

F.H.D. van Batenburg

V.W.A. Bos

J.P. Abrahams

M. van den Berg

S. Gulyaev

C.W.A. Pleij

J.J. Riethoven

Acknowledgements

STAR is developed by:

©dr.F.H.D. van Batenburg ¹⁾	dr.C.W.A. Pleij ²⁾
(1989) J.P. Abrahams ^{1) 2)}	(1990) M. van den Berg ¹⁾
(1990) V.W.A. Bos ¹⁾	(1991) J.J.M. Riethoven ¹⁾
(1992) M. Weynen ¹⁾	(1993) J. Th. Voskuilen ¹⁾
(1994) dr. S. Gulyaev ^{1 2)}	

Typing of the manual is done by H.Regeer and Y.Zitman.

We appreciate it if you refer to *STAR* and the theories behind it (greedy folding³⁾, stochastic folding⁴⁾ genetic algorithm folding⁵⁾) whenever you use *STAR* predictions.

Every effort has been made to ensure the accuracy of this document, but no responsibility can be accepted for any errors or omissions. *STAR* is licensed on an as-is basis; no warranties or liabilities as to any incidental or consequential damages, loss of revenues, reputations, grants or data apply.

We think that more than one copy of *STAR* is not unreasonable as long as only one copy of *STAR* is actively used at a time. Moreover, it may be used on more than one machine for educational purposes only and if the machines are at a single location (classroom). Reselling or redistribution of *STAR*, or its parts, in another program system is forbidden. We have no intention of enforcing our copyrights by legal procedures, but appeal instead to your honesty to comply with these rules.

This software was developed using APL.68000, a proprietary product of MicroAPL Ltd, which has given permission for a runtime version of APL.68000 to be included with the software. Copyright and all intellectual property rights in APL.68000 remain vested in MicroAPL Ltd. APL.68000 is a trademark of MicroAPL Ltd.

Apple and Macintosh are trademarks of Apple Computer Inc.

Atari and ST are trademarks of Atari Corp.

The PC version is implemented in APL*PLUS-III, a product of Manugistics.

Manual version 2.0.

Program version 3.0 March 1994; version 4.0 March 1995.

¹ Institute of Evolutionary and Ecological Sciences, Department of Theoretical Biology, University of Leiden, the Netherlands.

² Department of Biochemistry, Gorlaeus Laboratories, University of Leiden, the Netherlands.

³ Abrahams, J.P. & Berg, M.v.d. & Batenburg, F.H.D. & Pleij, C. (1990): Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. *Nucleic Acids Res.* 18(10)3035-3044.

⁴ Gulyaev, A.P. (1991): The computer simulation of RNA folding involving pseudoknot formation. *Nucleic Acids Research* Vol.19(9)2489-2494.

⁵ Batenburg, F.H.D. van & Gulyaev, A.P. & Pleij, C.W.A. (1995): An APL-programmed Genetic Algorithm for the prediction of RNA secondary structure. *J.Theor.Biol.*, 174, 269-280.

Gulyaev, A.P. & Batenburg, F.H.D. van & Pleij, C.W.A. (1995): The computer simulation of RNA folding pathways using a genetic algorithm. *J.Mol. Biol.* 250, 37-51.

1. ABOUT THIS MANUAL

This chapter shows you the organization of this manual. Furthermore it introduces you to the conventions that we use in this manual.

Therefore it is useful to skim over this chapter before you read on and know what conventions are used. Don't study this chapter too hard but concentrate on the other chapters instead. You can always look back to this chapter if some conventions appear.

1.1 THE ORGANIZATION OF THIS MANUAL

We divided this manual into three parts.

First, the introductory part that consists of chapters 1 to 3. This gives some general information about *STAR* and its use.

Second, the reference part extends from chapter 4 to 9. This part explains all menus of *STAR* in detail.

Finally the appendices summarise information that you need frequently while using *STAR*.

At the start of each chapter we present a brief introduction about its contents and its purpose.

This manual uses several conventions. We list them in this remaining paragraphs.

1.2 NAMING OF KEYS

When working with *STAR* you can use several keys of the keyboard. Therefore we will explain which key corresponds with a certain name in the following paragraphs.

[CR]-key: this is the key marked with "ENTER", "RETURN" or a large hooked arrow: ↵

[Co]-key: this is the key marked with the cloverleaf-like sign on Macintosh computers (called "command"-key) and [Control] or [Ctl] on Atari and PC. Normally this key is used in combination with another key. We write this is [Co+V] to indicate that key [V] should be pressed while [Co] is being pressed.

[Backspace]: this is the key marked with "Backspace" or a large leftward pointing arrow.

[Delete]: this is the key marked "Delete", "Del" or with a big cross. On machines without that key you can move the cursor one position forward with []] and subsequently use [Backspace].

[Esc]-key: this is the key marked with "ESC" or "ESCAPE".

1.3 CONVENTIONS USED

When we talk about a "**key**", we mean a key from the keyboard. With a "**button**" we mean either the mouse-button (-key) or a button in a box-frame. An example of a button in a box-frame is the [OK] button.

For "ease of speak" we use shorthand "press [OK]" instead of "position the mouse pointer above the box-frame titled "OK" and press the left button on your mouse". We also use such shorthand as "press [Desk]" for clicking on a menu.

We write **keys**, **menus**, **menu-options** and **buttons**, like OK buttons, between []. For example, the [Files] menu and the [Backspace]-key.

We write **names** of things such as folders between "...", for example: "foldername".

We enclose **options** in a menu with []. We separate the menu and its option with /. For example the [Primary/Open...] option.

Most of the time we don't include the menu name for an option. In that case we mean the menu that we discuss in the current chapter. For example: we discuss the [Primary] menu in chapter 6. In that chapter we refer to the [Primary/Open...] option as [/Open...].

When we refer to a **box**, we mean the window that appears on the screen with information and buttons in it.

In order to press a **button**, you should point the mouse-arrow to that button and press the mouse-button. For example "press the [OK] button" means that you should point your mouse-arrow to the [OK] button and press the mouse-button.

Some boxes have only one button, for example an [OK] button. You can leave these boxes by pressing that button, but you can also press the [CR] key. This is the key marked with "Enter", "Return" or with a large hooked arrow.

Some boxes have two buttons to leave the box, for example an [OK]- and a [CANCEL]-button. In this case you can also leave the box by pressing the [CR]-key. This is equal to selecting the button with the bold edges. If both buttons look the same, pressing the [CR]-key is equal to selecting the leftmost button of the box.

For example: pressing the [CR]-key in the following boxes has the same effect as pressing the [CANCEL] button.

Pressing [CR] has the same effect as pressing [CANCEL]

CANCEL

OK

Pressing [CR] has the same effect as pressing [CANCEL]

OK

CANCEL

Finally we want to define two frequently used terms: sequence and structure.

With **sequence** we mean the nucleotide sequence of the RNA molecule, the primary structure of the RNA.

With **structure** we mean the secondary structure of the RNA (with some tertiary elements; see paragraph 2.3).

2. THE THEORY UNDERLYING *STAR*

This chapter explains the theory that is the basis of *STAR*. We advise you to read this chapter at least superficially. That enables you to make full use of the capabilities of *STAR*.

2.1 INTRODUCTION

The folding of a single stranded RNA molecule is, at least for a great part, determined by its nucleotide sequence. This is the *primary* structure of RNA.

The formation of base pairs like A-U, G-C and G-U gives rise to specific structural motifs. Such motifs are stem regions and single stranded regions like hairpin, bulge, multibranched and internal loops. The ensemble of these structural elements in a linear presentation is the *secondary* structure of RNA.

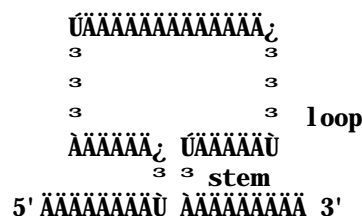
As soon as a set of free energy values for the various structural elements was available, attempts were made to predict the most probable secondary structure for a given RNA sequence.

Several algorithms have been developed for this purpose and they have become an additional tool in the determination of the structure of RNA. The most well-known are the minimal-energy algorithms. All algorithms of *STAR* are different in that they follow a “pathway” to reach the final structure. This can lead to the minimal energy configuration, but not necessarily so. We think that following the proper folding pathway is superior to the minimal energy approach, because this is what happens in nature.

2.2 FOLDING OF RNA

We consider the formation of a RNA secondary structure a stepwise process. During this process, intermediate structures evolve by subsequent addition of stems. Of all possible stems, only a few are realised.

The distance between two stemhalves is measured by counting the number of nucleotides in the loop which results upon the formation of a stem. This distance determines the free energy needed to bring the two stemhalves to each other. This free energy is an important factor in the rate of stem formation.



Once the stem is formed, it is stable by itself, or last long enough to be stabilized by the formation of another string. The free energy that is released by the formation of a stem determines the efficiency of nucleation and the stability of that stem.

The rate of formation of a stem thus depends on two types of free energy. These are the energy uptake for closing the loop and the energy release because of the pairing and stacking of the bases.

Energy values of various loops and stacks of basepairs have been published. Almost all RNA secondary structure programs use these energy values.

2.4 TERTIARY INTERACTIONS

Because the energy of stacking stems upon each other is important for the formation of the final structure, we enumerated all possible ways in which this stacking can be reached. You can see them in the figures below. In this figures the "3" represent the RNA sequence, and ":" the basepairings.

The diagram illustrates the difference between a fork and a bulge loop in RNA secondary structure. It shows two RNA sequences, UAAAAAAAAA, with different loop and bulge configurations.

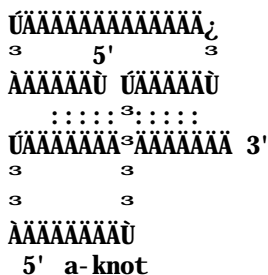
Fork: The sequence UAAAAAAAAA is shown with a loop of 3 nucleotides (3) and a bulge of 3 nucleotides (3). The fork structure is represented by the sequence UAAAAAAAAA with a loop of 3 nucleotides (3) and a bulge of 3 nucleotides (3).

Bulge loop: The sequence UAAAAAAAAA is shown with a loop of 3 nucleotides (3) and a bulge of 3 nucleotides (3). The bulge loop structure is represented by the sequence UAAAAAAAAA with a loop of 3 nucleotides (3) and a bulge of 3 nucleotides (3).

[illegible]

UAAAAAAAAAAAAAAAAA
 3' 3' 3'
 AAAAAA AAAAAA
 :::::3:::::
 5' AAAA3AAAAAAA
 3 3
 3 3
 AAAAAAAAAA
 3' a-knot

⁷ Freier, S.M. et al.(1986): Improve free-energy parameters for predictions of RNA duplex stability. Proc.Natl.Acad.Sci.USA Biochemistry 83: 9373-9377.



Note the 3' and the 5' a-knot. These pseudoknots are purely hypothetical so far; their existence in nature has not yet been proven.

In these two types of pseudoknots you can distinguish three different domains.

1. First, the quasi-continuous helical region.
2. Second, one single stranded stretch which doesn't cross any groove.
3. And third, one single stranded stretch crossing the shallow groove in the 5'a-knot, or crossing the deep groove in the 3'a-knot.

The energy values which the program assigns to these regions depend on:

1. the number of bases crossing a groove.
2. the number of base pairs to be crossed.
3. the type of groove to be crossed (deep, shallow or none).

The default energy values that are built in the program for these structures are based on an analysis of experimentally proven pseudoknots. We prohibit the formation of other pseudoknots by assigning a value of 99.9 kcal/mol to the single-stranded stretches crossing both grooves.

2.5 THE ALGORITHMS

Unlike most of the RNA folding programs, the algorithms in *STAR* do not necessarily calculate the most stable structure for a given RNA sequence. Instead of this, they simulate the folding of the RNA molecule.

Although a RNA structure can form rapidly in nature, it does not form instantaneously. Instead we envision a stepwise process. As we see it, the sequential formation of several intermediate structures gives rise to the final, native structure.

STAR supports 3 algorithms: greedy folding, stochastic folding and folding by genetic algorithm.

2.5.1 Greedy Folding

The greedy algorithm simulates the formation of a RNA structure by adding one stem at a time to the growing structure. Each incorporated stem is the one that adds most to the stability of the structure at that point. For this reason P.Higgs (Univ. Of Sheffield) coined this algorithm the "greedy algorithm".

A true simulation should take both the rate of formation and the rate of destruction of stems into account. For the sake of simplicity the greedy algorithm presumes that each intermediate structure is well defined and more or less stable.

In this way you can describe the addition of stems to the intermediate structure as a series of linked equilibria. As soon as a local equilibrium is reached, *STAR* seeks the next most favorable one.

Summarized, the algorithm of greedy folding is as follows:

1. Find all possible stems of the given RNA sequence.
2. Add stems to the structure until the free energy of the structure can not decrease anymore:
 - 2.1. Determine the next best stem.
 - 2.2. Add this stem to the structure.
 - 2.3. repeat 2.1.
3. Output the structure.

2.5.2 *Stochastic Folding*

The stochastic algorithm also folds RNA through a procedure of stepwise improving local equilibrium structures. Similarly to greedy folding, stochastic folding does not (explicitly) take disruption of stems into account. The essential difference is in the way to add stems to the structure, that is, to determine the stem most likely to form at any subsequent folding step.

In greedy folding (2.5.1) each step adds the stem with the greatest energy contribution. This ignores the possibility that a particular ensemble of stems could add more to the stability of structure than the stem with the best energy gain.

The stochastic algorithm tests if some stem *combinations* are better than the (energetically) best stem. Thus this algorithm applies a Monte Carlo procedure. At every step, the addition of stems to the final structure starts with the generation of a "population" of randomized structures. This population uses the most stable stems that are compatible with the previously formed (intermediate) structure.

Next, stems are added with probabilities depending on their energy values. In the current version, the probability to add particular stems is proportional to kinetic parameter K:

$$K = \frac{\text{energy gain}}{\text{destabilizing energy of loop.}}$$

After the random structure generation, the best structures (currently, three) are taken. Those stems that are common in all are added to the final structure. The addition uses a priority of descending order for kinetic parameter K.

In the next steps this whole procedure is repeated. Newly formed random structures add new stems that are compatible with the previously formed (intermediate) structure.

Furthermore, the algorithm can simulate RNA synthesis during folding. This is done by starting the folding in a small initial part at the 5' end of the RNA chain and "grow" by increasing this string with small increments. "Growth" increments the chain at the end of iterations when the number of stems compatible with those previously included to structure is small.

Summarized, the algorithm of stochastic folding is as follows:

1. Find all possible stems of the given RNA sequence.
 2. For growth simulation, start with 5' subset of the chain, otherwise take whole sequence.
 3. Add stems to the structure until the free energy of the structure does not decrease anymore:
 - 3.1 Generate random structures using stems compatible with those already added to structure (if any).
 - 3.2 Determine which stems are present in all best random structures.
 - 3.3 If these stems decrease the free energy, add them to the structure.
 - 3.4 Determine new subset of stems compatible with those already added to structure, the 3'-ends of stems being restricted by current chain length.
 - 3.5 If the number of these stems is too small, increase the chain length.
 - 3.6 Repeat from 3.1.
- Output the structure.

2.5.3 Genetic Algorithm

"Genetic algorithm" (GA) is the name for a special type of searching algorithm. The name derives from the fact that the procedure simulates the fitness-determined competition between individuals of a population, similarly to the process of natural evolution.

In our application, the individuals of the population are the folded structures that evolve during the simulation. The main advantage of this procedure is that it considers the competition of structures rather than that of possible stems (like in previous options 2.5.1 and 2.5.2). This allows the *disruption* of some stems in intermediate steps. Thus the algorithm simulates RNA folding pathway in some "energy-landscape" which overcomes energy barriers by stem disruption and descends into energy minima by stem formation. Generally, this seems to be the best way to simulate the RNA folding process.

At any point in the simulation, GA deals with a "population" of structures, which are produced by a randomized procedure, somewhat similar to the stochastic folding. Initially the population contains only unfolded states. Every GA iteration (the unit of the main program cycle) produces some new structures by changing previous ones and selecting the fittest solutions, thus yielding a new population with the same number of structures.

The *selection* procedure usually improves the fitness (free energy) of structures in the population, finally producing relatively stable structures.

The *changes* in structures are "GA mutations" (disruption as well as formation of some stems) and "GA crossovers" (the randomized generation of a new structure by combining stems from several parental solutions into compatible combination).

The probabilities of stem addition and disruption depend on energy contributions of stems: the more stable the stem, the greater the probability to include it to the structure and the less the probability to remove it.

Every structure in the population is mutated once and the crossover includes all stems of all current structures. So a population of N structures produces a temporary population of $2N+1$ possible structures. The new population is created by selecting the N best structures according to the fitness (free energy).

This completes one GA iteration. This procedure is repeated again until no improvement is found.

The GA also performs "growth" in a manner which is similar to that option in the Stochastic algorithm. The folding of the uncompleted RNA chain is simulated by starting with a small part at the 5' end of the chain (about 100 nucleotides). The chain length increases with increments of which the size depends on the free energy of structures in current population.

Summarized, the algorithm is as follows:

- 1 Find all possible stems of the given RNA sequence.
2. Start with 5' subset of the chain.
3. Form population of (10) random determined structures in the subset.
4. Perform Genetic Algorithm: iterate until the free energy doesn't increase for some number of iterations (e.g. 5).
 - 4.1 Mutate each structure.
 - 4.2 Disrupt stems in each structure.
 - 4.3 Add stems to each structure.
 - 4.4 Perform crossovers between the structures.
 - 4.5 Form new population by selecting the fittest structures.
 - 4.6 If energies of structures are not considerably improved, increase chain length.
 - 4.7 Repeat from 4.1.
5. Output the structure.

2.5.4 *Pros And Cons*

The greedy algorithm is the quickest of all algorithms. It yields rather nice predictions because it does not necessarily produces the structure with the lowest-energy value, but follows the folding pathway to reach its final prediction. The quality is comparable to that of minimum-energy programs. Sometimes minimum-energy programs are better, sometimes the greedy algorithm is better (in particular in case of short-range interactions and pseudoknots).

The stochastic algorithm is an improvement over the greedy algorithm. In particular if ensembles of moderate valued stems are better than the next-best stem in the folding process. This slight increase in quality is offset by a considerable increase of computer time though. The algorithm can require 10 times (or more) as much computer time as the greedy algorithm.

The genetic algorithm follows most faithfully the folding pathway of synthesized RNA. So its result is the most reliable of most algorithms. This can take considerable more computer time though, the algorithm is often 10 to 100 times as slow as the greedy algorithm.

3. THE PROCEDURE OF *STAR*

The purpose of *STAR* is to compute and show a RNA structure.

To do so, you should follow a number of consecutive steps. The order of those steps is not arbitrary. Therefore, we list these steps in the following paragraphs.

You can see this procedure summarised in the chart of appendix A.

3.1 STEP 1: GET YOURSELF A SEQUENCE

There are two ways to get a sequence in the computer.

The first method is to ***type*** in a sequence yourself. You can do this using the option [Primary/Edit]. Look in paragraph 7.6 for a detailed explanation of this option.

This will result in the information [...E/] in the rightmost menu of your screen.

The other method is to ***open*** a sequence file from disk use option [Primary/Open...]. This sequence file can either be a sequence you have saved once with *STAR*, or a file from a different origin. For example files from a text editor or a database⁸. You can use many file types here, as long as the contents of the file are coded in ASCII.

We explain this option in paragraph 7.2. Opening a sequence-file will result in the information [filename.1/] in the rightmost menu of your screen.

⁸ See Appendix D for detailed information.

3.2 STEP 2: GET YOURSELF A STRUCTURE

After you have a sequence you get yourself a structure. There are three different methods to do this.

The first method is to ***type*** one in yourself using the option [Secondary/Edit]. You find an explanation of this option in paragraph 8.8. This results in the information [.../E] in the rightmost menu of your screen.

The second method is to ***open*** a structure file using option [Secondary/Open...]. This sequence file can either be a sequence you have saved once with *STAR*, or a file from a different origin. For example files from a text editor or a database⁹. You can use any file type here, as long as the contents of the file are coded in ASCII. We explain this option in paragraph 8.2. This results in [.../2] being written in the rightmost menu of your screen.

The third method is to let *STAR* ***calculate*** a structure using option [Secondary/Calculate]. This option is what you want to do usually and it is the essence of *STAR*. The structure is then based on the previously acquired sequence (step 1). This results in [.../C] in the rightmost menu of your screen.

⁹ See Appendix D for detailed information.

If you want to use default parameters this is sufficient. If you want to change the default parameters you must specify this in [Secondary/Factors] before you start the calculation.
See paragraph 8.5 [/Factors] for an explanation of this option.

3.3 STEP 3: STUDY RESULT AFTER CALCULATION OF THE STRUCTURE

STAR displays a table with numbers that indicate which nucleotides form stacks.

To see the same information more visually, use [Secondary/View]. You can choose one view out of many. The most compact view is Lego-view, but this view cannot show pseudoknots. The most direct one is Span-view.

Pseudoknots are most easily spotted using Bracket-view.

If you want to study the results more quietly, use [Secondary/Print]. This will present you with the same assortment of views and will print the result.

If you have any printer problem you can use [Secondary/Print] to send the data to file. This enables you to print using your own textprocessor or it enables you to paste the view in your publication. Use a non-proportional font, to display the view properly.

4. THE *STAR* MENU

STAR presents the following menu-bar:

Desk	Files	Primary	Secondary	Energyrules	_. _
------	-------	---------	-----------	-------------	------

As the figure above shows, the *STAR* menu-bar consists of 6 menus: [Desk], [Files], [Primary], [Secondary], [Energyrules] and [_. _].

The [_. _] menu is not a normal menu because its name changes while you use *STAR*. The purpose of this menu is to show you the status in which *STAR* is at a certain moment. For example, if you have opened a sequence file, the name of this menu is the name of the file that you opened. Unlike the other menus we don't discuss this menu in a separate chapter. We explain it while we discuss the other menus.

first we discuss these menus briefly in this chapter, so you will understand their organization. Later on we elaborate on each menu in separate chapters.

4.1 THE [DESK] MENU

Desk	Files	Primary	Secondary	Energyrules
About Star				

The [Desk] menu contains the option [About *STAR*], which contains information about the program *STAR*. For more information see chapter 5.

4.2 THE [FILES] MENU

Desk	Files	Primary	Secondary	Energyrules
	Help ----- Rename... Delete... ----- Quit...			

The [Files] menu contains -as suggested by the name- some options that use or handle files. These are [/Rename...] and [Delete...].

Furthermore, this menu contains some miscellaneous options: [/Quit...] and [/Help].

The first option, [/Help] gives brief information about the options in this menu.

It goes without saying that you use option [/Quit] to end the program.

See chapter 6 for a detailed explanation of the other options in this menu.

4.3 THE [PRIMARY] MENU

Desk	Files	Primary	Secondary	Energyrules
		Help ----- Open... Save as... Save ----- View Edit ----- Print...		

The [Primary] menu contains all the options that operate on sequences, that is the primary RNA structure.

The first option, [/Help] gives brief information about the options in this menu.

The next three options, [/Open...], [/Save as...] and [/Save] let you retrieve or save sequences. You can for example edit a sequence and save the changed sequence for later use.

Use option [/View] to examine and [/Edit] to enter or change a sequence.

The last option, [/Print...] enables you to make a hardcopy (a print) of a sequence or write the sequence to a textfile for use in other programs.

You find a detailed explanation of the options in this menu in chapter 7.

4.4 THE [SECONDARY] MENU

Desk	Files	Primary	Secondary	Energyrules
			Help ----- Open... Save as... Save ----- Factors... Calculate ----- View... Edit ----- Print...	

The [Secondary] menu contains all the options that operate on the secondary structure of RNA.

The first option, [/Help] gives brief information about the options in this menu.

With the next three options ([/Open...], [/Save as...] and [/Save]) you can retrieve or save structure files; you can for example save calculated structures for later use.

You will use the following two options ([/Factors...] and [/Calculate...]) for the calculation (the folding) of the secondary structure. In [/Factors...] you choose the required algorithm. You can also deviate here from some default parameters. Option [/Calculate...] instructs *STAR* to do the calculations.

Use option [/View] to examine and [/Edit] to enter or change a structure.

The last option, [/Print...], enables you make a hard-copy (a print) of the structure or write the structure to a text file for use in other programs.

You find a detailed explanation of the options in this menu in chapter 8.

4.5 THE [ENERGYRULES] MENU

Desk	Files	Primary	Secondary	Energyrules
				Help

				Open...
				Save as...
				Save

				View...
				Edit...

				Print...

The options in the [Energyrules] menu involve the energyrules that *STAR* uses for the calculation of a secondary structure.

As you start *STAR* the program already contains default energyrules.

If you are satisfied with those energyrules, and you don't want to experiment with other energy values, you can ignore the options in this menu. In that case you can skip chapter 9, which explains that menu.

The first option in this menu, [/Help] gives brief information about the options in this menu.

The following three options, [/Open...], [/Save as...] and [/Save] allow you to retrieve or save energyrules files.

Use option [/View] to examine and [/Edit] to enter or change energyrules.

The final option, [/Print...], enables you to make a hard-copy (a print) of the energy values or write the energyrules to a textfile for use in other programs.

You find detailed explanations of these options in chapter 9.

4.6 SUMMARIZING: THE STRUCTURE OF THE [PRIMARY], [SECONDARY] AND [ENERGYRULES] MENUS

We hope that you have noticed the structure of the most important *STAR* menus: [Primary], [Secondary] and [Energyrules].

The top one is always a [/Help] option which gives a brief explanation about the other options in that particular menu.

The following three options are always the file-options [/Open...], [/Save as...] and [/Save]. Furthermore each menu contains the options [/View...] and [/Edit...] to examine and to edit.

And the last option in each menu is [/Print...] which allows you to make a hardcopy on paper or to create a text file.

The menu [Secondary] contains two "extra" options for the calculation of the structure: The first one is [/Factors...] to set several parameters. The other one is [/Calculate...] which instructs *STAR* to calculate a secondary structure.

4.7 THE [__.__] MENU

The right-most menu presents information about RNA data, available in the program.

If no RNA information is available, this menu shows "__.__".

If a primary sequence is available, this menu shows the name of the file, for example "STAR.1/_", or "E/_ if the sequence was entered using the editor.

If information about the secondary structure is available, the characters 2, C and E are used for read, computed or edited secondary structures. For example:

- STAR.1/2 means that the secondary structure was read from a structure file STAR.2.
- STAR.1/C means that the secondary structure was computed from a sequence file STAR.1.
- STAR.1/E means that the secondary structure was entered using the editor.

5. THE [DESK] MENU

Desk	Files	Primary	Secondary	Energyrules
About Star				

The [Desk] menu-title only has option [/About].

5.1 [/ABOUT STAR]

The [/About STAR] option displays information about the program and the team behind it.

Notice that *STAR* is copyrighted, so please don't make any illegal copies. As we wrote in the acknowledgement we have no intention to enforce this request in court, but we appeal to your honesty instead.

When you close this informative window, you return to the main menu.

6. THE [FILES] MENU

Desk	Files	Primary	Secondary	Energyrules
	Help			

	Rename...			
	Delete...			

	Quit...			

In this chapter, we explain all the options of the [Files] menu in detail.

6.1 [/HELP]

The [/Help] option displays brief information about the options in this menu.

The main purpose of this option is to give you some help at your fingertips, without having to consult the manual. Now don't throw away this manual as this option is only a brief reminder.

When you close this information window, you leave this option and return to the main menu.

6.2 [/RENAME...]

The [/Rename...] option enables you to rename a file that is on disk.

When you activate this option a file-selector appears to select the file that you want to rename.

After this the file-selector appears again and you can enter the new filename.

1 Possible errors

ALREADY EXISTING FILENAME:

When the new filename already exists, *STAR* warns you with the following box:

New filename already exists!
Do you wish to continue

OK

CANCEL

When you press [OK], *STAR* replaces the old file by the new one. If you don't want to replace the file, press [CANCEL].

INSUFFICIENT MEMORY:

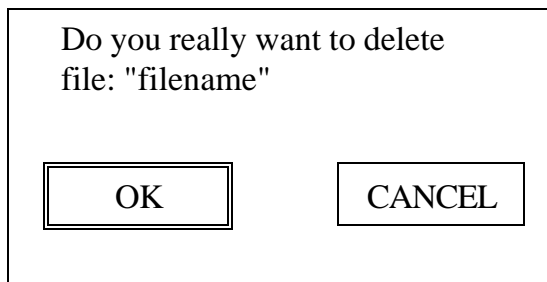
As *STAR* renames files by making a copy and after that destroying the original, memory can be a problem. In this case your original file still exists.

6.3 [DELETE...]

The [Delete...] option enables you to delete (erase, remove) a file that is on disk.

When you activate this option a file-selector appears. With this file-selector you select the file that you want to delete.

After this, the warningbox shown below appears. With this box, *STAR* asks you to confirm the deletion of the file.



A rectangular warning box with a thin black border. Inside, the text "Do you really want to delete file: 'filename'" is centered. Below the text, there are two buttons: "OK" on the left and "CANCEL" on the right, both with thin black borders.

When you press [OK], *STAR* deletes the file. If you don't want to delete the file after all, press [CANCEL].

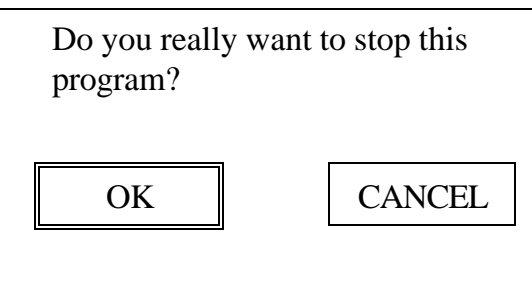
6.4 [/QUIT...]

The [/Quit...] option allows you to stop using *STAR* the smooth way.

This in contrast to the rough way: pressing the reset key or, even cruder, turn off the power.

When you activate this option the warningbox shown below appears.

With this box, *STAR* asks you to confirm the termination of the program.



A rectangular warning box with a thin black border. Inside, the text "Do you really want to stop this program?" is centered. Below the text, there are two buttons: "OK" on the left and "CANCEL" on the right, both with thin black borders.

If you don't want to stop *STAR*, press [CANCEL]. If you do want to stop *STAR*, press [OK] and you will return to the desktop.

If you changed your sequence, structure or energyrules, make sure that you saved them before you press [OK]. Otherwise you will lose them.

7. THE [PRIMARY] MENU

Desk	Files	Primary	Secondary	Energyrules
		Help ----- Open... Save as... Save ----- View Edit ----- Print...		

In this chapter we explain all the options of the [Primary] menu in detail.

7.1 [/HELP]

The [/Help] option displays brief information about the options in this menu.

The main purpose of this option is to give you some help at your fingertips, without having to consult the manual. Now don't throw away this manual for this option is only a brief reminder.

When you close this informative window, you return to the main menu.

7.2 [/OPEN...]

The [/Open...] option enables you to open a sequence file from disk.

This means that *STAR* transfers a copy of a specified sequence from disk into computer memory.

When you activate this option a file-selector appears to select the sequence file you want to open.

Note that you can open any file. However not all files contain valid sequences. So be sure that the file you are opening contains a valid sequence. To avoid confusion, we strongly suggest that you give all your sequence files a recognizable name, for example with the extension ".1". The extension is the part of a filename after the point ".". Examples are STAR.1, TMVRNA.1 and TRNATHRT.1.

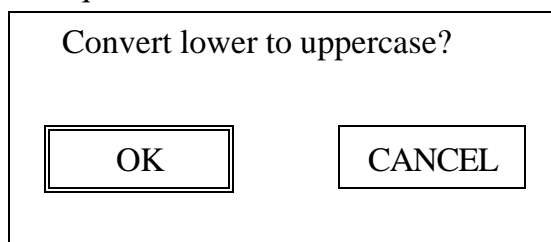
STAR can read EMBL files as well as files that contain ACGU characters exclusively ¹⁰.

¹⁰ See Appendix D for detailed information.

Now you might think "why all this trouble in preventing me to open a non-sequence file, why doesn't *STAR* do that for me?". The answer is: in this way you are able to open sequence files originating from databases or from other editors than the *STAR* editor.

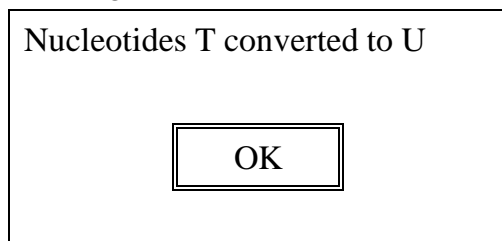
7.2.1 Lower case characters

If your sequence file contains lowercase characters *STAR* asks you what to do. It displays the question:



7.2.2 Thymine

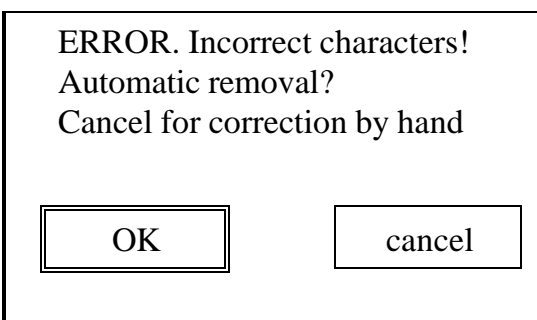
If your sequence file contains Thymine (besides A, C, G, or U) *STAR* converts T automatically to U. It will inform you of this action by the message box



You can only agree with this action by pressing [OK]. If you don't agree, you still have to press [OK]. Afterwards you always have the opportunity to activate the editor, using [Primary/Edit] and make changes.

7.2.3 Incorrect characters

If your sequence file contains any other character besides TACGU, *STAR* asks you what to do. It displays the question:

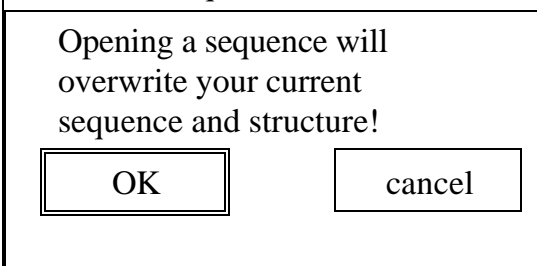


If you choose [OK], *STAR* will remove all offensive characters, but this is very dangerous: are you absolutely sure that you are not converting rubbish?

If you choose CANCEL, *STAR* will place the data in the editor and enable you to correct offensive characters. These characters are indicated by a caret “^” underneath. Correct them and exit the editor by closing the window.

7.2.4 Old and new sequence

If you want to open a sequence file while you already have a sequence and/or a structure in your computer, the following box appears. With this box, *STAR* asks you to confirm the opening of a (new) sequence file.



If you press [OK], *STAR* erases the sequence and the structure that are in your computer and you are able to open another sequence.

This means you loose the sequence and the structure in the computer! So, please be sure that you have *saved* your current sequence and structure or that they are *useless* before you press [OK].

After you have succesfully opened a sequence file, the status-indicator menu shows the name of the opened file. As explained in chapter 4, this is the rightmost *STAR* menu.

7.2.5 Save sequence

If *STAR* changed something (convert T to U, or convert lowercase to uppercase) or if you changed something yourself (correction by hand or adding comments) the original file is **not** changed. Everything was only done in computer memory. If you want to fix those changes in the original file, you should save your sequence with the options [/Save] or [/Save as...].

1 Possible errors

1. INSUFFICIENT MEMORY:

A very obvious error occurs when you open a sequence file that is too big for the memory of your computer. This results in a "INSUFFICIENT MEMORY" error.

Unfortunately you can do nothing else than allot more memory to *STAR* if available and start *STAR* again or start *STAR* all over again on a computer with more memory.

2. TOO MANY CHARACTERS:

When you open a sequence-file, and subsequently try to edit it in the *STAR* editor, it can happen that there are too many characters in a line. If that is the case, *STAR* warns you with the following box:

Some lines have been truncated.
Select cancel to quit from
the editor

CANCEL

CONTINUE

When you press [CANCEL], nothing will happen and you will return to the main menu. When you press [CONTINUE] also nothing serious will happen. The only thing *STAR* does, is using a few extra lines. You will not loose any nucleotide or comment.

7.3 [/SAVE AS...]

The [/Save as...] option enables you to save a sequence to file. This means that *STAR* transfers a copy of the sequence in the computer's memory to disk. You must enter the name of the file you want the sequence to save to.

As suggested earlier, we recommend that you use a name that ends with ".1" for a primary sequence. For example TMVRNA.1 or TMVRNA.1A or TMVRNA.1R1 .

When you activate this option a file-selector appears which enables you to enter the filename under which *STAR* should save the sequence.

Of course there must be a sequence in the computer's memory before you can save it to disk. Therefore you can only activate this option if the computer's memory holds a sequence.

This means that either you must have opened (option [/Open...]) or you must have entered (option [/Edit]) a sequence before you can use this [/Save as...] option.

If you enter a filename that already exists, the following box appears. With this box *STAR* asks you if you want to overwrite the file.

Replace existing "filename"?

Cancel

Replace

If you press[/Replace], *STAR* saves the sequence to disk. It overwrites the sequence on disk. Otherwise press [Cancel]. Then *STAR* overwrites nothing and you return to the main menu.

After you have saved your sequence, the status-indicator menu replaces the old filename and shows the new name you entered.

7.4 [/SAVE]

The [/Save] option enables you to save a sequence file for later use.

This means that *STAR* transfers a copy of the sequence in the computer's memory to disk.

The difference with the option [/Save as...] is that *STAR* saves the sequence to a file with the same name as the one that you opened or the one you saved to earlier, using the option [/Save as...]. That is why you must have opened a sequence file (option [/Open...]), or entered one with the editor and saved it with the option [/Save as...] before you can activate this [/Save] option.

STAR replaces the sequence on disk by the sequence that you [/Save] without warning. So please be careful: be sure that the sequence on disk is no longer useful before you use this option.

7.5 [/VIEW]

The [/View] option enables you to examine the sequence which is in the computer's memory.

This means, of course, that the computer must hold a sequence before you can [/View] one. So you must have opened (option [/Open...]) or have entered (option [/Edit]) a sequence before you can activate this option.

Now look as an example at the following sequence:

5	10	15	20	25	30	35	40	45	50	
UAGAG	CCGAG	CCUUG	GUAAA	AGGGU	GAGGU	CCCCG	GCAGU	UCCAA	UCUGC	50
CUAAU	UUUCA	GCACC	AGGGG	GGGGU	UUU					100

As you see, *STAR* divides the sequence into parts of 5 nucleotides. On each line *STAR* shows 50 nucleotides. The numbers start with 1 at the 5' part of the sequence, and increase until the 3' part of the sequence is reached.

7.6 [/EDIT]

The [/Edit] option enables you to change an existing sequence, or to create a new sequence. We explain the *STAR* editor in the following paragraphs in detail.

You can also edit sequences with a regular text editor before you open the sequence in *STAR*. Appendix D describes formats. After creation you can import such a file using option [/Open...] (see paragraph 7.2).

You enter the *STAR* editor by activating the option [/Edit] from the [Primary] menu.

When you enter the editor a new window appears. This window can be moved and resized, just like you usually do.

The editor shows you the sequence almost similar as in the [/View] option (paragraph 7.5) does, but it has not clustered the sequence in chunks of 5 nucleotides, nor does it display numbers.

STAR does not layout your text in the editor. Choose the best layout you want. You can use blanks, returns and empty lines whenever that suits you.

If you need these numbers while you are editing, you can first type them in the edit window. If you precede those numbers with an asterisk, *STAR* will ignore them. For example, you can type:

```
*   5   10   15   20 ... etc.
UAGAG CCGAG CCUUG GUAAA ... etc.
```

Everything typed after "*" will be ignored by *STAR* and considered as your personal comment.

In the following paragraphs we will describe how to:

1. Add nucleotides.
2. Add comments.
3. Delete nucleotides.
4. Move & copy nucleotides.
5. Leave the editor.

7.6.1 Add Nucleotides

You can add nucleotides by typing the letter corresponding with that particular nucleotide.

The keys that correspond with each type of base are:

a or A: Adenine
c or C: Cytidine
u or U: Uracil
g or G: Guanine

You can enter the nucleotides both in lowercase and uppercase.

However, *STAR* calculates a structure of uppercase nucleotides only!

This feature is particularly useful if you want to exclude a certain part of a sequence to basepair. You exclude that part by typing it in lowercase.

Normally you are advised to type with the [CapsLock] key pressed.

7.6.2 Comments

It is possible to add comments to your sequence while you are editing. These comments can only be viewed in the editor. When you select the option [Primary/View] the comments are left out.

Adding comments is quite simple. The only thing you have to do is typing an asterisk (="*"). To the right of that asterisk you can type any comment you want.

For example:

```
* Turnip yellow mosaic virus
*   5   10   15   20   25
CCCCC CGCAT CGACC TGCAT AATAA * 25
                                * CODON START AT 1
ctgta tcagg aactc tctcg atgc * lowe case
excluded
                                from calculation
```

Notice that blanks, empty lines and text after "*" are available for you to enhance readability.

7.6.3 Delete Nucleotides

If you want to delete one or a few nucleotides there are two ways to do so.

- If you press the [Delete]-key, *STAR* deletes (erases) the nucleotide at the cursor position.
- Another way is to press the [Backspace]-key. *STAR* deletes the nucleotide just left of the cursor.

If you want to delete a large string of nucleotides the following procedures are faster.

- First define a block of nucleotides by dragging the mouse over the sequence.
- After this you erase the block with nucleotides by [Co+X].

7.6.4 Move & Copy Nucleotides

Moving and copying nucleotides is a two-step procedure. First you transfer (a copy of) the particular nucleotides to the clipboard, secondly you paste the clipboard contents to the new, required position.

To **move** you drag the mouse over the nucleotides that you want to move.

Then you type [Co+X] to move the selection to the clipboard.

Finally you set the cursor at the new location and type [Co+V] to insert the selection to that new spot.

To **copy** some substring to another place, first drag the mouse over those nucleotides that you want to have a copy of.

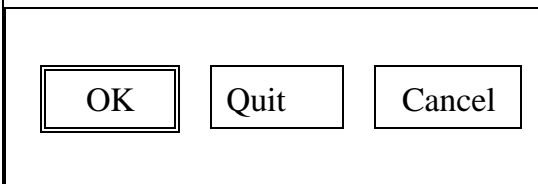
Then you type [Co+C] to place a copy of those nucleotides in the clipboard.

Finally, you set the cursor at the new location and type [Co+V] to insert the copy at that new spot.

7.6.5 Leaving The Editor

Leave the editor by clicking in the left top corner of your edit window.

You will see the following box¹¹.



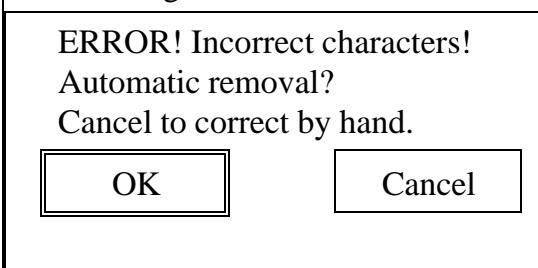
Choose [OK] if you are satisfied with your editing. Choose [Quit] if you want to discard your editing work. Choose [Cancel] if you want to resume editing.

The status indicator menu shows "___E". This is how *STAR* shows you that you entered a sequence in the memory of your computer with the editor.

1 Possible errors

ILLEGAL CHARACTERS:

If you want to leave the editor, and there are other characters in your sequence (no problem in comments) than A,C,G,U,a,c,g,u or spaces, then *STAR* warns you with the following box:



Just press the [Cancel]-button, and you will return to the editor- screen, where you can correct the wrong characters (maybe you did forget an asterisk before your comments). The wrong characters are easy to find, as *STAR* places a caret "^" underneath in a comment line.

¹¹ PC versions shows the [Exit] and [Quit] button underneath the edit-window.

7.7 [/PRINT...]

Use the [/Print...] option to make a hard-copy (a print) of your sequence, or to save the sequence to file in standard *STAR* layout. That standard layout is the layout that you see using [/View] ¹². You can only activate this option if you have a sequence in your computer's memory, of course.

When you activate this option one the following box appears:

PRINT OPTIONS

Page length ___ lines

Page width ___ columns

Top margin ___ lines

Bottom margin ___ lines

Print to: ☐File ☐Printer

OK

CANCEL

In the four fields marked by ____, you may enter the length and width of the paper you use, the top margin and the bottom margin. *STAR* already provided default values that are fine in most cases.

The top margin is the blank part above the text on a page. The bottom margin is the blank part below the text on a page.

For instance, if you use paper 71 characters long, enter "71" at the second field.

For a standard A4 paper format, you enter "66" for the pagelength, and "80" for the page width. For standard continuous forms paper, enter "71" for the pagelength and "80" for the pagewidth. We advise for both top and bottom margin a length of 2 lines.

You can specify whether to print to a printer or to a file. Printing to a printer means that *STAR* outputs directly to a printer. In this case a printer must be attached to your computer and it must be switched on.

7.7.1 Print to file

If you have any problem with printing, you can print to file. *STAR* asks you to enter a filename using the file selector. You can print this file separately after leaving *STAR*.

If you want to include the *STAR* output in a publication you can also use "print to file". The file can be read in any text processor. In this case you might prefer to set both top and bottom margin to 0 lines. Use a non-proportional font in your textprocessor to display numbers above the sequence properly.

1 Possible errors

1. NO PRINTER ATTACHED:

If you specify you want to print while there is no printer attached to your computer, you can not use *STAR* for about 30 seconds.

2. PRINTER BUSY:

This message can happen in network printers. *STAR* detects a busy signal and will not print. Unfortunately this is a network problem, not a *STAR* bug. Your best solution is to use option "Print to: File" and to print the file with another program (textprocessor).

¹² If you want your sequence saved the way you typed it or edited it (with comments and all) you should use [Primary/Save].

8. THE [SECONDARY] MENU

Desk	Files	Primary	Secondary	Energyrules
			Help ----- Open... Save as... Save ----- Factors... Calculate ----- View Edit ----- Print...	

In this chapter we explain all the options of the [Secondary] menu in detail. The most essential one is [/Calculate] to make a structure prediction.

8.1 [/HELP]

The [/Help] option displays brief information about the options in this menu.

The main purpose of this option is to give you some help at your fingertips, without having to consult the manual. Now don't throw away this manual as this option is only a brief reminder.

When you close this informative window, you return to the main menu.

8.2 [/OPEN...]

The [/Open...] option enables you to open a structure file from disk. This means that *STAR* transfers a copy of a specified structure from disk to the computer. Then you can use it with *STAR*.

These stems are regarded as "forced" stems. This means that they are always incorporated in a newly calculated structure. You can also change these stems in the editor ([Secondary/Edit]).

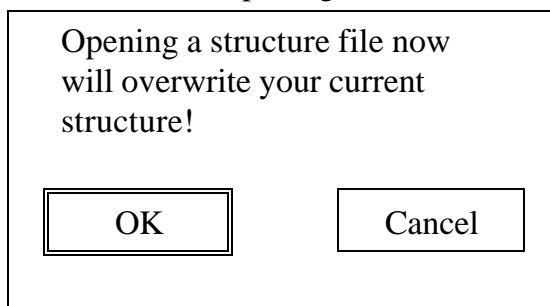
When you activate this option a file-selector appears by which you can select the structure file that you want to open.

You can open structurefiles that are made by *STAR*, but also ASCII-structurefiles are allowed.

¹³ A structure is closely related to the sequence from which it originates. Therefore you can only open structure files with a name that is the same as the filename of the sequence that you have opened. Furthermore, the name of the structure file should always have an extension that begins with ".2".

For example: you opened the sequence file "TMV.1". Now you can open only structure files beginning with "TMV.2". The remainder of the filename is not limited. For example: you can open the structure file "TMV.2A" or "TMV.2GA".

If you want to open a structure file, while you already have a structure in your computer, the box shown appears. With this box *STAR* asks you to confirm the opening of a (new) structure file.



If you press [OK], *STAR* erases the structure that is in your computer and you are able to open another structure. This means you lose the structure in the computer memory (not on disk of course)! So, please be sure that you have **saved** or **printed** your current structure or that it is **useless** before you press [OK]. You can save your structure with the options [/Save] or [/Save as...] or with [/Print] to file.

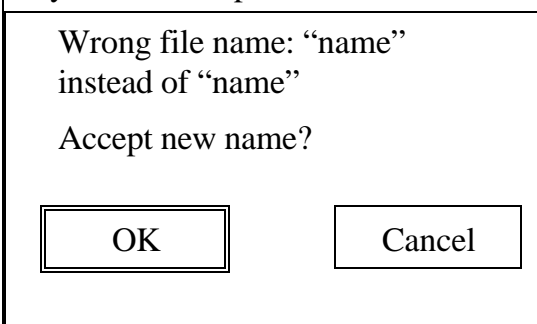
If you are not sure press [Cancel]. Then *STAR* erases nothing and you return to the main menu.

1 Possible errors

1. **WS-FULL:** A very obvious error is when you want to open a structure file that is too big for the memory of your computer. This results in a "WS FULL" error.

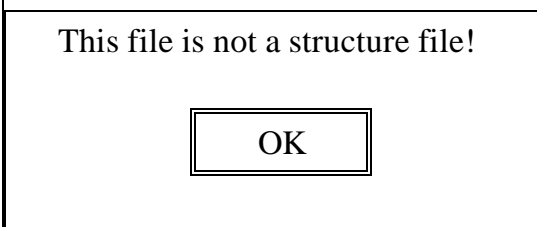
Unfortunately you can do nothing else than start *STAR* all over again on a computer with sufficient memory.

2. **WRONG FILENAME:** If you try to open a file with a wrong filename, the box shown below appears. With this box *STAR* warns you that you can not open the file.



If you are really sure that this file contains a structure of that sequence, change the filename.

3. **NO STRUCTURE FILE:** If you try to open a file that is not a structure file, the box shown below appears. With this box *STAR* warns you that the file you try to open is not a structure file.



All you can do is to press [OK], and you return to the main menu. View the file with a normal editor to see what is wrong with the file format ¹⁴.

¹³ See Appendix D for detailed information.

¹⁴ See Appendix D for detailed information.

8.3 [/SAVE AS...]

The [/Save as...] option enables you to save stems to file for later use. This file can be read using [Secondary/Open] in another session.

If you want to save a structure in a special view, you should use the option [Secondary/Print]. Usually, such a file cannot be read by *STAR*.

When you activate this option a file-selector appears that enables you to enter the filename under which *STAR* will save the structure.

As a structure is closely related to the sequence from which it originates, you can only save structures to a file with a special name.

The first part of the filename must be equal to the first part of the name of the sequence file that you opened. Furthermore, the extension of the filename must begin with ".2".

For example: if you opened the sequence "TMV.1", you can save your structure only to a file beginning with "TMV.2". The remainder of the filename is not limited. For example: you can save the structure to a file named "TMV.2A" or "TMV.2GA".

After you have successfully saved your structure, the rightmost part of the status-indicator menu shows "/2..".

For example, you opened the sequence file "TMV.1" and you saved a structure to a file named "TMV.2A". Then the status-indicator menu shows "TMV.1/2A".

1 Possible errors

1 WRONG FILENAME:

If you enter a filename that has not the same name as your sequence, the box shown below appears. With this box *STAR* warns you that the filename is not allowed and that it did not save the structure.

Filename: "wrong filename"
is not allowed.
Structure not saved.

OK

All you can do is to press [OK] and you return to the main menu. Then you can use the option [/Save as...] again to save to a file with a correct filename.

2 WRONG FILENAME:

If you enter a filename with an extension that does not start with ".2", the box shown below appears. With this box *STAR* warns you that the extension is not allowed and that it did not save the structure.

Extension not correct.
Use .2xx.
Structure not saved

OK

All you can do is to press [OK] and you return to the main menu.

Then you can use the option [/Save as...] again to save to a file with a correct filename.

3. ALREADY EXISTING FILENAME:

If you enter a filename that already exists, the following box appears.

Replace existing "filename"?

Cancel

Replace

If you press [Replace], *STAR* saves the structure to disk. It overwrites the structure on disk. Otherwise press [Cancel]. Then *STAR* overwrites nothing and you return to the main menu.

8.4 [/SAVE]

The [/Save] option enables you to save stems to file for later use. This file can be read using [Secondary/Open] in another session.

If you want to save a structure in a special view, you should use the option [Secondary/Print]. Usually, such a file cannot be read by *STAR*.

The difference with the option [/Save as...] is that *STAR* saves the structure to a file with the same name as the one that you opened, or saved earlier. That is why you should have opened a structure file (option [/Open...]), or saved earlier to a file with the option [/Save as...] before you can activate this option.

Furthermore, *STAR* replaces any existing structure file on disk without warning. So be careful with this option: be sure that the structure on disk is no longer useful before you use this option.

8.5 [/FACTORS...]

The [/Factors...] option enables you to change the default values for parameters that *STAR* uses for the calculation of a structure.

You can change these parameters using the box shown below. This box appears when you activate the [/Factors...] option.

You can enter the parameters on the fields marked by _____. As you can see we have already set parameters as default: for example 999 for number of potential stems.

To change the parameters you change some values and press [OK]. If you don't want to change them after all, press [Cancel].

ERROR. Error in numeric input!
Try again?

OK

CANCEL

When you press [OK], the box appears again with the offensive field replaced by "0" and you can change your input. When you select [Cancel] you return to the main menu, and *STAR* uses the values that were in the box before you started to change them.

GENERAL FACTORS

Analyse complete sequence:[NO]

Stop after stem ____999

SPECIFIC FACTORS

Analyse

From: ____1 To: ____76

Greedy folding

Change

No folding

____0 ____0

Stochastic folding

Change

No folding

____0 ____0

Genetic algorithm

Change

No folding

____0 ____0

No folding

____0 ____0

OK

Cancel

In the following paragraphs we will explain:

1. general factors
2. greedy folding factors
3. stochastic folding factors
4. genetic algorithm factors

1 Possible errors

ILLEGAL VALUES:

When you type any illegal value (i.e. the number of potential stems must be positive, and integer), *STAR* gives you the following warning:

8.5.2 [/Factors...]: General Factors

We will discuss 3 general factors:

1. "potential stems",
2. "stop after stem" and
3. "analyse complete sequence"/From-to.

1 Potential stems

As we discussed in paragraph 2.5, the first step of *STAR* is the generation of all possible stems in the given sequence. However the memory of a computer is limited. So if *STAR* would generate all possible stems this would certainly exceed the available memory; especially when *STAR* folds a large sequence. Therefore *STAR* limits the maximum number of generated stems.

The stems that *STAR* generates are the ones with the lowest free energy. In determining these stems, *STAR* assumes that the RNA has no secondary structure. However, it is possible that a stem that is excluded could have a very favourable energy value after some stems are formed. Nevertheless, once excluded it remains excluded forever. Therefore you should always set this parameter as high as possible regarding the memory of your computer.

Let us try make this clear with an example. Suppose you let *STAR* calculate (fold) a hypothetical sequence. You set the "Potential stems" parameter to 5. Suppose furthermore that *STAR* finds the following five stems:

```

26 34 70 78   P
30 34 67 71   increasing free energy
30 34 68 72   (ΔG lower)
30 34 69 73   P
35 39 47 51   ↓

```

If you don't know the meaning of these values, check paragraph 8.7.1. (the energy values of the hypothetical stems are not included).

Because some stems overlap, *STAR* incorporates the first and the last stem into the final structure. This structure is now:

```

26 34 70 78
35 39 47 51

```

Now you let *STAR* fold the same sequence, but with the "Potential stems" parameter set to 4. In this case *STAR* finds the following four stems:

```

26 34 70 78   P
30 34 67 71   increasing free energy
30 34 68 72   (ΔG lower)
30 34 69 73   ↓

```

Again, most stems overlap, but now *STAR* incorporates only one stem in the final structure. This structure is:

```

26 34 70 78

```

So the final structure is different from the one in the previous example: one stem is "missing". Furthermore, because one stem that could contribute to the free energy of the structure is not included, the total structure is less stable. This means that the chance that the calculated structure is the same as the one found in nature is smaller. In other words: the reliability is smaller.

The default for the "Potential stems" parameter is 1000.

For sequences of 250 nucleotides or less, 1000 was satisfying in our experience. For sequences upto 500 we suggest at least 2000.

In general we can say that 2, 3 or 4 times as long as 250 nucleotides requires 4, 9 or 16 times 1000 nucleotides in worst cases. Unfortunately, increase of this number also increases computer time and computer memory. So we often used 1000 and still got satisfying results in most cases.

2 Stop after stem

If you are interested in the most important stems, you can limit the calculation. Both "greedy folding", as well as "stochastic folding" will end if the number of predicted stems reach the limit that you specify previously in "Stop after stem".

As the Genetic Algorithm adds and removes stems continuously, this option does not work for that algorithm.

If you enter forced stems and you only want to know their energy values without prediction of new stems, you should set "stop after stem" to 0.

3 *Analyse complete sequence*

STAR will analyse the whole sequence by default. This is indicated by [YES] after "Analyse complete sequence".

If you don't want some regions to basepair, you can exclude regions in two ways.

1. Ending regions can be excluded by specifying other limits after "Analyse". For example in a RNA sequence of 76 nucleotides, changing 1 to 76 into 10 to 60 will exclude regions 1-9 and 61-76 from basepairing.
2. Alternatively you can define regions that should be excluded after "No folding:". For example, specifying in one of the "No folding:" lines the numbers 30 and 40 will exclude the region 30-40 from basepairing.
3. You can also use [Primary/Edit] and change uppercase to lowercase characters. See paragraph 7.6

8.5.3 *[/Factors...]: Greedy Folding Factors*

In most cases, the general factors ("Analyse complete sequence", "Potential stems" and "Stop after stem") are sufficient for this algorithm to work.

If you want to change the default factors for the greedy folding algorithm, you should press button [Change] after "greedy folding". Here you find two options for modification. Both were designed as an initial attempt to simulate influences on the folding pathway direction. They are very experimental and could be removed in a future version.

You see the following dialog box:

FACTORS FOR GREEDY ALGORITHM

Growth _____0
Temperature _____0

OK

CANCEL

As you see *STAR* provides 2 ways to influence the greedy folding algorithm:

1. Growth factor and
2. Temperature folding factor.

1 *Growth factor*

If a RNA molecule is synthesized the 5' terminus is the first one formed. You can imagine that stems that are formed near to 5' terminus are favored above stems that are formed near the 3' terminus.

STAR can deal with this using a bonus for stems near to the 5' terminus so that *STAR* adds them earlier to the structure.

You can define this bonus by entering a value bigger then 0 for the growth factor. The bigger you define this factor, the bigger the bonus.

STAR calculates this bonus B as follows:

$$\mathbf{B} = (\mathbf{GV} / 100) \times \mathbf{N}$$

where

GV=growth value you entered

N = nucleotide number

STAR adds this bonus to the free energy of the basepair of the stems (the fifth column in the definition of a stem; see paragraph 8.7.1). The resulting extra energy is used only for the determination of the next stem to be added.

The energy which is printed after the program is finished does not include this bonus.

2 *Temperature folding factor*

During and after the time a RNA molecule is synthesized, the RNA folds into its native structure. You can imagine that the chance that two nucleotides in a sequence meet each other to form a basepair depends on two factors:

1. the distance (in number of nucleotides) between the two nucleotides. This is the length of the loop formed.
2. a hypothetical factor that describes the flexibility of the RNA molecule. As we see it, this factor depends on the temperature of the RNA.

The first factor, distance, is dealt with in the energy gain/loss calculations. The energy needed to form a hairpin, bulge or internal loop is higher if the loop length is greater.

The second factor, "the temperature folding factor", is theoretical. Default we have set this factor to 0 so as to have no effect.

When you set this factor to a value higher than 0, long distance interactions are less favorable. When you set the factor to a value lower than 0 these interactions are more favorable.

STAR uses this factor during the prediction of a structure by multiplying the energy of the single stranded stretch with MF as follows:

$$\mathbf{MF} = \mathbf{EXP}(\mathbf{TV} / 1000)$$

where...

EXP=exponential

TV=Temperature value that you entered

8.5.4 [/Factors...]: Stochastic Folding Factors

In most cases, the general options ("Analyse complete sequence", "Potential stems" and "Stop after stem") are sufficient for the stochastic algorithm to work.

If you want to change the defaults for the stochastic folding algorithm, you should press button [Change] after "Stochastic folding".

This gives you the following dialog box:

FACTORS FOR STOCHASTIC ALGORITHM

Initial length __ 9999

Increment __ _ 25

Population size __ _ 10

Alternative computations

☐ 1 ☐ 2 ☐ 3

OK

CANCEL

As you see *STAR* provides 4 ways to influence the stochastic folding algorithm:

1. initial length and
2. increment,
3. population size and
4. alternative computation.

The first one determines if you deal with a fullgrown molecule (initial length is 9999) or with folding during synthesis (initial length is small, say 100). Using the latter often yielded better predictions.

1 Initial length

If a RNA molecule is synthesized the 5' terminus is the first one formed. You can imagine that stems that are formed near to 5' terminus are favored above stems that are formed near the 3' terminus. So the RNA starts the folding at the 5' terminus of the molecule.

By default *STAR* sets parameter "initial length" to a very high number. This ensures that *STAR* will immediately use the complete sequence for its algorithm.

A promising variant of the stochastic algorithm is this "growth" variant. This will steer the calculation along a simulated folding pathway by increasing the sequence gradually.

You invoke this "growth" variant by setting the "Initial length" to a small number, (we found numbers less than 100 unsatisfactory, so *STAR* always uses minimum 100)

The stochastic algorithm will select that many nucleotides from the 5' part of the sequence first and calculate its structure.

In subsequent steps *STAR* will increase the (initial) length of the sequence with the amount of nucleotides specified in "Increment". This continues until the sequence is complete.

We think that this growth variant can be more reliable because it mimics more or less the folding pathway.

2 Increment

The increment value is used for the amount that the sequence should increase when using the growth variant.

The value is highly empirical. If you choose a very small value, like 10, computation is very slow because it takes many cycles before the RNA sequence is full grown.

If you choose a very high value, like 100 or 200, the growth will not influence the calculation very much.

You should experiment with this value yourself for each RNA. We found a value of 25 or 50 often quite satisfactory.

Thus, if you specify computation 1, you select the minimal number of stems; if you specify computation 3, you select more stems.

This gives you the following dialog box:

FACTORS FOR GENETIC ALGORITHM

Barrier 100 [X] Growth

Stems taken for mutation
 ○50 ○40 ○30 ○20 ○10

Stop after ____3 ○ nochange
 ○ iterations

Population size ____5

OK

CANCEL

The default value is relatively high (100 kcal/mol), so that essentially any barrier is considered. Our preliminary studies show that most barriers are in the range 10-20 kcal/mol. Decrease of the parameter to lower values can sometimes "spoil" the predictions.

2 *Growth*

If a RNA molecule is synthesized the 5' terminus is the first one formed. You can imagine that stems that are formed near to 5' terminus are favored above stems that are formed near the 3' terminus. So the RNA starts the folding at the 5' terminus of the molecule.

The best results of the Genetics Algorithm were obtained when applying this "growth" variant. This will steer the calculation along a simulated folding pathway by increasing the sequence gradually.

So *STAR* applies the growth variant by default. You can switch it off when you want to simulate folding of a synthesised RNA.

When applying the growth variant, the genetic algorithm will select about 20 nucleotides from the 5' part of the sequence first and calculate its structure.

In subsequent steps *STAR* increases the (initial) length of the sequence with an amount that depends on the improvement of recent iterations. This continues until the sequence is complete.

We think that this growth variant can be more reliable because it mimics more or less the folding pathway.

3 *Stems Taken*

You can choose one of 5 alternative computations. The difference between them is the number of the most stable stems, that are taken in every GA-mutation for possible addition to those that are previously formed. This number is varied between 50 and 10. Variant 50 is the default value; others could be faster, but are sometimes less reliable.

4 *Stop mode*

The end of simulation is determined by two different ways: a "no changes" way or an "iterations" way.

Both ways use the number that you specify right after "Stop after". The meaning of this number is different, depending on the two alternative radio-buttons next to that number.

If you choose [*no change*], the calculation halts if there is no improvement in # consecutive computations (and if the RNA chain is completed, see "growth..."). You specify that number # at the left of this radio-button, in general 3 is sufficient.

If you choose [*Iteration*], the calculation halts after every #-th iteration. You can specify that number # at the left of this radio-button. You can use this option to have the opportunity to stop the program without running it to the end (for example, you have already noticed the folding of the most likely structure, or, the opposite, you see the pathway that contradicts completely with some of your data and you want to change parameters).

Whatever you choose, after reaching the "stop-criterion" *STAR* halts and asks for instruction. You can tell to iterate another number of cycles, or to continue until some iterations do not change. This goes on until you tell *STAR* to stop.

5 *Population Size*

You can change the number of structures that constitute the population in GA procedure.

We found that a population size of about 5 yielded quite satisfactory results. Smaller numbers tend sometimes to decline the quality of results, being, however, much faster. Bigger numbers tend to increase the time of calculations without sufficient improvement.

8.6 [/CALCULATE]

[/Calculate...] is the essential option of *STAR*. It calculates (predicts, folds) a structure for you.

You should have a sequence in your computer's memory before you can use this option.

In the following paragraphs we follow the calculation step by step.

In the first paragraph we tell you what to do before you start the calculation. In the second paragraph we tell you how to start the calculation. In the third paragraph we explain the common preliminaries of all folding algorithms, then we explain each algorithm. Finally in the last paragraph we tell you how *STAR* shows the calculated structure.

In summary:

1. before you start calculation
2. starting the calculation
3. generation of all stems
4. greedy folding
5. stochastic folding
6. genetic algorithm folding
7. show predicted structure

8.6.1 Before You Start Calculation

Normally you choose the algorithm to calculate the structure without any preliminaries.

However, if you are more experienced you have a number of alternatives that you can consider before you start the calculation.

1. *correct sequence*

Before you start the calculation, be sure that you have the correct sequence in the computer. If you don't, get one using the options [Primary/Open...] or [Primary/Edit].

2. *folding factors*

Also, be assured that you are satisfied with the folding factors. See the previous paragraph (paragraph 8.5).

3. *energyrules*

Furthermore, are you satisfied with the default energyrules? If you are not, you can change them by opening an existing energyrules file with option [Energyrules/Open...] or edit the default energyrules with option [Energyrules/Edit...].

4. *forced stems?*

When you are sure that certain stems must exist in the structure, you can force *STAR* to incorporate those stems in the structure. If you want to do that, you simply define those "forced stems" before you start the calculation. We explain this procedure in paragraph 8.8 where we discuss the option [Secondary/Edit].

If you already worked on the same sequence in a previous session, you can load a structure file from disk with the option [Secondary/Open] to force these stems into the structure.

Alternatively, if you forced stems in a previous run, you must delete them if you don't want them now. For deletion use [Secondary/Edit] and delete all stems in the window that appears.

5. *non-pairing*

When you are sure that a certain part of the sequence does not basepair with the rest of the sequence, you can prevent the basepairing of this region. If you want to do that, you should change the nucleotides of that part of the sequence into lowercase a,g,c or u. You can do this in [Secondary/Factors...] in the fields below the text "Analyse complete sequence"..

6. *circular RNA*

If you want to predict circular RNA you should know one reliable hairpin (if you don't, you could try the most stable one using a linear prediction first). Let us say for example 27-32:38-43 in tRNATHRT.

Rearrange the nucleotides in the sequence using [Primary/Edit] such that the new ends fall within the loop. In our example remove nucleotides 1-32 and add them to the right of the last nucleotide (number 76).

Force the hairpin in this new configuration using [Secondary/Edit]. In the example: 6-11:71-76 you have circular RNA you can bind the first nucleotide to the last.

8.6.2 *Starting The Calculation*

You start calculation by activating option [/Calculate...].

You see the following dialog box:

PERFORM ALGORITHM:

- ☐ Greedy folding
- ☐ Stochastic folding
- ☐ Genetic Algorithm

Cancel

Press the required algorithm.

If you already made a calculation before, the box shown below appears. With this box *STAR* warns you that it will overwrite your current structure if you let it calculate a new one.

Predicting a structure now
will overwrite your
current calculated structure

OK

Cancel

When you are sure that the stems of your old structure can be overwritten, press [OK].

Otherwise press [Cancel]. Then you will return to the main menu so you can first save that structure for later use. You can do this by the option [Secondary/Save].

Remember that *STAR* will forced stems (the “forced stems” are those that you edited or opened, not a previously calculated structure) in the new structure (see previous paragraph).

If you don't want that, you should erase the stems of that edited structure before you start the calculation. Use option [Secondary/Edit] and remove all stems in the appearing window (see paragraph 8.7).

If you *do* want to incorporate previously **calculated** stems into the new calculated structure, it suffices to press [Secondary/Edit]. *STAR* inserts the calculated structure in the edit window and all you have to do is exit the editor.

8.6.3 *Preliminary Generation Of All Possible Stems*

As mentioned in paragraph 2.5, *STAR* first generates all possible stems of your sequence. Remember that "all possible stems" also depends on the "Potential stems" parameter; see paragraph 8.5.1.

STAR generates those stems by storing all possible direct stackings of a sequence in a stacking table.

Because the memory of a computer is limited, *STAR* doesn't necessarily create this table in one run. During this process, *STAR* displays the size of the subtable that is formed at a certain moment. For example:

Size of stack-table searched: 90 59 (run 1)
Size of stack-table searched: 32 32 (run 2)

When *STAR* has found all possible stems, it tells you how many stems were found, for example:

229 potential stems found, prediction starts.

8.6.4 Greedy Folding Calculation

After generation of all possible stems the greedy folding algorithm adds the most stable stems one by one to the structure. During this process, *STAR* displays each stem it adds at a certain moment, for example:

```
26 34 70 78 paired, ΔG = -16.9 kcal/mol
      (first stem added)
35 39 47 51 paired, ΔG = -16.3 kcal/mol
      (second stem added).
```

In paragraph 8.7.1 we explain the meaning of these values.

8.6.5 Stochastic Folding Calculation

The stochastic folding algorithm indicates the RNA chain length which is currently analyzed and the number of nucleotides left for subsequent folding in growing chains.

nucleotides are analyzed, # left.

Next the number of stems taken for Monte Carlo generation of random structures is indicated:

stems are considered.

STAR shows how many random structures are generated:

```
Random structure generated: 1
Random structure generated: 2
...
```

After processing these structures, the program shows stems added into the structure which will be the final one and its current free energy

```
5 10 15 20 paired, ΔG = -5.2 kcal/mol
```

If the stochastic procedure did not select any stem, stems can be added (not always) by stability criterion only:

```
stems selected without Monte Carlo
21-25 31-25 paired, ΔG = -8.2 kcal/mol
```

At the end of each Monte Carlo cycle, the program estimates the time needed to complete the folding. This is *very* tentatively; just to give you a crude indication.

8.6.6 Genetic Algorithm Calculation

After the generation of lists of possible stems the genetic algorithm processes them by constructing intermediate structures.

For every iteration, its number and the current RNA chain length are indicated:

Iteration: 1... chain length:...

during “growth” the GA displays the current length of the synthesised sequence. If the energy of the structures improve very rapidly, growth is retarded a bit in order to reach a relatively stable folding.

The program shows also the energies of new structures produced by GA operations:

```
Mitating structure 1: -15.3 kcal/mol
...
Mitating structure...: -7.8 kcal/mol
Crossover:             -15.8 kcal/mol
```

In the end of iteration possible changes in the most stable intermediate structure are shown:

```
Added:      Removed:
3  8  83  88  6  8  15  17
```

Free energy of the best structure and the mean energy of the current population of structures are shown:

```
Best free energy: -15.7
Mean free energy: -12.4 (quality=0.8)
```

STAR compares the mean energy with that of a random sequence. This ratio is used for “quality”. Values below 1 are low-energetic foldings that could unfold later to enable more stable stems, often in a quick “zippering” process. Despite their low “quality” such structures may have a real purpose in the folding process.¹⁵ A low value in the final result may indicate either low prediction quality or the absence of a functional important structure.

Every 5-th iteration also outputs all stems of the best structure:

Best structure:

```
3  8 84 88
20 24 30 34
. . . . .
```

At the end of each cycle, the program estimates the time needed to complete the folding. This is *very* tentatively; just to give you a crude indication.

There are two alternatives to stop the GA-calculations. You choose in [Secondary/Factors] “GeneticAlgorithm[Change]”.

If you selected “**o**nchange”, the program will end after some number of iterations # that did not improve the structure. This number # is specified in front of “**o**nchange”.

If you choose “**o**iterations”, the program will pause after the selected number of iterations. . This number # is specified in front of “**o**nchange”.

However you wanted to stop (after a fixed number of iterations or after some iterations without change), when the stop-criterion is reached the program halts and asks for instructions. It presents the following box:

Stop criterium reached

Stop after __3__ **o** nochange
o iterations

OK

Stop now

- if you choose [iterations] [OK] STAR will continue for another 3 (you can change this number too) iterations.
- if you press [nochange] [OK] STAR will continue until another 3 (you can change this number too) iterations in a row do not improve.
- if you are satisfied with the latest prediction press [Stop now].

¹⁵ Gulyaev, A.P. & Batenburg, F.H.D. van & Pleij, C.W.A. (1995): The computer simulation of RNA folding pathways using a genetic algorithm. J.Mol. Biol. 250, 37-51.

8.6.7 Final Show Of Predicted Structure

When *STAR* has processed all stems, the calculation process is finished. Now *STAR* shows the calculated structure in the following way:

```

26 34 70 78 -12.0 9.0 -3.0 * GAGGUCCCC
                        * UUUUGGGGG
35 39 47 51 -11.8 5.1 -5.7 * GGCAG
                        * CCGUC
11 14 20 23 -5.5 4.9 -0.6 * CCUU
                        * GGAA
  5  7 15 17 -6.5 4.2 -3.0 * GCC
                        * UGG
44 46 54 56 -3.6 5.5 -0.2 * AAU
                        * UUA
57 59 67 69 -2.8 4.099 2.1 * UUC
                        * GGG

```

5

We explain the meaning of the values in this table in paragraph 8.7.1

8.7 [/VIEW...]

The [/View...] option enables you to examine the structure that is in memory. Therefore, you can only activate this option if there is a sequence as well as a structure present in the computer.

After activating this option the following box appears. With this box you select how you want to view the structure.

Choose display:

Table	Span
Lego	Mountain
Bracket	ConnecT
Cancel	

There are 6 ways to view a structure: 1) by table, 2) by span-view, 3) by lego-view, 4) by mountain-view, 5) by bracket-view and 6) by connect-view.

N.B. Only table- span- bracket- and connect-view show pseudo-knots correctly, lego- and mountain-view do not!

Press buttons [Table], [Span], [Lego] [Mountain], [Bracket] or [Connect] to select how you want to view the structure. If you press [CANCEL] you return to the main menu.

The following paragraphs illustrate these outputs with the example of the sequence from file STAR.1.

8.7.1 View Structure As Table

When you view a structure as [Table], the following table appears on screen:

26	34	70	78	-12.0	9.0	-3.0	*	GAGGUCCCC	
							*	UUUUGGGGG	
35	39	47	51	-11.8	5.1	-5.7	*	GGCAG	
							*	CCGUC	
11	14	20	23	-5.5	4.9	-0.6	*	CCUU	
							*	GGAA	
5	7	15	17	-6.5	4.2	-3.0	*	GCC	
							*	UGG	
44	46	54	56	-3.6	5.5	-0.2	*	AAU	
							*	UUA	5
57	59	67	69	-2.8	4.1	2.1	*	UUC	
							*	GGG	

In this table, the first four columns specify the location of a stem:

1. Column 1 contains the base number of the most 5' nucleotide of the stem.
2. Column 2 contains the base number of the 3' end of the 5' stem-half.
3. Column 3 contains the base number of the 5' end of the 3' stem-half.
4. Column 4 contains the base number of the most 3' nucleotide of the stem.

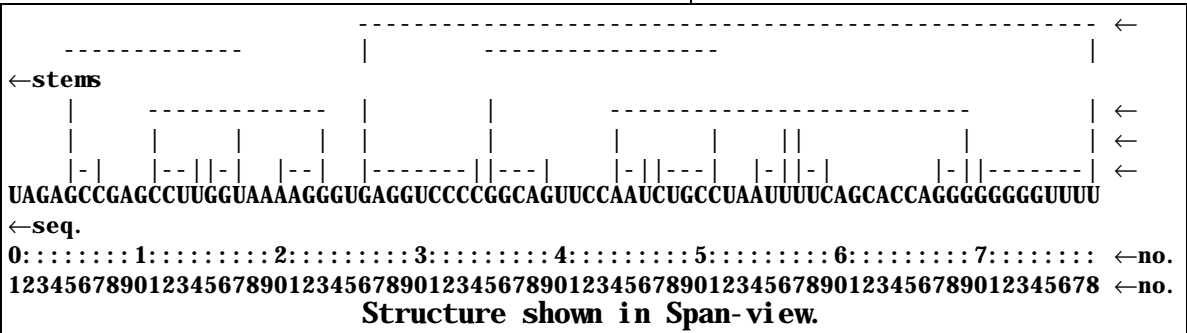
The next two columns show the energy content of the stem:

5. Column 5 contains the Gibbs free energy of the base pairs in kcal/mol.
6. Column 6 contains the Gibbs free energy of the single stranded stretch of RNA located directly between the corresponding bases of the second and third column in kcal/mol.
7. Column 7 contains the Gibbs free energy that the stem contributes to the structure of previous stems. Each value is the sum of col 5 and 6 unless there is a loop outside the stem. The sum of column 7 yield the total energy of the whole structure.

At the right side of the table you see the basepairings of the corresponding stem. Rightmost of the table you see a "5". This indicates the fifth (tenth, fifteenth, etc.) stem of the structure.

8.7.2 View Structure By Spannings

When you view a structure by [Span], (after a while) the following picture appears on screen:



In the middle row of this picture, you see the nucleotides. And the two bottom lines are the nucleotides numbers: in the bottom line the units and one line above the tens. The nucleotides are numbered starting with 1 at the 5' end and increase until the 3' end is reached.

Above the sequence you see the stems drawn. *STAR* draws a stem by connecting two single-stranded stemhalves.

STAR draws a stemhalf as:

| - - - |

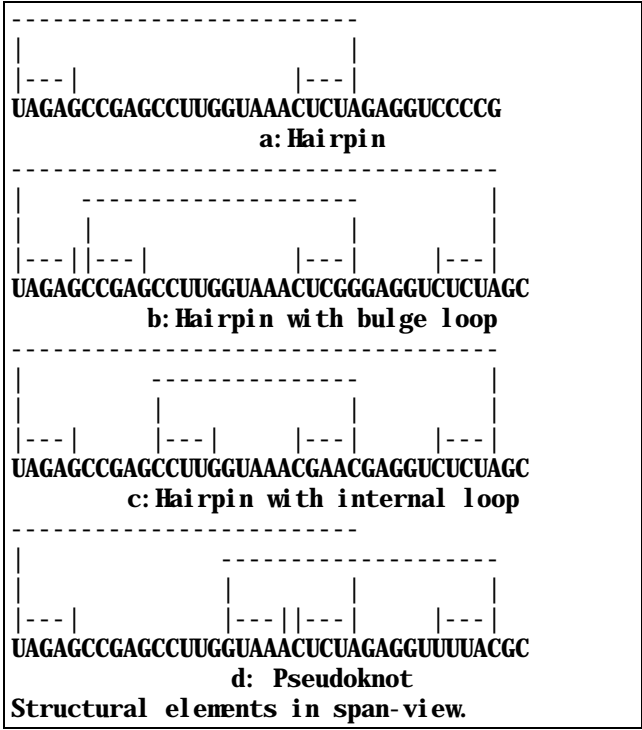
STAR draws a stem as:

|
| - - - | | - - - |
|

This presentation is more visual than the view as [Table]. Just imagine that you "pull the strings together". The disadvantage of this "graphic view" is that *STAR* takes more time to calculate it. Furthermore it gives less information than the table as the energy values are not included.

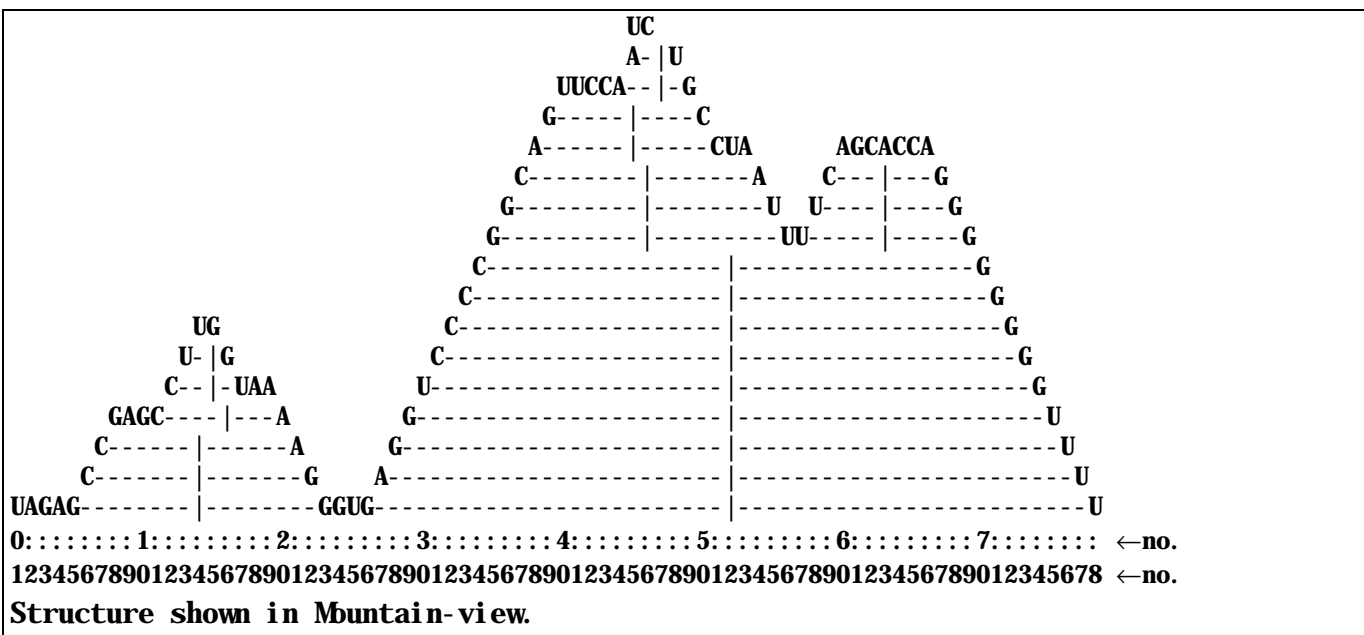
Notice that stem 5-7:15-17 forms a pseudoknot with stem 11-14:20-23. Another pseudoknot is formed by stem 35-39:47-51 and stem 44-46:54-56.

In the next figure we show how various structural elements look like in span-view.



8.7.3 View Structure As Mountain

When you show a structure by mountain-view,
you see the picture shown.



Basically this is the representation as proposed
by Hogeweg & Hesper¹⁶. Pairing nucleotides are
connected by horizontal lines; the centre is
indicated by a "|".

***N.B. Pseudoknots will display incorrectly in
mountain-view.*** You see this problem for
pseudo-knot 5-7:15-17 with 11-14:20-23. The
mountain view incorrectly suggests three stems
instead: 5-7:21-23 and stem 11-11:20-20 and
stem 12-13:16-17!

¹⁶ Hogeweg, P. & Hesper, B. (1984): Energy directed folding of
RNA sequences. Nucleic Acids Research 12, p.67-74.

8.7.4 View Structure As Lego

When STAR shows a structure as Lego-view, you see the following picture.

```

          UC
          AU
        UUCCAG
        G--|-C
        A--|-CUAAGCACCA
        C---|--AC---|--G
        G---|--UU---|--G
        G---|--UU---|--G
        C-----|-----G
        C-----|-----G
        C-----|-----G
        UG      C-----|-----G
        UG      C-----|-----G
        CUAA    U-----|-----G
        GAGC-|A   G-----|-----U
        C--|--A   G-----|-----U
        C--|--G   A-----|-----U
        UAGAG--|--GGUG-----|-----U
0: : : : : 1: : : : 2: : : 5: : : 6: : : 7: : : :
12345901589345601234727123456

```

This lego-view was proposed by F.H.D.van Batenburg. It is a compression of the mountain-view in order to reduce the mountainous size of that view. Nevertheless it presents the same information as the classical mountain-view.

N.B. Pseudoknots will display incorrectly in lego-view. You see this problem for pseudoknot 5-7:15-17 with 11-14:20-23. The mountain view incorrectly suggests three stems instead: 5-7:21-23 and stem 11-11:20-20 and stem 12-13:16-17!

After some use, you can very easily recognise the structural components. In the next figure we show how various structural elements look like in lego-view.

```

CCGAGCCUUGGUA
G-----|-----C
A-----|-----U
G-----|-----C
A-----|-----U
U-----|-----AGAGGUCCCCGGCAGUCCAAUC
                                     a: Hairpin
CCUUGGUA
G-----|-----C
A-----|-----U
G-----|-----C
C-----|-----G
C-----|-----GGAGGU
G-----|-----C
A-----|-----U
G-----|-----C
A-----|-----U
U-----|-----AGCAGUCCAAUCUGCCUAAUUU
                                     b: Hairpin with bulge loop
      GUAAA
      G-|-C
      U-|-A
      U-|-A
      C-|-G
CCGAGC-|-GGAGGU
G-----|-----C
A-----|-----U
G-----|-----C
A-----|-----U
U-----|-----AGCAGUCCAAUCUGCCUAAUUU
                                     c: Hairpin with internal loop
Structural components in Lego-view.

```

8.7.6 View Structure As Connect

When *STAR* shows a structure as *Connect* view, you see the following output.

1	U	0	2	0	1
2	A	1	3	0	2
3	G	2	4	0	3
4	A	3	5	0	4
5	G	4	6	17	5
6	C	5	7	16	6
7	C	6	8	15	7
8	G	7	9	0	8
9	A	8	10	0	9
10	G	9	11	0	10
.....					
71	G	70	72	33	71
72	G	71	73	32	72
73	G	72	74	31	73
74	G	73	75	30	74
75	U	74	76	29	75
76	U	75	77	28	76
77	U	76	78	27	77
78	U	77	0	26	78

Notice that pseudoknots are very obvious by the use of brackets instead of parentheses. For example, in the first pseudoknot brackets are used to indicate basepairing 5-7 : 15-17 and parenthesis for basepairing of 11-14 : 20-23.

This view is not particularly useful for personal inspection but more intended for migration of data to other programs.

8.8 [/EDIT]

The [/Edit] option enables you to force *STAR* to incorporate certain stems into the predicted structure (see paragraph 8.6.1).

You can activate this option only if there is a sequence present in the computer. A structure may or may not be present. If it is, you get the opportunity to change the present structure. If it is not, you can enter a new structure.

You can also edit a structure with a regular text editor. Appendix D describes the format. After creation, you can import such a file using [Secondary/Open] (see paragraph 8.2).

You enter the *STAR* editor by activating the option [/Edit] from the [Secondary] menu.

When you enter the editor a new window appears. This window can be moved and resized, just like you usually do.

Usually there is no previous structure. Then you only see a blank screen when you enter the editor.

If there is already structure when you enter the editor, you see something like:

```
(0) RNASTRUCTURE
(1) 26 34 70 78
(2) 35 39 47 51
(3) 11 14 20 23
(4) 5 7 15 17
(5) 44 46 54 56
(6) 57 59 67 69
```

The four columns are:

1. Column 1 contains the base number of the most 5' nucleotide of the stem.
2. Column 2 contains the base number of the 3' end of the 5' stem-half.
3. Column 3 contains the base number of the 5' end of the 3' stem-half.
4. Column 4 contains the base number of the most 3' nucleotide of the stem.

The following paragraphs will describe how to:

1. Add stems.
2. Type comments
3. Delete stems.
4. Leave the editor
5. Possible errors.

8.8.1 Add A Stem

You can add a stem by simply typing the 4 numbers (in increasing order) that describe the stem.

For example:

```
26 34 70 78
44 46 54 56
5 7 12 14
```

8.8.2 Type comment

Similar to editing sequences you can add comment wherever you like by preceding it with an asterisk. For example:

```
* stem 5' part | 3' part
* start | end | start | end
   26   34   70   78 * first stem
44      46      54      56 *second stem
```

However, unlike comments in sequences, these comments are not stored.

8.8.3 Delete Stems

If you want to delete one or a few stems there are two ways to do so.

- If you press the [Delete]-key, *STAR* deletes (erases) the number at the cursor position.
- Another way is to press the [Backspace]-key. *STAR* deletes the number just left of the cursor.

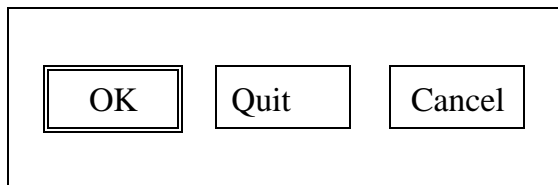
If you want to delete many stem-numbers the following procedure is faster.

- First you define a block of stems by dragging the mouse over the structure. After this you erase the selected block with [Co+X].

8.8.4 Leave The Editor

Leave the editor by clicking in the left top corner of your edit window.

You see the following dialog box¹⁷.



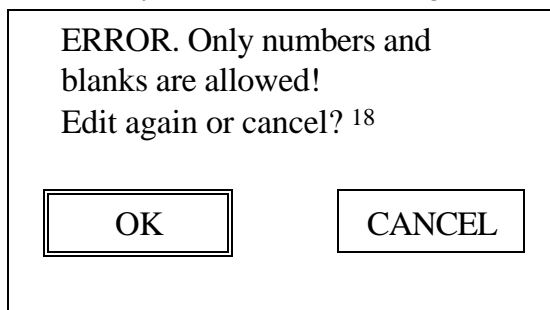
Choose [OK] if you are satisfied with your editing. Press [Quit] if you want to discard your editing work. Choose [Cancel] if you want to resume editing.

The status indicator menu shows "___/E". The "/E" shows you that you have a sequence in the memory of your computer, although you did not [/Open...] a sequence file, but entered it with the editor.

1 Possible Errors

1. WRONG CHARACTERS:

If you want to leave the editor, and there are other characters in your structure (note: to the right of the asterisks you can type anything you want) then numbers and blanks, *STAR* warns you with the following box:



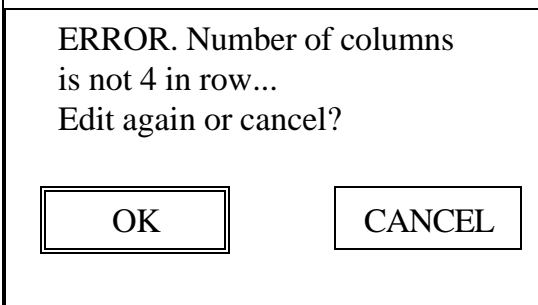
Just press the [OK]-button, and you will return to the editor-screen, where you can correct the wrong characters.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old

structure from before you started editing will remain in the computer.

2. WRONG FORMAT:

It is also possible that one of your structure-stems does not consist of four numbers. Now the following box will appear:

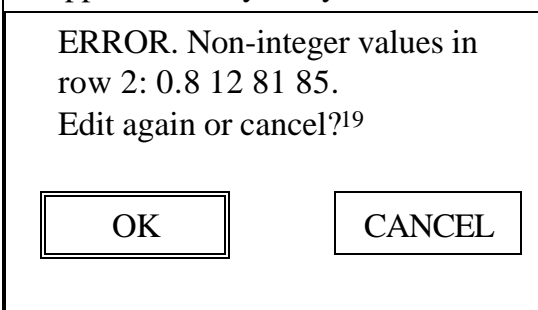


Just press the [OK]-button, and you will return to the editor- screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the last edit-session. The old structure from before you started editing will remain in the computer.

3. NON-INTEGGER VALUES:

If you used points in your stem-numbers to make non-integer values, the following box appears when you try to leave the editor:



¹⁷ PC version shows the [Exit] or [/Quit] buttons underneath the edit window.

¹⁸ Points are only allowed when they are used in numbers, not when they are used on their own.

¹⁹ Example of a wrong stem

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws all your changes from the edit-session away. The old structure from before you started editing will remain in the computer.

4. OVERLAPPING STEMS:

When several stems of your structure overlap each other, the following box appears:

ERROR. Stem not allowed:

24 26 66 68,

(overlap with other stem).

Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor- screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old structure from before you started editing will remain in the computer.

5. UNEQUAL LENGTH OF STEM-HALVES:

When the two stem-halves of one stem are not of the same length, *STAR* warns you with the following box after leaving the editor:

ERROR. Stem not allowed:

13 14 20 25,

(unequal length stemhalves).

Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor- screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old structure from before you started editing will remain in the computer.

6. STEMNUMBERS NOT IN SUCCEEDING ORDER:

When the stemnumbers of the first column of the structure-table are not in succeeding order, *STAR* warns you with the following box after leaving the editor:

Stemnumbers not in
succeeding order!

Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editorscreen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old structure from before you started editing will remain in the computer.

8.9 [/PRINT...]

You can use the [/Print...] option to make a hard-copy (a print) of your calculated structure.

Alternatively you use this option to save the structure to file in a *STAR* layout of your choice; this is the layout that you see using [/View] ²⁰. You might do this to export the structure to a textprocessor for publication, or for export to a drawing program like LoopDloop. Obviously you can only activate this option if you have a structure in your computer's memory.

When you activate this option the following box appears:

CHOOSE DISPLAY

Table	Span
Lego	Mountain
Bracket	Connect
<div style="border: 1px solid black; padding: 5px; display: inline-block;">Cancel</div>	

In this box, you choose what view of the structure you want to print. These views are the same as described by the [/View...] option; see paragraph 8.7.

After you have chosen the view the following box appears. With this box you can specify where you want to print to, and enter the size of your printer paper.

PRINT OPTIONS

Page length ___ lines
 Page width ___ columns
 Top margin ___ lines
 Bottom margin ___ lines

Print to: ☐File ☐Printer

OK

CANCEL

In the four fields marked by ____, you enter the length and width of the paper, and the top and bottom margin. *STAR* already provided default values that are fine in most cases.

The top margin is the part above the text on a page. It contains the header. The bottom margin is the blank part below the text on a page.

For instance, if you use paper 71 characters long, enter "71" at the first field.

For a standard A4 paper format, you enter "66" for the pagelength, and "80" for the page width. For standard continuous forms paper, enter "71" for the pagelength and "80" for the pagewidth. We advise for both top and bottom margin a length of 2 lines.

You can specify whether to print to a printer or to a file.

Printing *to a printer* means that *STAR* outputs directly to a printer.

8.9.1 Print to file

If you have any problem with printing, you can print to file. *STAR* asks you to enter a filename using the file selector. You can print this file separately after leaving *STAR*.

Printing *to file* also allows you to store a view of the structure for later use.

²⁰ For standard lay-out to read back in *STAR* you should use [Secondary/Save].

Another application for file output might be the drawing package LoopDloop²¹. For LoopDloop you should use [Bracket]-view for output. Be aware that currently LoopDloop does *not* handle pseudoknots. If you have pseudoknots, you should first remove one of the two stems of that pseudoknot.

If you want to include the *STAR* output in a publication you can also use "Print to File". The file can be read in any text processor. In this case you should set top and bottom margin to 0 lines. Use a non-proportional font in your textprocessor to display the structure properly.

1 Possible errors

1. NO PRINTER ATTACHED:

If you specify you want to print to a printer while there is no printer attached to your computer, you can not use *STAR* for about 30 seconds.

2. PRINTER BUSY:

This message can happen in network printers. *STAR* detects a busy signal and will not print. Unfortunately this is a network problem, not a *STAR* bug. Your best solution is to use option "Print to: File" and print the file with another program (textprocessor).

²¹ Author Don Gilbert, Biocomputing office biology dept. of Indiana University, Bloomington, IN 47405, USA. Available on FTP: ftp.bio.indiana.edu as /molbio/loopdloop/loopdloop.hqx.

9. THE [ENERGYRULES] MENU

Desk	Files	Primary	Secondary	Energyrules
				Help

				Open...
				Save as...
				Save

				View
				Edit

				Print...

In this chapter, we explain all options of the [Energyrules] menu in detail. Usually you will not need this feature. *STAR* uses the most recent energy values from literature. However, you can supply your own values if you don't agree with our choices. For example if you want to fold at another temperature as 37° C., or if you want to experiment, or if research suggests other values. In particular for pseudoknots our choices only allow a restricted class of pseudoknots. For research and experimentation with other pseudoknots you can change those energy values yourself.

9.1 [/HELP]

The [Help] option displays brief information about the options in this menu.

The main purpose of this option is to give you some help at your fingertips, without having to consult the manual. Now don't throw away this manual as this option is only a brief reminder.

When you close this informative window you return to the main menu.

9.2 [/OPEN...]

Every time you start *STAR* the internal energyrules will be activated.

The [/Open...] option enables you to open a file with energyrules from disk to replace those internal default values.

The set of energyrules of *STAR* consists of 11 types of energyrules (stacking, internal loop, bulge, hairpin, deep groove, shallow groove, quasi-junction, mismatches, 2 types of dangling ends and some constants). See paragraph 9.5 [/View] for a description. Each of these types can be opened separately.

STAR has already a complete set of energyrules as a default in its memory. When you open one particular type of energyrules, *STAR* replaces only this type of energyrules, the other types remain unchanged.

When you activate this option, the following box appears:

Opening set of energyrules
will erase that set in memory!
Continue?

OK

CANCEL

With this box *STAR* asks you to confirm the opening of an energyrules file.

If you press [OK], *STAR* will read new values from disk, and erase the energyrules in memory. This means that you lose that type of energyrules in the computer during that particular session! Don't panic, the standard energyrules are always reactivated when you start *STAR* again.

So, if you enter new values, please be sure that you *save* those energyrules using the options [/Save] or [/Save as...].

If you are not sure press [CANCEL]. Then *STAR* erases nothing and you return to the main menu.

After you pressed [OK] a file-selector appears by which you can select the energyrules file you want to open.

This is your last chance to back off by pressing [Cancel]; if you press [OK] your data are replaced.

Because these energyrules are very specific for *STAR*, you can not open just any file. You can only open files with specific extensions, like:

BUL or bul for bulge rules

for example RULES.BUL

DEE or dee for deep groove rules

for example ENGY.DEE

FOR or for for formula constants

for example FREIER.FOR

HAI or hai for hairp inloop rules

for example GIBBS.HAI

INT or int for internal loop rules

for example STOCKMAY.INT

MIS or mis for mismatches

for example RULS.MIS

NO3 or no3 for nonpaired 3' terminals
for example CHANGED.NO3

NO5 or no5 for nonpaired 5' terminals
for example TEMP.NO5

QJU or qju for quasi-junction rules
for example TRYOUT.QJU

SHA or sha for shallow groove rules
for example ENERGY.SHA

STA or sta for stacking energyrules
for example JACOBSON.STA

1 Possible errors:

1. WS-FULL:

A very common error is when you want to open an energyrules file that is too big for the memory of your computer. This results in a "INSUFFICIENT MEMORY" error.

Unfortunately you can do nothing else then start *STAR* all over again on a computer with sufficient memory.

2. WRONG FILENAME:

When you try to open a file with the wrong extension (Eg. when you want to open a stackingrulesfile with the name STACK.TXT), the following error-message appears:

ERROR. Only²²
of name²³ can be opened

OK

²² The dots stand for the type of energyrules you selected, like: "stackingrules", "internallooprules", "bulgerules", "hairpinlooprules", "deepgrooverules", "shallowgrooverules", "quasi-junctionrules", "mismatchrules" "nonpairedterminals" or "formulaconstants".

²³ The dots stand for the corresponding filename with extensions like: "*.sta", "*.int", "*.bul", "*.hai", "*.dee", "*.sha", "*.qju", "mis", "for", "no3", or "no5".

All you can do is press the [OK]-button after which you return to the main menu. The file is not opened. Now you can try to open another file, or rename the file. When you want to rename the wrong file, be sure it is indeed the correct file with just the wrong name, and not also with the wrong contents.

3. WRONG FORMAT:

The format of all the energyrules must obey some format-rules. They are listed here:

1. *Bulge loop energyrules:*

The table must always consist of 2 columns.

2. *Deep groove energyrules:*

You can make the table as large as you want, but all the columns and rows must be completely filled.

3. *Shallow groove energyrules:*

You can make the table as large as you want, but all the columns and rows must be completely filled.

4. *Internal loop energyrules:*

The table must always consist of 5 columns.

5. *Hairpin loop energyrules:*

The table must always consist of 5 double columns. (this means $5 \times 2 = 10$). The first value of each value-couple is the length of the hairpinloop, and these values must be in increasing order.

6. *Quasi-junction energyrules:*

The table must always consist of 2 columns.

7. *Nonpaired terminal nucleotides:*

These tables must always consist of 16 lines with 4 columns.

8. *Formula constants:*

This table requires 3 lines. The first and second line specifies 3 numbers, the last line one number.

9. *Stacking energyrules:*

The table must always consist of 16 rows and 16 columns.

10. *Mismatches:*

The table must always consist of 64 lines each with 4 columns.

If the format of your file with energyrules-table does not conform to the rules as described above, the following box appears:

ERROR. Illegal format.
File not opened!

OK

All you can do is press the [OK]-button, after which you return to the main menu. The file is not opened.

4. DOUBLE, NEGATIVE OR NON-INTEGGER VALUES IN COLUMN 1, OR VALUES NOT IN INCREASING ORDER:

If there are double values, negative values or non-integer values in column 1 of your energyrules-table you just tried to open, or if the values in the first column of this table are not in increasing order, the following box appears:

ERROR. Illegal values in column
1. File not opened!

OK

All you can do is press the [OK]-button, after which you return to the main menu. The file is not opened.

5. DOUBLE, NEGATIVE OR NON-INTEGERS IN ROW 1, OR VALUES NOT IN INCREASING ORDER:

If there are double values, negative values or non-integer values in row 1 of your energyrules-table you want to open, or if the values in the first row of this table are not in increasing order, the following box appears:

ERROR. Illegal values in row 1.
File not opened!

OK

All you can do is press the [OK]-button, after which you return to the main menu. The file is not opened.

6. DOUBLE, NEGATIVE, NON-INTEGERS OR NOT SUCCEEDING VALUES IN COLUMN 1 OF THE HAIRPINLOOP ENERGYRULES:

When the lengths of the hairpinloops contain a negative, non-integer or double value, the box shown below, appears. This box appears also when the values of the lengths of the hairpinloops are not in increasing order.

ERROR. Illegal values used in
file. File not opened!

OK

All you can do is press the [OK]-button, after which you return to the main menu. The file is not opened.

9.3 [/SAVE AS...]

The [/Save as...] option enables you to save your energyrules for later use. This means that *STAR* transfers a copy from the computer's memory to disk. You must enter the name of the file where you want to save the energyrules to.

The set of energyrules of *STAR* consists of 11 types of energyrules. Each of these types can be saved separately.

When you activate this option, the following box appears:

CHOOSE ENERGYRULES

Quasijunction

Hairpin loop

Bulge

Internal loop

Formula const.

Mismatches

Deep Groove

Stacking

Shallow Groove

No 3' pairing

No 5' pairing

Cancel

Use the buttons to select the type of energyrules you want to save. When you press [CANCEL] nothing will happen and you return to the main menu.

When you select a type of energyrules and press [OK] a file-selector appears by which you can enter the filename under which *STAR* should save the energyrules.

Because these energyrules are very specific for *STAR*, you can not save a file with just any file-name. You can only save to filenames with specific extensions like:

BUL or bul for bulgerules
for example RULES.BUL

DEE or dee for deep grooverules
for example ENGY.DEE

FOR or for for formula constants
for example FREIER.FOR

HAI or hai for hairpinlooprules
for example GIBBS.HAI

INT or int for intlooprules
for example STOCKMAY.INT

MIS or mis for mismatches
for example RULS.MIS

NO3 or no3 for nonpaired 3' terminals
for example CHANGED.NO3

NO5 or no5 for nonpaired 5' terminals
for example TEMP.NO5

QJU or qju for quasi-junction rules
for example TRYOUT.QJU

SHA or sha for shallow grooverules; for
example ENERGY.SHA

STA or sta for energyrules
for example JACOBSON.STA

If you enter a filename that already exists, the following box appears. With this box *STAR* asks you if you want to overwrite the file.

Replace existing "filename"?

Cancel

Replace

If you press[/Replace], *STAR* saves the energyrules to disk. It overwrites the energyrules on disk.

Otherwise press [Cancel]. Then *STAR* overwrites nothing and you return to the main menu.

1 Possible errors

WRONG FILENAME:

When you try to save the selected type of energyrules with a wrong filename (Eg. when you try to save the bulge loop energyrules with the name BULGE.TXT) the following error-message appears:

ERROR. Use ...²⁴ as extension

File²⁵ not saved!

OK

All you can do is press [OK], after which you return to the main menu.

²⁴ The dots stand for the corresponding filename with extensions like: "*.sta", "*.int", "*.bul", "*.hai", "*.dee", "*.sha", "*.qju", "mis", "for", "no3", or "no5".

²⁵ The dots stand for the wrong file-name you tried to save to.

9.4 [/SAVE]

The [/Save] option enables you to save your current set of energyrules for later use. This means that *STAR* transfers a copy of the set in the computer's memory to disk.

The set of energyrules of *STAR* consists of 11 types of energyrules. Each of these types can be saved separately.

The difference with the option [/Save as...] is that *STAR* saves the energyrules to a file with the same name as the one that you opened or saved to earlier with respectively the options [/Open...] or [/Save as...]. That is why you must have opened an energyrules file (option [/Open...]) before you can activate this option, or saved the energyrules earlier using the option [/Save as...].

Furthermore, *STAR* replaces the energyrules on disk by the energyrules that you saved without warning, so be careful with this option: be sure that the energyrules on disk are no longer useful before you use this option.

When you activate this option a box appears that looks like:

CHOOSE ENERGYRULES	
Quasijunction	Hairpin loop
Bulge	Internal loop
Formula const.	Mismatches
Deep Groove	Stacking
Shallow Groove	No 3' pairing
	No 5' pairing
Cancel	

Select the type of energyrules that you want to save.

9.5 [/VIEW...]

You can use the [/View...] option to view the energyrules that are in the computer's memory.

Because there are several energyrules, you must first select the type of energyrules you want to view. You do that using the following box that will appear immediately after you activate this option.

CHOOSE ENERGYRULES	
Quasijunction	Hairpin loop
Bulge	Internal loop
Formula const.	Mismatches
Deep Groove	Stacking
Shallow Groove	No 3' pairing
	No 5' pairing
Cancel	

Use the buttons to select the required type of energyrules.

When you press [CANCEL], nothing will happen and you will return to the main menu.

As you see there are 11 types of energyrules. We explain them now.

1. **Bulge:** the energy values for bulge loops.
2. **Deep groove:** one of the energy values for p-knots. See "shallow groove" in the Glossary for detailed explanation.
3. **Formula:** constants for multibranch loops, for asymmetric internal loops, and for tetraloops.
4. **Hairpin:** the energy values for hairpin loops.
5. **Internal loop:** the energy values for internal loops.
6. **Mismatches:** the free-energy increments for terminal mismatches and basepairs.
7. **Nonpaired terminals:** the free-energy increments for nonpaired terminal nucleotides. One table for 3' dangling ends and another for 5' dangling ends.
8. **Quasi-junction:** the energy values for a-junction in a pseudoknot. See the Glossary for a detailed explanation
9. **Shallow groove:** one of the energy values for p-knots. See "shallow groove" in the Glossary for detailed explanation.
10. **Stacking:** the energy values for all possible stackings of basepairs.

In the following paragraphs we discuss the display of each type of energyrules.

9.5.1 [/View...]: Bulge Loop

When you view the energyrules of a bulge loop, you see a two-column table, like:

BULGERULES²⁶

0	40
1	30
2	46 <===
3	55
4	62
6	71

1. In the first column specifies the length of the bulge loops.
2. The second column specifies the Gibbs Free Energy, in hcal/mol (*not kcal/mol!*) of the bulge loops.

For example, if you want to know the energy of a bulge loop 2 nucleotides long, you look in the second column in the row which starts with 2.

In this example the energy value is 46 hcal/mol.

When a value is not defined, *STAR* uses the most nearby lower value. In this example the energy of a bulge loop length of 5 nucleotides is 62 hcal/mol.

9.5.2 [/View...]: Deep Groove

When you view the energyrules of a deep groove, you see a table like:

DEEPGROOVERULES²⁷

0	1	2	15	16
0 +	---	---	---	---
2	999	999	999	999
3	999	42	42	999
4	42	42	42	999 <===
8	42	42	42	999
9	42	42	42	999
10	999	999	999	999

The top row specifies the number of nucleotides that cross the deep groove (L1 in next figure). The left column specifies the number of base pairs to be crossed (S2 in next figure).

For example you want to know the energy value of a connecting loop of 2 nucleotides crossing the deep groove over 4 base pairs. The energy value is then 42 hcal/mol.

²⁶ As we continuously update STAR, you might see different values in your program.

²⁷ As we continuously update STAR, you might see different values in your program.

9.5.3 [/View...]: Formula Constants

When you view the energyrules of formula constants, you see a table like:

FORMULA		28
46 2 1 *	Multibranched loops	
30 5 4 *	Asymmetric internal loops	
-20	* Tetraloops	

The first line specifies the formula to calculate the destabilizing energy for a multibranched loop. The reported values represent three coefficients (a, b, c) in the linear approximation of energy:

$$\Delta G = a + b \cdot n + c \cdot h$$

where n is the number of nucleotides in the loop, and h is the number of helices that form the loop. The table values are multiplied by 10.

The second line specifies the formula for penalties in asymmetric internal loops. The penalty in a loop with branches m and n is approximated by formula

$$\Delta G = \min(a, 3n - m^3 \times (b - \min(c, m, n)))$$

where “min” defines a minimum and a, b and c are values from the table (multiplied by 10).

The third line specifies the additional stability assumed to stabilize specific tetraloop sequences. Table value are multiplied by 10.

9.5.4 [/View...]: Hairpin Loop

When you view the energyrules of a hairpin loop, you see a ten-column table like:

HAIRPINLOOPRULES²⁹									
1	999	:	3	70	:	4	45	:	5 41 : 6 43 <==
=	7	47	:	8	50	:	9	53	: 10 55 : 11 57
12	59	:	13	61	:	14	63	:	15 64 : 16 66
17	67	:	18	68	:	19	70	:	20 71 : 21 72
22	73	:	23	74	:	24	75	:	25 76 : 26 77
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:

The table consists of five double columns. Two values between two colons form a data-pair.

1. The first value of each pair indicates the length of the hairpin-loop.
2. The second specifies the Gibbs Free Energy in hcal/mol (*not kcal/mol!*) of the hairpin-loop.

So the odd columns present the length of the hairpin-loops (= the number of free nucleotides in the string). The even columns represent the Gibbs Free Energies.

For example if you want to know the energy of hairpin loop of length 4, you look after “:4”. In this case it is the third pair, and the corresponding Gibbs Free Energy value is 45 hcal/mol.

When a value is not defined, *STAR* uses the most nearby lower value. In this example the energy of a hairpin loop with a length of 1, or 2 nucleotides is 999 hcal/mol.

²⁸ As we continuously update *STAR*, you might see different values in your program.

²⁹ As we continuously update *STAR*, you might see different values in your program.

9.5.5 [/View...]: Internal Loop

When you view the energyrules of internal loops, you get a two-column table like:

INTLOOPRULES³⁰

2	1
3	10
4	16 <====
5	22
7	29

1. The first column gives the length of the internal loop.
2. The second column gives the Gibbs Free Energy in hcal/mol (*not kcal/mol!*).

For example, if you want to know the energy value for an internal loop of length 4, you must look in the second column in the row which starts with 4. In this example the value is 16 hcal/mol.

When the values for a certain length are missing, *STAR* uses the most nearby lower value. In this example the value for an internal loop 6 nucleotides long is 22.

9.5.6 [/View...]: Mismatches

When you view the energyrules for mismatches, you see a table like:

MSMATCHRULES

X	Y: A	C	G	U

AA. yA	0	0	0	0
AC. yA	0	0	0	0
AG. yA	0	0	0	0
AU. yA	0	0	0	0
*-----				
AA. yC	0	0	0	0
AC. yC	0	0	0	0
AG. yC	0	0	0	0
AU. yC	0	0	0	0
*-----				
AA. yG	0	0	0	0
AC. yG	0	0	0	0
AG. yG	0	0	0	0
AU. yG	0	0	0	0
*-----				
AA. yU	-8	-10	-10	-9
AC. yU	-7	-7	- 21	-7 <==
AG. yU	-8	-17	-10	-9
AU. yU	-9	-8	-9	-8
*-----				
CA. yA	0	0	0	0
CC. yA	0	0	0	0
CG. yA	0	0	0	0
CU. yA	0	0	0	0
*-----				
CA. yC	0	0	0	0
CC. yC	0	0	0	0
CG. yC	0	0	0	0
CU. yC	0	0	0	0
*-----				
CA. yG	-19	-20	-19	-18
CC. yG	-10	-11	-29	-8
CG. yG	-19	-20	-19	-16
CU. yG	-17	-15	-19	-12
*-----				
CA. yU	0	0	0	0
CC. yU	0	0	0	0
CG. yU	0	0	0	0
CU. yU	0	0	0	0
*-----				
GA. yA	0	0	0	0
GC. yA	0	0	0	0
GG. yA	0	0	0	0
GU. yA	0	0	0	0
*-----				
GA. yC	-11	-13	-13	-23
GC. yC	-11	-6	-34	-5
GG. yC	-16	-29	-14	-14
GU. yC	-21	-8	-23	-7
*-----				
GA. yG	0	0	0	0
GC. yG	0	0	0	0
GG. yG	0	0	0	0
GU. yG	0	0	0	0

³⁰ As we continuously update STAR, you might see different values in your program.

The left column specifies four nucleotides involved in the mismatch interaction. The first and the last of them form the pair in a stem; the second (x) and the third (y) are in the mismatch upon the pair.

For example, the value for mismatch C-G upon the pair A-U, is found in the part of table with A and U being 1st and 4th nucleotides; C in the mismatch is found as X in second line of this part and G(Y) is determined in top row as column 3: the value is 21. Thus the energy is -2.1 kcal/mol.

9.5.7 0 [/View...]: Nonpaired Terminals

When you choose the “NO5PAIRED” you see the following table with free-energy increments for terminal nucleotides with 5’ dangling ends:

NO5PAIRINGRULES					
	x:	A	C	G	U
xA. A		0	0	0	0
xA. C		0	0	0	0
xA. G		0	0	0	0
xA. U		-3	-3	-4	-2
*					
xC. A		0	0	0	0
xC. C		0	0	0	0
xC. G		-5	-2	-2	-1
xC. U		0	0	0	0
*					
xG. A		0	0	0	0
xG. C		-2	-3	0	0
xG. G		0	0	0	0
xG. U		-2	-2	-2	-2
*					
xU. A		-3	-2	-2	-2
xU. C		0	0	0	0
xU. G		-2	-2	-2	-2
xU. U		0	0	0	0

When you choose the "NO5PAIRED" you see the alternative table for 3’ dangling ends:

NO3PAIRINGRULES					
	x:	A	C	G	U
xA. A		0	0	0	0
xA. C		0	0	0	0
xA. G		0	0	0	0
xA. U		-8	-5	-8	-6
*					
xC. A		0	0	0	0
xC. C		0	0	0	0
xC. G		-17	-8	-17	-12
xC. U		0	0	0	0
*					
xG. A		0	0	0	0
xG. C		-11	-4	-13	-6
xG. G		0	0	0	0
xG. U		-8	-5	-8	-6
*					
xU. A		-7	-1	-7	-1
xU. C		0	0	0	0
xU. G		-12	-5	-12	-7
xU. U		0	0	0	0

The top specifies the nonpaired nucleotide (either at 5’- or at 3’-end of base-pair). The left column defines the base-pair.

For example, to calculate the energy of dangling end with A at the 3’-end of G-C pair in the second table we find “G-C” in left column and A in top row: -11. The energy is -11 kcal/mol or -1.1 kcal/mol.

9.5.8 [/View...]: Quasi-junction

The quasi-junction table is still in an experimental stage. When you view the energyrules of a quasi-junction, you get a table like:

QUASI JUNCTION		³¹
0	0	
1	999	<===
2	999	

1. The first column gives the number of nucleotides in the quasi-junction (J in next figure).
2. The second column gives the Gibbs Free Energy in hcal/mol (*not kcal/mol!*) for that quasi-junction..

For example, row two shows you that only pseudoknots are predicted without a quasi-junction because a quasi-junction of 1 nucleotide has a penalty of 999.

When the values for a certain length are missing, *STAR* uses the most nearby lower value. In this example the value for a quasi-junction of 6 nucleotides long is 999.

³¹As we continuously update STAR, you might see different values in your program.

9.5.9 [/View...]: Shallow Groove

When you view the energyrules of a connecting loop spanning a shallow groove, you see a table like:

SHALLOWGROOVERULES³²

	0	1	2	3	15	16
0	+	---	---	---	---	---
1		999	999	999	999	999
2		999	999	42	42	999
3		999	42	42	42	999 <===
4		999	42	42	42	999
6		999	42	42	42	999
7		999	999	999	999	999

The top row specifies the number of nucleotides that cross the shallow groove (L2 in above figure of the previous paragraph). The left column specifies the number of basepairs to be crossed (S1 in next figure).

For example you want to know the energy value of a connecting loop of 2 nucleotides crossing the shallow groove over 4 base pairs. The energy value is then 42 hcal/mol.

9.5.10 [/View...]: Stacking Energyrules

The stacking energyrules include the energy values that follow from the stacking of basepairs. *STAR* displays them in a table, as shown below.

STACKINGRULES³³

	AA	AU	AG	AC	UA	UU	UG	UC	GA	GU	GG	GC	CA	CU	CG	CC
AA	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
AU	20	-12	20	20	-12	20	-3	20	20	-3	20	-21	20	20	-21	20
AG	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
AC	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
UA	20	-18	20	20	-12	20	-3	20	20	-3	20	-21	20	20	-21	20
UU	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
UG	20	-3	20	20	-3	20	-3	20	20	-3	20	-13	20	20	-13	20
UC	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
GA	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
GU	20	-3	20	20	-3	20	-3	20	20	-3	20	-13	20	20	-13	20
GG	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
GC	20	-21	20	20	-21	20	-13	20	20	-13	20	-48	20	20	-30	20
CA	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
CU	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
CG	20	-21	20	20	-21	20	-13	20	20	-13	20	-43	20	20	-48	20
CC	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

In this table, you find the energy of a particular stack by selecting the row with the bottom basepair in the left column and the top base-pair in the first line. The intersection is the Gibbs Free Energy of the stack in hcal/mol (*not kcal/mol!*). The positive values of 20 for non-Watson-Crick pairs are chosen for programming reasons only.

For example if a GC basepair stacks upon a UG basepair, the Gibbs Free Energy is -13 hcal/mol. See the underscored number in above table.

This explains the energy of a stack of two basepairs. To calculate the energy of a stack of more basepairs we give the following example:

GC	
CG	- 43
AU	- 21
GU	- 3
	----- +
	- 67 hcal/mol

³² As we continuously update STAR, you might see different values in your program.

³³ As we continuously update STAR, you might see different values in your program.

So you start at the top of the stack and work your way downwards until the stack has ended. For each stack of two basepairs, you search the energy in the table. Finally you add these values to get the Gibbs Free Energy of the stack.

9.6 [/EDIT...]

The [/Edit...] option enables you to change the default energy rules if you are not satisfied with them. After such changes you should save the new values for later use, as *STAR* always begins a session with the default values. You can also edit energyrules with a regular editor. Appendix D as well as this chapter show the proper format. You can import such a file using [Energyrules/Open] (see paragraph 9.2).

We will explain the editor in the following paragraphs.

You enter the editor by activating the option [/Edit...] from the [Energyrules] menu. Subsequently the following box appears.

CHOOSE ENERGYRULES	
Quasijunction	Hairpin loop
Bulge	Internal loop
Formula const.	Mismatches
Deep Groove	Stacking
Shallow Groove	No 3' pairing
	No 5' pairing
Cancel	

Use the buttons to select the type of energyrules you want to edit. If you don't understand the different types of energyrules, consult paragraph 9.5 as we explained them there.

When you press [Cancel], you will return to the main menu.

After you have entered the editor, you see the energyrules on screen.

There are 11 different types of energyrules and the editorscreen differs for all of them.

For explanation of general features like adding text, moving text, copying text and quitting the editor please consult paragraph 7.6 [Primary/Edit] or 8.8 [Secondary/Edit].

Adding/deleting rows/columns

In many tables you can add rows (or columns). If you do so, be sure to add as many values as the other rows (columns) contain (including the first column which is often not an energyvalue, but a type of entry specification).

For example

0	1	2	
0 +	---	---	
1	999	999	
2	999	15 999	
	Erroneous		

0	1	2	3
0 +	---	---	---
1	999	999	999
2	999	15	999
	Correct		

The tables are displayed in a regular layout. If you change values, or add rows or columns you are free to use as many blanks as you want (“free format”). As long as the number of values match with the other rows (and columns), *STAR* does not care about your layout.

For example, the following is quite correct:

```

0   1   2       3
0 + ---|---
1   | 999 999 999
2 | 999 15 999
```

Comment

Whenever you feel the need, you can add comment by separating the values left and the comment with an asterisk.

For example:

```

0 40 This is erroneous
1 30 * This is correct comment
```

9.6.1 Edit Bulge loop Energyrules

The table with bulge loop energyrules is the same as the one discussed by the [/View...] option.

You can change the energy values in column 2. You can also add or delete rows. Remember that by doing so you must complete or delete both values for that particular row.

9.6.2 Edit Deep groove Energyrules

The table of energyrules of a connecting loop spanning a deep groove is basically the same as the one discussed by the [/View...] option.

You can change all energyvalues in the table. You can also add or delete rows and/or columns. Remember that by doing so you must complete all values for that particular row including the heading value.

You can also add or delete a column. In that case you should add or delete all values for that column, including the heading value.

9.6.3 Edit Formula Constants

The table with formula constants is basically the same as the one discussed by the [/View] option.

You should only change values, not delete nor add values.

9.6.4 Edit Hairpinloop Energyrules

The table for energyrules of a hairpinloop is the same as the one discussed by the [/View...] option.

You can also add or delete one or more value pairs. Remember that you must always add or delete a pair, separated by a colon; a single value will cause an error.

The length in the odd columns represents the number of free nucleotides in the string.

9.6.5 Edit Internalloop Energyrules

The table with internalloop energyrules is basically the same as the one discussed in the [/View...] option.

You can change the energy values in column 2. You can also add or delete rows. Remember that by doing so you must complete or delete both values for that particular row.

9.6.6 Edit Mismatch Energyrules

The table with energyrules for mismatches is basically the same as the one discussed by the [/View] option.

You should only change values, not delete or add values.

9.6.7 Edit Nonpaired Terminal Nucleotides

The tables with energyrules for terminal nonpaired nucleotides is basically the same as the ones discussed by the [/View] option.

You should only change values, not delete nor add values.

9.6.8 Edit Quasi-junction Energyrules

This table is still in an experimental stage.

The table with energyrules of a quasi-junction is basically the same as the one discussed by the [/View...] option.

You can change all energyvalues in the table. You can also add or delete rows. Remember that by doing so you must complete both values for that particular row.

9.6.9 Edit Shallow groove Energyrules

The table with energyrules of a connecting loop spanning a shallow groove is basically the same as the one discussed by the [/View...] option.

You can change all energyvalues in the table. You can also add or delete rows and/or columns. Remember that by doing so you must complete all values for that particular row including the heading value.

You can also add or delete a column. In that case you should add or delete all values for that column, including the heading value.

9.6.10 Edit Stacking Energyrules

The stacking energyrules are collected in a table. This table is basically the same as the one discussed in the [/View...] option.

You should only replace values. You should not add rows or columns, nor delete rows or columns.

9.6.11 Possible errors

1. WRONG CHARACTERS:

If you want to leave the editor, and there are characters in your structure (note: to the right of the asterisks you can type anything you want) other than numbers, blanks, asterisks, and sometimes decimal points ³⁴ then *STAR* warns you with the following box:

Only numbers, blanks and points
are allowed! ³⁵
Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor-screen, where you can correct the wrong characters.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrules from before you started editing will remain in the computer.

1. WRONG FORMAT:

The format of all the energyrules must obey to some format-rules.

They are listed here:

1. Bulgeloop energyrules:

The table must always consist of 2 columns.

2. Deep groove energyrules:

You can make the table as large as you want, but all the columns and rows must be completely filled.

3. Hairpinloop energyrules:

The table must always consist of 5 double columns. (this means $5 \times 2 = 10$).

The first value of each value-couple is the length of the hairpinloop, and these values must be in increasing order.

4. Internalloop energyrules:

The table must always consist of 5 columns.

5. Quasi-junction energyrules:

The table must always consist of 2 columns.

6. Shallow groove energyrules:

You can make the table as large as you want, but all the columns and rows must be completely filled.

7. Stacking energyrules:

The table must always consist of 16 rows and 16 columns.

If the format of your edited energyrules-table is not as described above, the following box appears:

Format error! Edit

.....³⁶

again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor- screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws all your changes from the edit-session away. The old energyrules from before you started editing will remain in the computer.

³⁴ Points are only allowed when they are used in numbers, not when they are used on their own.

³⁵ The default colons in the hairpinlooprules, and the default dividing-lines in the headers of the energy-tables are also harmless.

³⁶ The dashes stands for the type of energyrules you were editing. (Eg. "stackingrules" or "hairpinlooprules", etc.).

3. DOUBLE, NEGATIVE OR NON-INTEGGER VALUES IN COLUMN 1:

If there are double values, negative values or non-integer values in column 1 of your edited energyrules-table, the following box appears when you try to leave the editor:

ERROR. Double, negative or
non-integer values in column 1:

0.8 32 41 49 56 ³⁷

Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrules from before you started editing will remain in the computer.

4. DOUBLE, NEGATIVE OR NON-INTEGGER VALUES IN ROW 1:

If there are double values, negative values or non-integer values in row 1 of your edited energyrules-table, the following box appears when you try to leave the editor:

ERROR. Double, negative or
non-integer values in row 1:

16 999 999 999 999 999 ³⁸

Edit again or cancel?

OK

CANCEL

³⁷ Example of a wrong energy-value (non-integer value in column 1).

³⁸ Example of a wrong energy-value (the value 16 appeared twice in the first row of the energyrules-table).

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrules from before you started editing will remain in the computer.

5. NUMBER OF NUCLEOTIDES CROSSING THE GROOVE ARE NOT IN INCREASING ORDER:

The values in the first row of the table must be in increasing order. If they are not, *STAR* gives the following message:

ERROR. Values in row 1 not in
increasing order!

Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrules from before you started editing will remain in the computer.

6. NUMBER OF NUCLEOTIDES BEING CROSSED, ARE NOT IN INCREASING ORDER:

The values in the first column of the table must be in increasing order. If they are not, *STAR* gives the following message:

ERROR. Values in column 1 not
in increasing order!
Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrules from before you started editing will remain in the computer.

7. LENGTH OF HAIRPINLOOPS NOT INCREASING, OR NEGATIVE OR NON-INTEGER:

The first value of each value-couple in the hairpinloop energyrules-table is the length of the hairpinloop, and these values must be in increasing order. They also must be positive and integer. If they are not, *STAR* gives the following message:

ERROR. Length of hairpinloops
not in increasing order, or
double or negative values!
Edit again or cancel?

OK

CANCEL

Just press the [OK]-button, and you will return to the editor-screen, where you can correct your mistake.

But if you made a terrible mess and you don't want to correct it anymore, just press the [CANCEL]-button. *STAR* throws away all your changes from the edit-session. The old energyrulese from before you started editing will remain in the computer.

8. TOO MANY CHARACTERS:

When you open a energyrules-file, and after that try to edit it in the *STAR* editor, it can happen that there are too many characters on a line. If that is the case, *STAR* warns you with the following box:

Some lines have been truncated.
Select cancel to quit from the
editor.

CANCEL

CONTINUE

When you press [CANCEL], nothing will happen and you will return to the main menu. When you press [CONTINUE] also nothing serious will happen. The only thing *STAR* does, is using a few extra lines. You will not loose any energy-value or comment.

9.7 [/PRINT...]

You can use the [/Print...] option to make a hard-copy (a print) of your energyrules.

When you activate this option the following box appears:

CHOOSE ENERGYRULES

Quasijunction	Hairpin loop
Bulge	Internal loop
Formula const.	Mismatches
Deep Groove	Stacking
Shallow Groove	No 3' pairing
No 5' pairing	
<div style="border: 1px solid black; display: inline-block; padding: 5px 20px;">Cancel</div>	

Use the buttons to select the type of energyrules you want to print. If you don't understand the different types of energyrules, consult paragraph [Secondary/View] as we explained them there.

When you press [CANCEL], you will return to the main menu.

After you have chosen which energy values to print, the following box appears. With this box you can specify where you want to print to, and enter the size of your printer paper.

PRINT OPTIONS

Page length ___ lines
 Page width ___ columns
 Top margin ___ lines
 Bottom margin ___ lines

Print to: ☐File ☐Printer

OK

CANCEL

In the four fields marked by ____, you may enter the length and width of the paper you use, the top margin and the bottom margin. *STAR* already provided default values that are fine in most cases.

The top margin is the blank part above the text on a page. The bottom margin is the blank part below the text on a page.

For instance, if you use paper 71 characters long, enter "71" at the first field.

For a standard A4 paper format, you must enter "66" for the pagelength, and "80" for the page width. For standard continuous forms paper, enter "71" for the pagelength and "80" for the pagewidth. We advise for both top and bottom margin a length of 2 lines.

You can specify whether to print to a printer or to a file.

Printing to a printer means that *STAR* outputs directly to a printer. In this case a printer must be attached to your computer and it must be switched on.

If you have any problem with printing, you can print to file. *STAR* asks you to enter a filename using the file selector. You can print this file separately after leaving *STAR*.

If you want to include the *STAR* output in a publication you can also use "Print to File". The file can be read in any text processor. Use a non-proportional font in your textprocessor to display the energyrules properly.

1 Possible errors

1. NO PRINTER ATTACHED:

If you specify you want to print to a printer while there is no printer attached to your computer, you can not use *STAR* for about 30 seconds.

2. PRINTER BUSY:

This message can happen in network printers. *STAR* detects a busy signal and will not print. Unfortunately this is a network problem, not a *STAR* bug. Your best solution is to use option "Print to: File" and print the file with another program (textprocessor).

[illegible][illegible]

10. APPENDIX A: STAR PROCEDURE

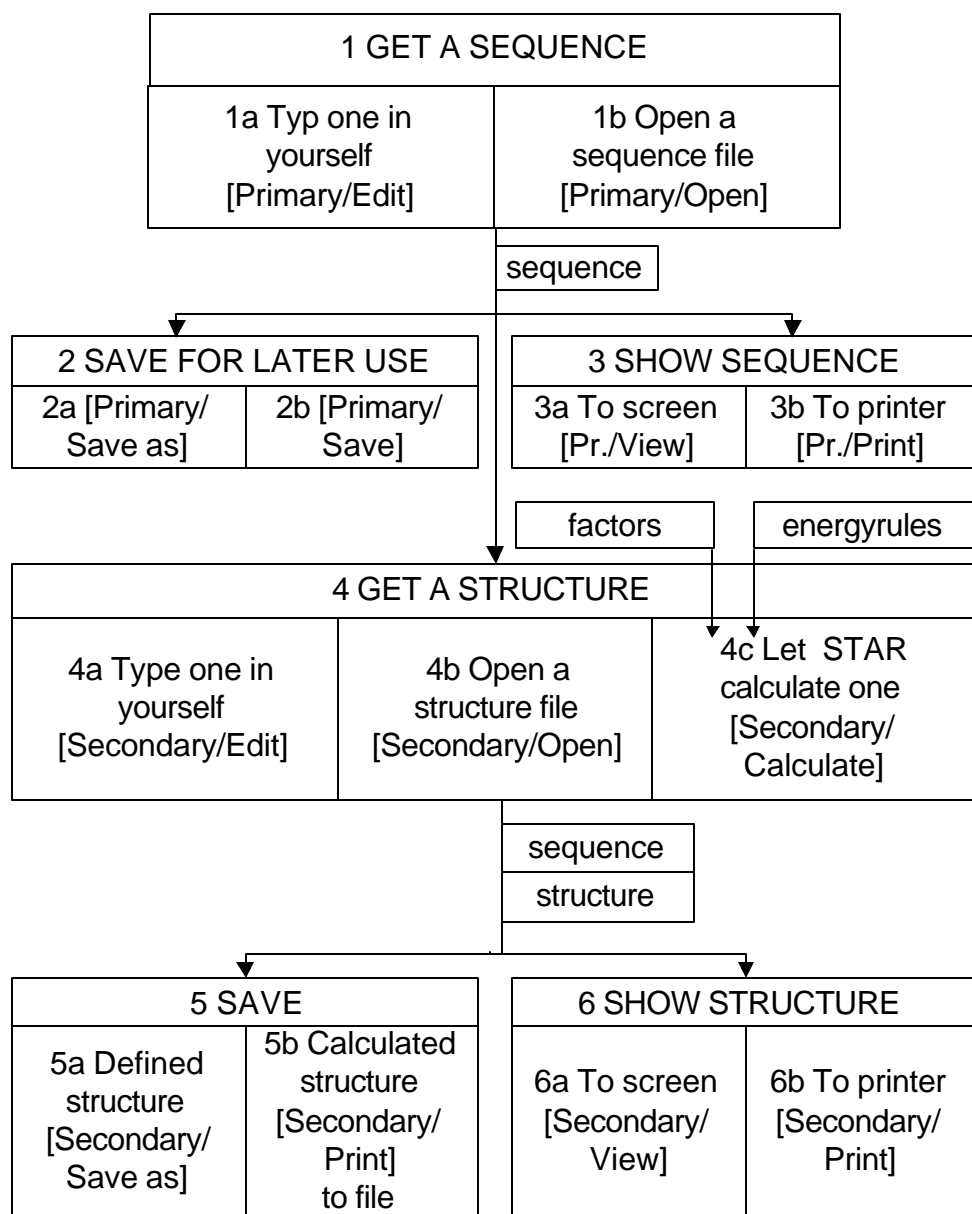
In this appendix we show the order in which you use the different options more visually.

The top part of the boxes represent actions you should take at a certain moment. We numbered them 1-6. You can perform most of these actions in various ways, which we numbered 1a,1b and so on.

The bottom part of the boxes represent the input for a certain action.

For example to let *STAR* calculate a structure (action 4c), the energyrules are input.

We show the options and the corresponding menu options between [...].



11. APPENDIX B: ERRORMESSAGES

In this appendix we list all errormessages alphabetically . For each message we give an explanation and what you should do if you meet that particular message.

-
- | | |
|--|--|
| <p>1. CAN NOT OPEN FILE: "wrong filename".
CAN ONLY OPEN FILE: "correct filename"
You tried to open a structure file with a name which is not allowed. Structure files must have the same name as the sequence you opened, except that the extension must be ".2" instead of ".1".
Remedy:
Check for mistake and enter correct name.</p> <p>2. DISK FULL
The disk you are trying to save a file to is full.
Remedy:
Save to another disk.</p> <p>3. DISK IS PHYSICALLY WRITE-PROTECTED
The disk you are trying to save a file to is write protected.
Remedy:
Remove the write protection or use another disk.</p> <p>4. ERROR. DOUBLE, NEGATIVE OR NON-INTEGER VALUES IN COLUMN 1:..... EDIT AGAIN OR CANCEL?
You were editing energyrules (intlooprules, bulgerules, deep grooverules, shallow grooverules or quasi-junction rules) and when you tried to leave the editor these energyrules contained double, negative or non-integer values in the first column.
Remedy:
Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and <i>STAR</i> discards all the changes made in this edit-session.</p> | <p>5. ERROR. DOUBLE, NEGATIVE OR NON-INTEGER VALUES IN ROW 1:..... EDIT AGAIN OR CANCEL?
You were editing energyrules (deep grooverules, shallow grooverules or quasi-junction rules) and when you tried to leave the editor these energyrules contained double, negative or non-integer values in the first row.
Remedy:
Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and <i>STAR</i> discards all the changes made in this edit-session.</p> <p>6. ERROR. ILLEGAL FORMAT! FILE NOTOPENED
You tried to open a structure- or energyrules-file, of which the format was not correct. A structure-file should consist of a six-column table. For the format of the specific energyrules see paragraph 9.6.5 as we explained it there.
Remedy:
All you can do is press [OK], after which you return to the main menu. The file is not opened.</p> <p>7. ERROR. ILLEGAL VALUES IN COLUMN 1! FILE NOT OPENED!
You tried to open a energyrules-file (intlooprules, bulgerules, deep grooverules, shallow grooverules or quasi-junction rules), but the first column of that file contained double, negative or non-integer values, or these values were not in increasing order.
Remedy:
All you can do is press [OK], after which you</p> |
|--|--|

return to the main menu. The file is not opened.

8. ERROR. ILLEGAL VALUES IN ROW 1!
FILE NOT OPENED!

You tried to open a energyrules-file (deep grooverules, shallow grooverules or quasi-junction rules), but the first row of that file contained double, negative or non-integer values, or these values were not in increasing order.

Remedy:

All you can do is press [OK], after which you return to the main menu. The file is not opened.

9. ERROR. ILLEGAL VALUES USED IN
FILE. FILE NOT OPENED!

You tried to open a hairpinloop-energyrules file, and the lengths of the hairpinloops contained double, negative or non-integer values. It is also possible that the lengths of these loops were not in increasing order.

Remedy:

All you can do is press [OK], after which you return to the main menu. The file is not opened.

10. ERROR. IN SEQUENCE ONLY
A,C,G,U,a,c,g,u ALLOWED! EDIT AGAIN
OR CANCEL?

You were editing a sequence and when you tried to leave the editor, your sequence contained other characters than A,C,G,U,a,c,g,u and spaces. (NOTE: not in your comments, there you can type anything you want).

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes (*STAR* replaces wrong characters by "", so they are easy to find), or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

11. ERROR. LENGTH OF HAIRPINLOOPS
NOT IN INCREASING ORDER, OR
DOUBLE, NEGATIVE OR NON-INTEGERS
VALUES! EDIT AGAIN OR CANCEL?

You were editing hairpinloop-energyrules and when you tried to leave the editor the lengths of the hairpinloops contained double, negative or non-integer values.

It is also possible that the lengths of the hairpinloops were not in increasing order.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

12. ERROR. NON-INTEGERS VALUES IN ROW
...: EDIT STRUCTURE AGAIN OR
CANCEL?

You were editing a structure, and used periods in your stem-numbers to make non-integer values.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

13. ERROR. ONLY OF NAME *.... CAN
BE OPENED!

You selected a type of energyrules to open, and entered a filename with the wrong extension. (Eg. you selected the bulgerules to open, and entered the filename BULGE.TXT as filename, which is not allowed). See also paragraph 9.2.

Remedy:

All you can do is press [OK], after which you return to the main menu. The file is not opened.

14. ERROR. STEM NOT ALLOWED
(OVERLAP WITH OTHER STEM). EDIT
AGAIN OR CANCEL?

During the editing of a structure you made a

stem which overlaps with another stem.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

15. ERROR. STEMS NOT ALLOWED
(UNEQUAL LENGTH STEMHALVES).
EDIT AGAIN OR CANCEL?

During the editing of a structure you made a stem with two stemhalves that are unequal in length.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

16. ERROR. STEMNUMBERS NOT IN
SUCCEEDING ORDER! EDIT AGAIN OR
CANCEL?

During the editing of a structure you made a stem of which the stemnumbers were not in increasing order.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

17. ERROR! THIS IS NOT A SEQUENCE-FILE!
FILE NOT OPENED.

You tried to open a non-sequence file as a sequence-file.

Remedy:

Press [OK] and you return to the main menu. Now try to open another sequence-file.

18. ERROR. USE .PRT AS EXTENSION. FILE
"filename" NOT SAVED

You tried to make a printerfile with another extension then ".prt" or ".PRT".

Remedy:

All you can do is press [OK], after which you return to the main menu. The file is not saved.

19. ERROR. USE ... AS EXTENSION. FILE
NOT SAVED

You tried to save a energyrules-file with a filename which had the wrong extension. Valid names are: "*.sta", "*.int", "*.bul", "*.hai", "*.dee", "*.sha", "*.qju", "mis", "for", "no3", or "no5".

Remedy:

All you can do is press [OK], after which you return to the main menu. The file is not opened.

20. ERROR. VALUES IN COLUMN 1 NOT IN
INCREASING ORDER! EDIT AGAIN OR
CANCEL?

During the editing of energyrules the values in column 1 were not in increasing order.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

21. ERROR. VALUES IN ROW 1 NOT IN
INCREASING ORDER! EDIT AGAIN OR
CANCEL?

During the editing of energyrules the values in row 1 were not in increasing order.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

22. ERROR. WRONG INPUT! TRY AFGAIN
OR CANCEL?

You changed the factors in the secondary menu, but you entered illegal values.

Remedy:

Press [OK], after which you re-enter the factors-box, where you can correct your

mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in the factors.

23. FILENAME "filename" IS NOT ALLOWED.

You used the option [/Save as...] to save a structure. The filename you entered is not allowed. The first part of the filename must be equal to the name of the *sequence file* that you opened. With the first part of a filename we mean the part of the name preceding the extension. The extension of the filename must begin with "2". The extension of a filename is part of the name following the point ".".

For example: if you have opened the sequence file "TMV.1" you can save your structure only to filenames beginning with "TMV.2".

Remedy:

All you can do is press [OK] and return to the main menu. Now try to save again with a correct filename.

24. FORMAT ERROR! EDIT AGAIN OR CANCEL?

You were editing a structure or energyrules (this is indicated on the dots), and when you tried to leave the editor the format was not correct. Structures should consist of a four-column table, for the format of the specific energyrules see paragraph 9.6.5.

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

25. NEW FILENAME ALREADY EXISTS! DO YOU WISH TO CONTINUE?

When you tried to rename a file, the filename you selected for the new file already existed.

Remedy:

Select [OK] and *STAR* replaces the old file by the new one, or select [CANCEL] and

nothing happens, you return to the main menu

26. ONLY NUMBERS, BLANKS AND POINTS ARE ALLOWED! EDIT AGAIN OR CANCEL?

You were editing a structure, or energyrules, and when you tried to leave the editor the structure or energyrules contained other characters than numbers, blanks, and points. (Points are only allowed in numbers, colons are allowed in the hairpinloop-energyrules).

Remedy:

Press [OK], after which you re-enter the editor, where you can correct your mistakes, or press [CANCEL] after which you return to the main menu, and *STAR* discards all the changes made in this edit-session.

27. PREDICTING A STRUCTURE NOW WILL ADD STEMS TO YOUR CURRENT STRUCTURE

You edited a structure, and after that you selected the option [Calculate]. Now *STAR* warns you that the edited stems will be forced into your structure.

Remedy:

If you think this is alright, you can press [OK], and the calculation starts. If you on the other hand don't want these stems to be incorporated, then press [CANCEL], and you return to the main menu. There you can save your old structure and/or erase the already existing stems in the editor.

28. PREDICTING A STRUCTURE NOW OVERWRITE YOUR CURRENT STRUCTURE

You already had a (calculated, not edited) structure in the computers memory when you selected the option [/Calculate]. Now *STAR* warns you that the old stems will be erased by the new ones.

Remedy:

If you think this is alright, you can press [OK], and the calculation starts. If you on the other hand don't want the old stems to be

overwritten then press [CANCEL], and you return to the main menu. There you can save your old structure.

29. REPLACE EXISTING "filename"?

The filename you selected to save to, already exists.

Remedy:

If you want to replace that file by the new one, select [YES]. If you don't want to replace the old file, then select [NO], after which you return to the main menu. There you can save again with another filename, or rename the old file with the option [Files/Rename...] (see paragraph 6.2).

30. QUIT THIS EDIT WINDOW & ABANDON CHANGES?

You left the editor by selecting the option [Edit/Quit unchanged] in the menu of the macintosh. Now *STAR* asks you if you are sure you don't want to save the changes made.

Remedy:

Press [OK] if you don't want to save the changes made. But if you want to save your changes made while you were editing, then press [CANCEL], after which you re-enter the editor. There you can leave the editor in another way.

31. SOME LINES HAVE BEEN TRUNCATED. SELECT CANCEL TO QUIT FROM THE EDITOR.

You opened a sequence, structure, or energyrules, and in these file there were too many characters on a line. When you enter the editor *STAR* warns you, with this error-message.

Remedy:

When you select [CANCEL] nothing happens, and you return to the main menu. When you select [CONTINUE] also nothing serious happens, *STAR* spreads your sequence, structure or energyrules over some extra lines. You will not loose anything of your sequence, structure or energyrules.

32. THIS IS NOT A STRUCTURE-FILE!

You tried to open a non-structure file as a structure-file.

Remedy:

All you can do is press [OK] after which you return to the main menu. Now you can try to open another structure-file.

33. WARNING - THE FILENAME YOU HAVE SELECTED ALREADY EXISTS. DO YOU WISH TO CONTINUE?

The filename you selected to save to, already exists.

Remedy:

If you want to replace that file by the new one, select [YES]. If you don't want to replace the old file, then select [NO], after which you return to the main menu. There you can save again with another filename, or rename the old file with the option [Files/Rename...] (see paragraph 6.2).

34. WS FULL

The computer you use has not enough memory for the operation.

Remedy:

You must restart *STAR* with more memory.

12. APPENDIX C: GLOSSARY

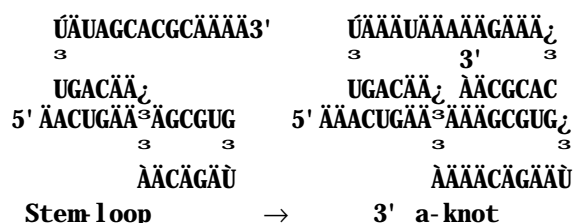
In this appendix you find alphabetically ordered information about some technical computer words relevant for RNA structure.

1. *A-knot*

We use the name "a-knot" for a type of indirect stacking of stems.

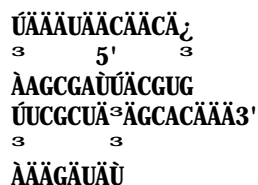
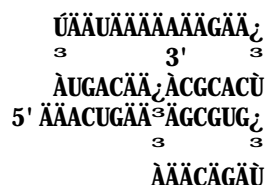
This type of stacking is hypothetical: the a-knot has not been found in nature so far.

You can see an a-knot as an ordinary stem-loop structure of which the free 3' end or the free 5' end crosses the backbone of the stem and basepairs with the loop. You see this in the figure shown below which shows the formation of a 3' a-knot. In this figures, 3 nucleotides cross the backbone while spanning 9 base pairs.



Both a 5' and a 3' a-knot can exist.

We show examples in the figures below.



2. *APL*

The programming language in which we wrote *STAR*. For Atari and Macintosh we used APL.68000 from ©MicroAPL, for the PC we used APL*PLUS from ©Manugistics.

3. *ASCII*

American Standard Code for Information Interchange: a certain code that many computers use so they can communicate with each other.

STAR can open sequence-, structure- or energyrules-files originating from other programs or databases, if they are in ASCII.

4. *Asymmetric internal loops*

Any internal loop is formed by two stems and have two parts. It was suggested that loops with equal parts (symmetric) are more favourable compared to asymmetric ones.

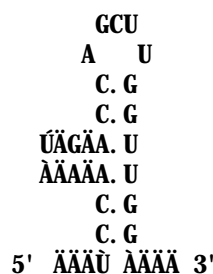
This is taken into account by *STAR*: the more the difference between two parts of a loop, the more the destabilizing penalty.

5. *Bulge*

"Bulge" is the name for an internal loop where one part does not have nucleotides.

Both 3' and 5' bulges can exist.

You can see an example of a small 5' bulge in the figure below.



The names of the energyrules that *STAR* uses to predict a bulge are "Stacking" and "Bulge loop".

6. *Bulge loop*

"Bulge loop" is the name of the table with energy values for bulge loops in RNA strings.

7. Crossover

Here crossover is used in the context of GA. For every iteration, a new structure is created from a random mix of all structures in the population. See also Genetic Algorithm, mutation, population and selection.

8. Deep groove

"Deep groove" is the name of the table for energy values of the connecting loop crossing the deep groove in a pseudo knot. See explanation and figure at shallow groove and pseudoknot table.

9. Direct stacking

Two base pairs can stack directly or indirectly on each other. The base pairs stack directly if the stacking nucleotides are next to each other in the RNA sequence. Compare the figure below with the one of "indirect stacking".

```

3 3
A. U
3 3
C. G
3 3

```

10. File-selector

A box which you use to select files.

11. Fork

"Fork" is the name for one of the types of indirect stacking of two base pairs. You can see an example of a fork in the figure below.

```

      UAG
      A   C
      U. A
      U. A
      G. C
5'  AAAAC. G
3'  AAAA. U
      U. A
      G. C
      C   C
      GUG

```

Currently, *STAR* does not take the indirect stacking into account, but just deals with the direct stacking using tables "Stacking" and "Hairpin loop".

12. Formula constants

Some energy values are calculated by *STAR* using formulas. The user can change the values of adjustable parameters in these expressions.

13. Genetic Algorithm

A Genetic Algorithm (abbreviated as GA) is an iterative procedure that finds its final solution by improving its initial solution gradually. Improvements are accomplished by steps that are called "mutation", "crossover" and "selection" (see those terms) for each iteration. So the algorithm folds RNA through a procedure of stepwise improving local equilibrium structures. We feel that those steps mimic the folding pathway better than those in the other algorithms. See also mutation, crossover, selection, population, greedy algorithm, and stochastic algorithm.

14. Greedy Algorithm

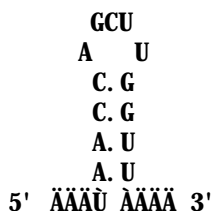
The greedy algorithm folds RNA through a procedure of stepwise improving local equilibrium structures. In each step it adds the stem with the greatest energy contribution. Therefore, a much better description is "greedy algorithm" (so coined by Paul Higgs).

15. Growth

All algorithms can simulate RNA synthesis during folding. For the greedy algorithm we use a rather experimental formula. The other algorithms mimic the synthesis much more closely. This is done by starting the folding using a small initial part at the 5' end of the RNA chain only. In subsequent iterations this initial parts "grows" by increasing it with small increments.

16. Hairpin

"Hairpin" is the name for one of the types of loops. You can see an example of a hairpin loop in the figure below.



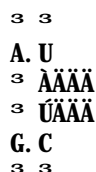
The names of the energyrules that *STAR* uses to predict a hairpin are "Stacking" and "Hairpin loop".

17. *Hairpin loop*

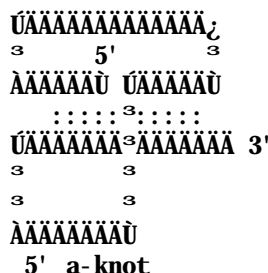
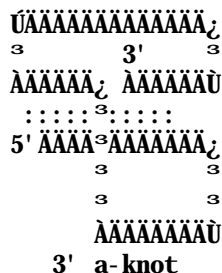
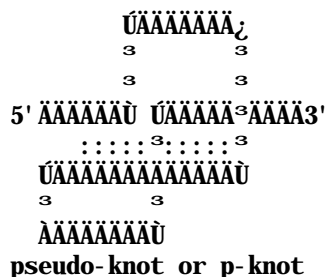
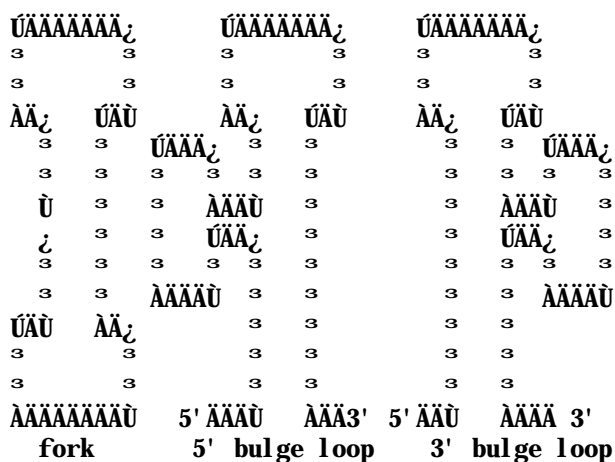
"Hairpin loop" is the name of the table with energy values for hairpin loops in RNA strings.

18. *Indirect stacking*

Two base pairs can stack directly or indirectly on each other. The base pairs stack indirectly if nucleotides are next to each other in one strand only. Compare the figure below with the one of "direct stacking".

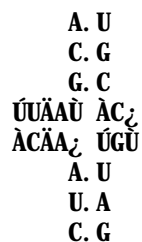


In principle, one can distinguish 6 types of indirect stacking. Two of them, the 3' a-knot and the 5' a-knot, are not known to exist so far. You can see the six types in the figures shown below.



14. *Internal loop*

"Internal loop" is the name of the table with the energy rules for internal loops in RNA strings. You can see an example of an internal loop in the next figure.



Note that the internal loop degenerates to a bulge loop when one of the 2 strands reduces to zero-length.

15. Minimal loop length

RNA needs a minimum length to form a hairpin-loop. This minimum length is 3 nucleotides. We have defined this by setting the energy values for loop lengths smaller than and equal to 3 nucleotides to 99.9 kcal/mol. You can find the energy values for various loop lengths in the energyrules called "hairpin loop". You can change these values so that the minimum loop length will be bigger (or even smaller) than 3 nucleotides.

16. Mismatches

Mismatches are formed by two nucleotides adjacent to a stem. They are assumed to have favourable energy contribution (due to the stacking), being located in hairpins and internal loops.

```

      3 3
      C. G
      U. A
      C   A <===
      N     N
      A   G <===
      C. G
      C. G
      3 3

```

17. Multibranched loops

Multibranched loops are formed by several (more than 2) stems. The energies for these loops are poorly investigated and usually the linear dependence on loop size is assumed.

```

      3 3
      C. G
      U. A
      N   N
      N   ACGACGÄÄ
      N   UGCUGCÄÄ
      N   G
      C. G
      C. G
      3 3

```

18. Mutation

Here mutations are used in the context of GA. For every iteration, all structures of a population are subjected to mutation. This means that for each structure some randomly chosen stems are added and/or disrupted. The probabilities of stem addition and disruption depend on energy contributions of stems: the more stable the stem, the greater the probability to include it to the structure and the less the probability to remove it. See also Genetic Algorithm, crossover, population and selection.

19. Nucleotides

You can add nucleotides by typing the letter corresponding with that particular nucleotide.

The keys that correspond with each type of base are:

a or A: Adenine
c or C: Cytidine
u or U: Uracil
g or G: Guanine

You can enter the nucleotides both in lowercase and uppercase.

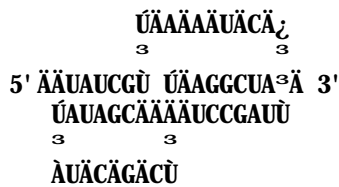
Remember however that *STAR* calculates a structure of uppercase nucleotides only!

This feature is particularly useful if you want to exclude a certain part of a sequence to basepair. You exclude that part by typing it in lowercase.

Normally you are advised to type with the [CapsLock] key pressed.

20. P-knot

"P-knot" is the name we used for one of the ways in which two stems can stack indirectly on each other (see also "indirect stacking"). You can see an example of a p-knot in the figure below.



You can see it as an ordinary stem-loop structure of which the free 3' end crosses the shallow groove of the stem and basepairs with the loop, or the free 5' end crosses the deep groove of the stem and basepairs with the loop.

We summarized the energy values used to calculate the energy it cost for the free end to cross either the deep groove or the shallow groove of a stem in the energyrules called "deep groove" and "shallow groove" respectively.

Population

Here selection is used in the context of GA. Each iteration creates not 1, but several alternative structures. This set of structures is called the population. After applying mutation and crossover, a new population is formed for the next iteration by selection of the “best” structures.

Pseudoknot tables

STAR uses 3 energyrule tables for pseudoknots: Deep groove table, Shallow groove table and Quasi-junction. The relation between rows and columns in the tables is as follows:

Quasi-junction

"Quasijunction" is the name of the table with energy values for a junctions in a pseudoknot. See the figure below with a quasi-junction J.

24.*Secondary structure*

The "secondary structure" of RNA is the assembly of structural elements which are formed by intramolecular basepairing. Besides the classical elements like stems, hairpin loops, buldge loops, internal loops and multiple loops, *STAR* is able to predict some types of tertiary interactions (see also "structure" and "tertiary structure").

25.*Selection*

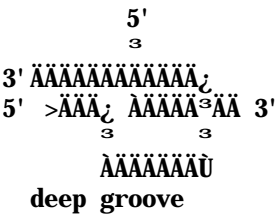
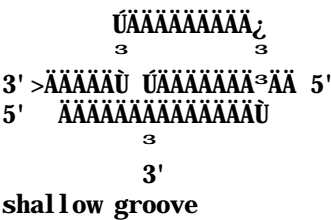
Here selection is used in the context of GA. For every iteration, all structure in the population is mutated once and the crossover includes all stems of all current structures. So a population of N structures produces a temporary population of 2N+1 possible structures. Then a new population is created by selecting the N best structures according to the fitness (free energy).

26.*Greedy algorithm*

The greedy algorithm folds RNA through a procedure of stepwise improving local equilibrium structures. In each step it adds the stem with the greatest energy contribution. Therefore, a much better description is "greedy algorithm" (so coined by Paul Higgs). Unlike GA, this algorithm does not take disruption of stems into account.

27.*Shallow groove*

"Shallow groove" is the name of the energyrules for one sort of p-knots. In a p-knot one loop crosses the deep groove, the other crosses the shallow groove. Following the RNA strand in a p-knot from the 3' side, the first loop you met crosses the shallow groove. Following the RNA strand in a p-knot from the 5' side, the first loop you met crosses the deep groove. We illustrate this in the figure shown below.



28.*Stacking*

"Stacking" is the name of the energyrules for the stacking of basepairs upon each other. For an example of how to calculate the stacking energy of a stem see the figure below.

3	3	
C	G	energy of C G over A U: -2.1
A	U	energy of A U over C G: -2.1
C	G	energy of C G over G C: -3.0
G	C	-7.2
5'	ÄÜ ÄÄ 3'	

29.*Stacking table*

During the calculation process, *STAR* first accumulates all possible stackings in a stacking table. If memory suffices the size of this table is the squared length of the sequence.

30. Stochastic Algorithm

The stochastic algorithm folds RNA through a procedure of stepwise improving local equilibrium structures. The stochastic algorithm tests if some stem combinations are better than the (energetically) best stem. Thus this algorithm applies a Monte Carlo procedure. At every step, the addition of stems to the final structure starts with the generation of a "population" of randomized structures. This population uses the most stable stems that are compatible with the previously formed (intermediate) structure. Next, stems are added with probabilities depending on their energy values. Unlike Genetic Algorithm, stochastic folding does not take disruption of stems into account.

31. Structure

The structure of a RNA molecule is defined in *STAR* by the nucleotide numbers of the bottom and top nucleotide pair of a stem. For example the stem in the figure below is defined by "14 17 23 26".

```

      AUG
    A   U
      C. G
      C. G
      U. A
      C. G
5' AGACUCGAUAGCGÄÜ ÄÄ AUCUCÄÄÄ3'

```

32. Tertiary structure

Here the term "tertiary structure" is reserved for those interactions which are formed by Watson-Crick basepairing of a loop region with a complementary region outside that loop (see for example "a-knot" and "p-knot"). *STAR* is capable of dealing with such interactions. Because relatively little is known about such interactions, we assigned a standard positive value to the free energy of the single stranded regions they create upon their formation.

33. Tetraloops

Some hairpin loops of 4 nucleotides have unusually low destabilizing energy due to a favourable sequence in a loop. Known tetraloops can be described as motifs GNRA or UA(U)CG, where N defines arbitrary nucleotide and R is purine. The program takes these motifs into account by separate energy contribution.

34. Nonpaired terminal nucleotides

Terminal nucleotides are dangling ends adjacent to a stem. They have small favourable stacking contribution and are taken into account in multibranched loops. If some nucleotide can stack on either of two helices, the better value is taken.

3 3	3 3
C. G	C. G
A. U	A. U
C. G	C. G
5' ÄÄÄÄ ÄÄÄÄ3'	5' ÄÄÄÜ ÄÄÄÄ3'
5' nonpaired	3' nonpaired

13. APPENDIX D: FILE FORMAT DOCUMENTATION

You might have noticed that the *STAR*-editor lacks a myriad number of functions and options that other editors offer. The *STAR*-editor was incorporated in this package mainly to have a simple editor available if you decide to make a quick adjustment in your sequence, structure or energyrules. Probably you find it more convenient to use your favourite editor for large-scale modifications. This is easily done, because all files are plain ASCII files.

In the sequel we inform you about the layout of those data files.

13.1 STAR FILES

In the sequel we inform you about the layout of those data files in order to use other editors instead of the primitive *STAR* editor.

We will discuss:

- sequence files
- structure files
- energyrules files

In the examples we terminate each line with an [enter]-key “[CR]”; don’t type this in.

13.2 SEQUENCE FILE

There are three rules to consider when editing/creating a sequence with your own editor:

1. **Nucleotides.** First, you may only use the characters "A", "C", "U", "G" (and the lowercase alternatives) and " " (a blank). All other characters will be removed -without warning- when opening (option [Primary/Open...]) your sequence. The lowercase characters "a", "c" etcetera can be used as NON-PAIRING nucleotides (see Chapter 7.6).

2. **Comment.** Second, there is an option to include comments within your sequence file (see also paragraph 7.6.4). An asterisk marks the beginning of a comment in a line. Everything after the asterisk in that line will be discarded by *STAR*.
3. **Blanks.** Empty lines and blanks can be used to separate functional groups in your sequence, or to make your sequence easier to check. For example (remember, the [CR] is the [Enter]-key):

GCGG	* 5' end of the 3' noncoding [CR]
	* region of the [CR]
	* tobacco mosaic virus (TMV) [CR]
GUCAA	* bulge loop [CR]
AUGUAUA [CR]	
UGguuCA	* hairpin loop [CR]
UAUAUCAUCCGC [CR]	

or:

*	4	7	11	14	17	20	23	[CR]
GCG	GGU	CAA	AUG	UAU	AUG	guu	CAU	[CR]
AUA	UCA	UCC	GC	[CR]				

or:

GCGGGUCAA AUGUAUAUGguuCAUAUAUCAUCCGC [CR]

13.3 STRUCTURE FILE

The structure file has the following layout:

```

STRUCTURE
stem 1[CR] (line 1)
stem 2 * a comment[CR] (line 2)
.
.
stem n[CR] (line n-1)

```

Each line specifies the positions (and optionally the energy's) of a stem. Thus, one line is made up of:

1. the base number of the most 5' nucleotide of the stem
2. the base number of the 3' end of the 5' stem-half
3. the base number of the 5' end of the 3' stem-half
4. the base number of the most 3' nucleotide of the stem

For example:

```

STRUCTURE
27 32 38 43 *RNA structure Fo virus[CR]
49 53 61 65 [CR]
1 7 66 72 [CR]
11 13 22 24 [CR]

```

When *STAR* writes a structure File, all your comments will be lost. So keep a copy of your original structure file.

You don't need to type in a regular layout. Just have 4 numbers on each line.

13.4 THE ENERGYRULE FILES

If you are not completely satisfied with the default energyrules, you might consider changing them.

The format of all the energyrules must obey some format-rules. They are listed here:

1. Bulge loop energyrules:
Line 1 is text: BULGERULES.
The table must always consist of 2 columns.
2. Deep groove energyrules:
Line 1 is text: DEEPGROOVERULES.
You can make the table as large as you want, but all the columns and rows must be completely filled.
3. Formula energyrules:
Line 1 is text: FORMULA.
The following lines must specify the following sets of numbers: 3, 3 and 1 numbers.
4. Hairpinloop energyrules:
Line 1 is text: HAIRPINLOOPRULES.
The table must always consist of pairs of values.
The first value of each value-pair is the length of the hairpinloop, and these values must be in increasing order.
5. Internalloop energyrules:
Line 1 is text: INTERNALLOOPRULES.
The table must always consist of 5 columns.
6. Mismatch energyrules:
Line 1 is text: MISMATCHRULES.
The table must contain 64 rows of 4 values.
7. Nonpaired terminal 3' dangling nucleotides:
Line 1 is text: NO3PAIRINGRULES.
The table must contain 16 rows of 4 values.
8. Nonpaired terminal 5' dangling nucleotides:
Line 1 is text: NO5PAIRINGRULES.
The table must contain 16 rows of 4 values.

9. Quasi-junction energyrules:

Line 1 begins with text: QJUNCTIONRULES.

The table must always consist of 2 columns.

10. Shallow groove energyrules:

Line 1 is text: SHALLOWGROOVERULES.

You can make the table as large as you want, but all the columns and rows must be completely filled.

11. Stacking energyrules:

Line 1 is text: STACKINGRULES.

The table must always consist of 16 rows and 16 columns.

Adding/deleting rows/columns

In many tables you can add rows (or columns).

If you do so, be sure to add as many values as the other rows (columns) contain (including the first column which is often not an energyvalue, but a type of entry specification).

For example

0	1	2	
0	+	---	---
1		999	999
2		999	15 999
Erroneous			

0	1	2	3
0	+	---	---
1		999	999 999
2		999	15 999
Correct			

The tables are displayed in a regular layout. If you change values, or add rows or columns you are free to use as many blanks as you want ("free format"). As long as the number of values match with the other rows (and columns), *STAR* does not care about your layout.

For example:

0	1	2	3
0	+	---	---
1		999	999 999
2		999	15 999
Correct			

Comment

Whenever you feel the need, you can add comment by separating the values left and the comment with an asterisk.

For example:

```
0 40 This is erroneous
1 30 * This is correct comment
```

Similar to the structure file, if you write an energyrules file, it will be transformed to the format *STAR* uses, without comments. So keep a copy of your original energyrules file safe if you prefer to have the comments included.

After you saved your ASCII-file, you can start *STAR*, activate the option [/Open...] in one of the appropriate menu's, and load your file. That's all there is to it!

C O N T E N T S

[illegible]

1. ABOUT THIS MANUAL	7
1.1 THE ORGANIZATION OF THIS MANUAL	7
1.2 NAMING OF KEYS	7
1.3	8
1.3 CONVENTIONS USED	8
2. THE THEORY UNDERLYING STAR	9
2.1 INTRODUCTION	9
2.2 FOLDING OF RNA	9
2.3	10
2.3 LOOP ENERGIES	10
2.4 TERTIARY INTERACTIONS	10
2.5 THE ALGORITHMS	11
2.5.1 Greedy Folding	11
2.5.2 Stochastic Folding	12
2.5.3 Genetic Algorithm	13
2.5.4	15
2.5.4 Pros And Cons	15
3. THE PROCEDURE OF STAR	17
3.1 STEP 1: GET YOURSELF A SEQUENCE	17
3.2 STEP 2: GET YOURSELF A STRUCTURE	17
3.3 STEP 3: STUDY RESULT AFTER CALCULATION OF THE STRUCTURE	18
4. THE STAR MENU	19
4.1 THE [DESK] MENU	19
4.2 THE [FILES] MENU	20
4.3 THE [PRIMARY] MENU	21
4.4 THE [SECONDARY] MENU	22
4.5 THE [ENERGYRULES] MENU	23
4.6 SUMMARIZING: THE STRUCTURE OF THE [PRIMARY], [SECONDARY] AND [ENERGYRULES] MENUS	24
4.7 THE [_.] MENU	24
5. THE [DESK] MENU	25
5.1 [/ABOUT STAR]	25
6. THE [FILES] MENU	27
6.1 [/HELP]	27
6.2 [/RENAME...]	27
6.3	28
6.3 [DELETE...]	28
6.4 [/QUIT...]	28
7. THE [PRIMARY] MENU	29
7.1 [/HELP]	29
7.2 [/OPEN...]	29
7.2.1 Lower case characters	30
7.2.2 Thymine	30
7.2.3 Incorrect characters	30
7.2.4 Old and new sequence	30
7.2.5 Save sequence	31
7.3 [/SAVE AS...]	31
7.4	32
7.4 [/SAVE]	32

7.5	[/VIEW].....	32
7.6	33
7.6	[/EDIT].....	33
7.6.1	Add Nucleotides	33
7.6.2	Comments.....	33
7.6.3	Delete Nucleotides	34
7.6.4	Move & Copy Nucleotides	34
7.6.5	Leaving The Editor.....	34
7.7	35
7.7	[/PRINT...]	35
7.7.1	Print to file.....	35
8.	THE [SECONDARY] MENU	37
8.1	[/HELP].....	37
8.2	[/OPEN...].....	37
8.3	39
8.3	[/SAVE AS...]	39
8.4	40
8.4	[/SAVE].....	40
8.5	[/FACTORS...].....	41
8.5.2	42
8.5.2	[/Factors...]: General Factors	42
8.5.3	[/Factors...]: Greedy Folding Factors ...	43
8.5.4	45
8.5.4	[/Factors...]: Stochastic Folding Factors	45
8.5.5	[/Factors...]: Genetic Algorithm Factors	46
8.6	48
8.6	[/CALCULATE]	48
8.6.1	Before You Start Calculation.....	48
8.6.2	Starting The Calculation.....	49
8.6.3	Preliminary Generation Of All Possible Stems 50	
8.6.4	51
8.6.4	Greedy Folding Calculation.....	51
8.6.5	Stochastic Folding Calculation.....	51
8.6.6	52
8.6.6	Genetic Algorithm Calculation.....	52
8.6.7	53
8.6.7	Final Show Of Predicted Structure.....	53
8.7	[/VIEW...].....	53
8.7.1	54
8.7.1	View Structure As Table.....	54
8.7.2	View Structure By Spannings.....	55
8.7.3	56
8.7.3	View Structure As Mountain.....	56
8.7.4	View Structure As Lego.....	57
8.7.5	58
8.7.5	View Structure As Bracket.....	58
8.7.6	View Structure As Connect.....	58
8.8	59
8.8	[/EDIT].....	59
8.8.1	Add A Stem.....	59
8.8.2	Type comment.....	59
8.8.3	Delete Stems	59
8.8.4	Leave The Editor.....	59
8.9	62
8.9	[/PRINT...].....	62

8.9.1	<i>Print to file</i>	62
9.	<i>THE [ENERGYRULES] MENU</i>	65
9.1	[HELP]	65
9.2	[OPEN...]	65
9.3	[SAVE AS...]	68
9.4	70
9.4	[SAVE]	70
9.5	71
9.5	[VIEW...]	71
9.5.1	72
9.5.1	[/View...]: <i>Bulge Loop</i>	72
9.5.2	[/View...]: <i>Deep Groove</i>	72
9.5.3	73
9.5.3	[/View...]: <i>Formula Constants</i>	73
9.5.4	[/View...]: <i>Hairpin Loop</i>	73
9.5.5	74
9.5.5	[/View...]: <i>Internal Loop</i>	74
9.5.6	[/View...]: <i>Mismatches</i>	74
9.5.7	0 [/View...]: <i>Nonpaired Terminals</i>	75
9.5.8	[/View...]: <i>Quasi-junction</i>	76
9.5.9	77
9.5.9	[/View...]: <i>Shallow Groove</i>	77
9.5.10	[/View...]: <i>Stacking Energyrules</i>	77
9.6	[EDIT...]	78
9.6.1	<i>Edit Bulge loop Energyrules</i>	79
9.6.2	<i>Edit Deep groove Energyrules</i>	79
9.6.3	<i>Edit Formula Constants</i>	79
9.6.4	<i>Edit Hairpin loop Energyrules</i>	79
9.6.5	<i>Edit Internal loop Energyrules</i>	79
9.6.6	<i>Edit Mismatch Energyrules</i>	80
9.6.7	<i>Edit Nonpaired Terminal Nucleotides</i> ..	80
9.6.8	<i>Edit Quasi-junction Energyrules</i>	80
9.6.9	<i>Edit Shallow groove Energyrules</i>	80
9.6.10	<i>Edit Stacking Energyrules</i>	80
9.6.11	81
9.6.11	<i>Possible errors</i>	81
9.7	84
9.7	[/PRINT...]	84
10.	<i>Appendix a: STAR procedure</i>	89
11.	<i>Appendix b: ERRORMESSAGES</i>	91
12.	<i>appendix c: Glossary</i>	97
13.	<i>Appendix d: FILE FORMAT</i>	
DOCUMENTATION	105
13.1	STAR FILES	105
13.2	SEQUENCE FILE	105
13.3	106
13.3	STRUCTURE FILE	106
13.4	THE ENERGYRULE FILES	106

(18-nov-00) Sometimes your name and address differ from that in the order form. For example because you are the final user, but the financial department ordered the program, or because you moved in the mean time. If your address is different, please fill out this registration form and return it to us. This will enable us to notify you of updates of *STAR*.

Name: _____

Company: _____

Address: _____ City: _____

Country: _____

STAR licence _____ for Atari/PC/Macintosh (please cross out)

_____ fold here _____

Don't hesitate to write us for making suggestions, reporting problems or for comments.

Comment: _____

Stamp
here

F.H.D. van Batenburg
Institute for Theoretical Biology
University Leiden
Kaiserstraat 63
2311 GA Leiden
the Netherlands

----- fold here -----