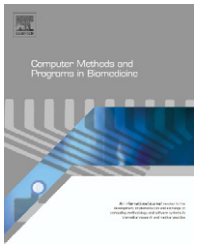
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 90 (2008) 89–94

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



www.rnaworkbench.com: A new program for analyzing RNA interference

Radka Svobodová Vařeková[a], *Ivan Bradáč*[a], *Martin Plchút*[a], *Michal Škrdla*[d], *Michael Wacenovsky*[b], *Helmuth Mahr*[b], *Georg Mayer*[b], *Herbert Tanner*[b], *Hermann Brugger*[b], *Josef Withalm*[b], *Peter Lederer*[b], *Heinrich Huber*[b], *Gerhard Gierlinger*[a,b], *Ronald Graf*[b], *Hakim Tafer*[c], *Ivo Hofacker*[c], *Peter Schuster*[c], *Martin Polčík*[a,*]

[a] *ANF DATA (Siemens), Heršpická 5, 639 00 Brno, Czech Republic*
[b] *Siemens, Gudrunstrasse 11, 1101 Vienna, Austria*
[c] *Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstrasse 17, 1090 Vienna, Austria*
[d] *Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic*

ARTICLE INFO

Article history:
Received 5 June 2007
Received in revised form
30 October 2007
Accepted 1 December 2007

Keywords:
RNA interference
siRNA design
mRNA secondary structures
siRNA sequence rules

ABSTRACT

RNA interference (RNAi) has become an important tool to study and utilize gene silencing by introducing short interfering RNA (siRNA). In order to predict the most efficient siRNAs, a new software tool, RNA Workbench (RNAWB), has been designed and is freely available (after registration) on http://www.rnaworkbench.com. In addition to the standard selection rules, RNAWB includes the possibility of statistical analyses of the applied selection rules (criteria). The role of RNA secondary structures in the RNA interference process as well as the application of sequence rules are discussed to show the applicability of the software.

1. Introduction

The RNAi [1,2] is a highly regulated enzyme-mediated process in which a short interfering RNA (siRNA) suppresses the expression of a given gene. RNAi has already significantly advanced our understanding of the functions of several genes (e.g. [3]) and future applications for drug development are envisaged (e.g. [4]). The process is triggered by the transfection of a double strand RNA (dsRNA) which is cleaved by the enzyme Dicer into 21–23 nucleotide segments, the siRNAs [5]. They are subsequently incorporated into the RNA-induced silencing complexes (RISC) [6]; the active RISC includes only the siRNA antisense strand which hybridizes with the complementary target region on the messenger RNA (mRNA) and prevents thus the gene expression. The efficiencies different target regions are very different [7] and one would like to find the most efficient siRNA in a fast and straightforward way. In order to select the siRNA with the highest repression efficacy, which in a wet-lab is a costly and time-consuming process, several computational tools have been introduced (e.g. [8–10]) and recently [11–13]. They model the siRNA–mRNA hybridization based on selection criteria, either empirical or constructed from basic principles (e.g. free energy consideration).

* *Corresponding author*. Tel.: +420 543106202; fax: +420 543106226.
E-mail address: martin.polcik@siemens.com (M. Polčík).
0169-2607/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2007.12.001

The aim of this work is to present our new software tool RNA Workbench (RNAWB). It consists of two parts. The first one is the Designer which carries out the design of the most efficient siRNAs. Several thermodynamic and sequence based rules provided by most existing tools are available. Moreover, one can use the sequence design rules, which are either built in or can be inserted by means of regular experessions. Also a new feature—the so-called Target pattern criterion [14] has been built into the Designer. The other part of RNAWB is the Analyzer the goal of which is to perform statistical analyses of the applied design criteria on already existing experimental data.

The RNAWB has been applied to analyze experimental data and some new results have been achieved and published in Ref. [14]. This article serves primarily as an application note about the new software to all potential users.

As there has been some discussions in the literature about the role of the secondary structures of the mRNA (see computational details and theory, part average accessibility), the first test example of the application of this software tackles this issue. The other example is related to the sequence rule testing which is presented in more detail in our publication [14].

2. Computational details and theory

The criteria (rules) according to which the efficacy of siRNAs are rated are introduced in this chapter. The RNAWB provides all the options available in already existing RNAi computational tools and a new rule—Target Pattern has been included. These design rules can easily be chosen in the graphical user interface (GUI) and they are described in the help options.

2.1. *Energy and sequence rules*

The free energies of the duplex, the 5′ and the 3′ ends, the middle region and the free energy of siRNA breaking are used as thermodynamical criteria for the efficacy rating of a chosen siRNA. The values of the free energies are obtained by the RNAduplex program from the Vienna RNA Package (VRNAp) [15]. The user can also choose the C/G percentage criterion in which the targets are valued according to their cytosine and guanine contents [16].

The sequence strategies offer the user a choice in applying either the Tuschl rules [17] or to define their own rules using regular expressions. One example how the regular expressions can be used is shown on the three Tuschl rules:

Tuschl Rule 1 :	AA(N19)UU,	regular expression :	AA. ∗ UU
Tuschl Rule 2 :	NA(N21),	regular expression :	.A.∗
Tuschl Rule 3 :	NAR(N17)YNN, R ∈ {A, G}, Y ∈ {C, U},	regular expression :	.A[AG]. ∗ [CU]..

(1)

Here *N* is any from the four nucleotides (C/G/U/A).

A detailed description of the regular expressions is described in the help option.

Based on a systematic sequence analysis, the so-called Target Patterns criterion can be chosen in the GUI. Its design and significance is analyzed in Ref. [14].

2.2. *Average accessibility*

The role of the secondary structure of the mRNA for the interference has been discussed in the literature. While no relationship between Mfold-predicted [18] mRNA structures and the potency of the siRNA has been concluded in Ref. [19] and little importance of target secondary structures in Ref. [7], the effect of the secondary structures of the target for gene silencing has been shown in Refs. [20–24]. Here we treat the issue by introducing average accessibility and looking on its significance for the siRNA design. This accessibility is a result of mRNA folding due to hydrogen bonds leading to inaccessible regions of the mRNA strand. For value calculation of the average accessibility, the Vienna RNA Package [15] has been used.

The accessibility, which takes this effect into consideration, is implemented the following way: the Boltzmann-weighted ensemble of suboptimal secondary structures is generated. The accessibility, which is defined as the percentage of nucleotides not bonded via hydrogen bond, is calculated for every member of the ensemble. Finally, by averaging the accessibilities over the ensemble, the average accessibility is calculated for the target. The ensemble can be produced by stochastic backtracking as introduced in SFOLD [8], here we use the implementation in the RNAsubpot program from VRNAp. It should be pointed out that the average accessibility is a continuous parameter ranging from 0 to 100. This seems a bit controversial assuming rather small number of the target nucleotides, which may imply only discrete values of the average accessibility. Its value, however, is calculated from a large number of secondary structures from the ensemble and the averaging then produces the continuous distribution of values.

2.3. *Access to the experimental data*

The RNAWB part – Analyzer makes it possible to perform analyses on two public databases – CGB siRNA [25] and siRecords [26]. The databases contain experimental RNAi results of about 420 (October 2005) and 1220 (February 2006) siRNAs, respectively.

3. System description

The flow chart process diagrams of the two main parts of the RNAWB system are shown in Fig. 1a and b. The architecture of the system is presented in Fig. 1c. Tomcat has been used as the application server. From the architectural point of view, the web (application) server part of the application can be further divided into:

- *JavaServer Pages layer* (*JSP*) which provides the HTML code to be displayed in the Web pages.

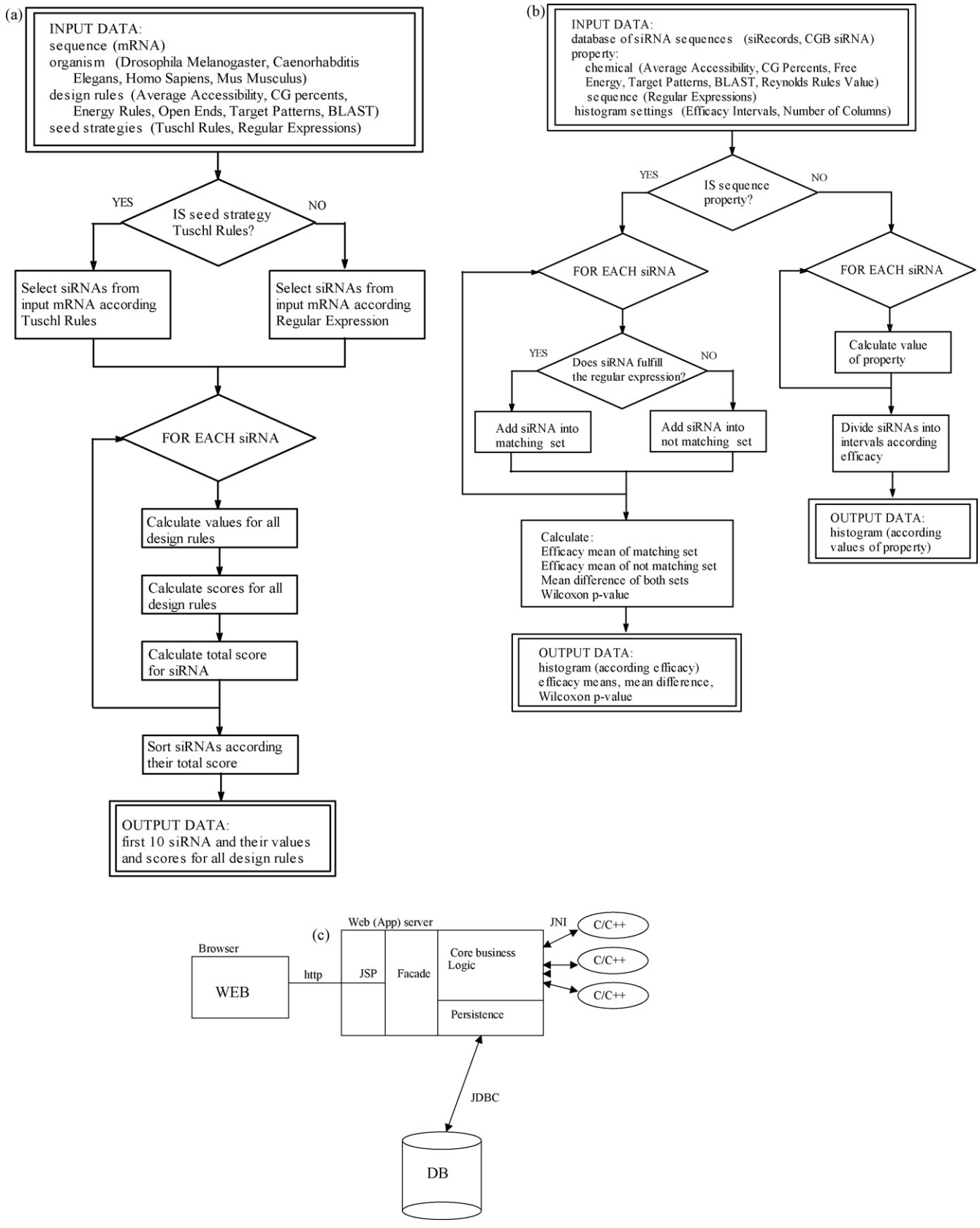


Fig. 1 – (a) The flow process diagram of the RNAWB system to calculate the siRNA efficacy. (b) The flow chart of the Analyzer part of the RNAWB. (c) The architecture of the RNA Workbench system.

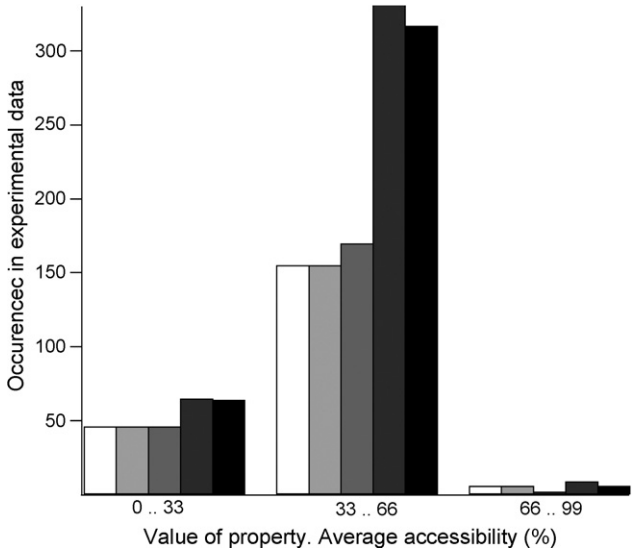


Fig. 2 – Histogram showing the distribution of experimental siRNAs from both databases together for different average accessibility and for different efficiencies—the bars correspond to efficacies 0–20%, 20–40%, 40–60%, 60–80% and 80–100% as the collor gets darker.

- *Facade*: the layer which connects the JSP and Core Logic. It is not desirable that the JSP pages access the core logic classes directly out of the following reasons: JSP pages should contain just the display logic. The Facade layer can be adjusted, e.g. for the EnterpriseJava Bean (EJB) or Struts framework in the future if needed.
- *Core logic*: implements the main algorithms, including calling of the C/C++ functions and programs and cooperation with the Persistence layer.
- *Persistence*: provides the access to the database through Java objects. This layer is further divided into the Java Persistence layer – the general Java-DB mapping framework – and the RNA WB Persistence layer – the extension of the Java Persitence layer fitted for the RNA Workbench project.

4. Results and discussion

To test the RNAWB we performed a rather thorough comparison with the published results by Gong et al. [11]. Good agreement with the presented data has been achieved. The results have been summarized in Ref. [14]. The following part concentrates on the practical application of the RNAWB and in particular on the role of the average accessibility and the sequence rules, which were in the centre of our attention.

4.1. *Average accessibility*

Fig. 2 shows the distribution of siRNAs from both databases together [25,26] as a function of the average accessibility. The picture indicates that most efficient siRNAs are located in the central part of the histogram (33–66% average accessibility) and that the accessibility may thus be a significant parameter as the efficient siRNAs in the central region clearly dominate the unefficient. To look into this effect in more detail, two statistical analyses have been carried out. The analysis of variance (one way ANOVA test) gives us the value of the statistic $F = 7.51$ (the ratio mean square between and mean square error) for the siRNAs efficiencies in the central region compared to those from the side regions, clearly rejecting the null hypothesis (i.e. one distribution for siRNAs for all average accessibilities). The probability of the F value being random is 0.0061, well bellow the significance at 1% level. The mean efficacy in the central region is 58.89% while the global mean value for both side regions (0–33% and 66–99% average accessibility) is 53.9%. Similarly high significance can be achieved by applying the Wilcoxon rank sum test [27] (called also Mann–Whitney test, Mann–Whitney U-test) which has been used to compare the medians of the investigated sets. The Wilcoxon p-value, which characterizes the statistical significance of the difference of the mean efficiencies, is 0.0148, which is well bellow the 5% significance level.

The results shown in Fig. 2 can be used as a test case, because the histogram can be produced by the user of the RNAWB. The example shows that the average accessibility is a significant parameter in the efficient siRNA design.

Table 1 – Some of the statistically significant sequence criteria for the CGB siRNA and siRecords databases

Description of criterion		CGB siRNA database		siRecords database	
Position in siRNA	Nucleotide	Mean efficacy difference	Wilcoxon p-value	Mean efficacy difference	Wilcoxon p-value
−2 (first in overhang)	A	11.01	0.001474	7.54	5.42×10^{-6}
	A/U	9.58	0.002206	7.42	1.21×10^{-5}
	No C	8.26	0.01133	4.93	0.02871
1	G	13.14	3.397×10^{-4}	6.10	2.609×10^{-4}
	C/G	10.48	9.712×10^{-4}	7.11	3.847×10^{-5}
	No U	9.89	0.01295	5.59	0.01163
9	No U	9.96	0.01334	4.22	0.01292

Mean efficacy difference is the difference of mean efficiencies of the siRNAs to fullfil and fail the particular sequence rule given in column 2.

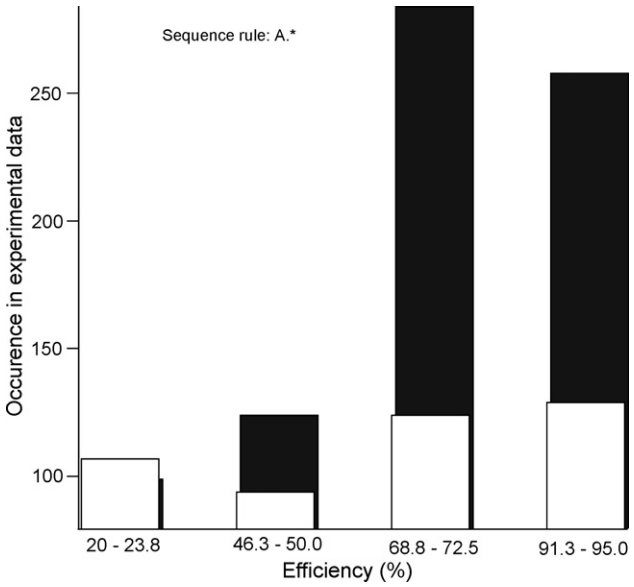


Fig. 3 – Statistical analysis of the sequence criterium A.* in the siRecord database. The dark bars show the number of siRNAs that match, the bright that do not match the criterium. The data stored in the siRecord database have discrete efficacy values.

4.2. *Sequence rules*

The next example presents the design of the sequence rules. The sequence rules can be inserted by regular expessions, see Eq. (1).

One interesting application shown here is the systematic search for new sequence rules. All single nucleotides (C/G/U/A) at all positions (−2 to 21, −2 means the first position in the overhang) in the siRNA have been tested in both databases. Table 1 shows several of the applicable rules, which we analyzed. For each rule, Table 1 contains the mean efficacy differences between the siRNAs that fulfill and fail the rule. Fig. 3 shows the result of the calculation in the first row of the table when the A.* (i.e. nucleotide A at position −2) rule has been tested. From the 1220 targets in the siRecord database a histogram in Fig. 3 has been constructed which clearly shows that for high silencing efficacy the targets that match (dark bar) dominate those that do not match the selection rule (bright bars). This fact is also reflected in the very low value of the Wilcoxon parameter (p-value) which indicates that the difference between the two sets (bright and dark) is significant.

Fig. 3 shows very clearly what can be found in Table 1, all those chosen rules are statistically significant, because Wilcoxon p-value is smaller than 0.05. This allows us to use the rules in the search for new siRNAs by means of RNAWB or other available tools. This example is only one of several more that have been presented in a more thorough analysis of the RNAi databases using the RNAi Analyzer [14].

Further development of the RANWB will concentrate on making the access to more experimental data and possibly tackle issues like off-target effects and/or look in more detail on the secondary structure role.

5. Mode of availibility

The system can be accessed on the web address http://www.rnaworkbench.com. It requires registration.

Acknowledgements

The financial support from the Austrian Zentrum für Innovation und Technologie, Wiener Innovationsförderungsprogramm für Betriebliche Forschung & Entwicklung, for the RNA Workbench project, is acknowledged.

REFERENCES

[1] A. Fire, S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, C.C. Mello, Potent and specific genetic interference by doublestranded RNA in Caenorhabditis elegans, Nature 391 (1998) 806–811.
[2] P.A. Sharp, RNA interference—2001, Genes Dev. 15 (2001) 485–490.
[3] R. Barstead, Genome-wide RNAi, Curr. Opin. Chem. Biol. 5 (2001) 63–66.
[4] M. Zacharias, Perspectives of drug design that targets RNA, Curr. Med. Chem.-Anti-Infect. Agents 2 (2003) 161–172.
[5] P.D. Zamore, T. Tuschl, P.A. Sharp, D.P. Bartel, RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, Cell 101 (2000) 25–33.
[6] E. Bernstein, A.A. Caudy, S.M. Hammond, G.J. Hannon, Role for a bidentate ribonuclease in the initiation step of RNA interference, Nature 409 (2001) 363–366.
[7] A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, A. Khvorova, Rational siRNA design for RNA interference, Nat. Biotechnol. 22 (2004) 326–330.
[8] Y. Ding, C.Y. Chan, C.E. Lawrence, Sfold web server for statistical folding and rational design of nucleic acids, Nucl. Acids Res. 32 (2004) W135–W141.
[9] Y. Naito, T. Yamada, K. Ui-Tei, S. Morishita, K. Saigo, siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference, Nucl. Acids Res. 32 (2004) W124–W129 (Web Server issue).
[10] B. Yuan, R. Latek, M. Hossbach, T. Tuschl, F. Lewitter, siRNA Selection Server: an automated siRNA oligonucleotide prediction server, Nucl. Acids Res. 32 (2004) W130–W134.
[11] W. Gong, Y. Ren, Q. Xu, Y. Wang, D. Lin, H. Zhou, T. Li, Integrated siRNA design based on surveying of features associated with high RNAi effectiveness, BMC Bioinformatics 7 (2006) 516.
[12] S. Iyer, K. Deutsch, X. Yan, B. Lin, Batch RNAi selector: a standalone program to predict specific siRNA candidates in batches with enhanced sensitivity, Comput. Methods Prog. Biomed. 85 (2007) 203–209.
[13] O. Matveeva, Y. Nechipurenko, L. Rossi, B. Moore, P. Saetrom, A.Y. Ogurtsov, J.F. Atkins, S.A. Shabalina, Comparison of approaches for rational siRNA design leading to a new efficient and transparent method, Nucl. Acids Res. 35 (2007) e63.
[14] I. Bradac, R. Svobodova Varekova, M. Wacenovsky, M. Skrdla, M. Plchut, M. Polcik, siRNA selection criteria—statistical analyses of applicability and significance, Biochem. Biophys. Res. Commun. 359 (2007) 83–87.

[15] I. Hofacker, W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, P. Schuster, Fast folding and comparison of RNA secondary structures, Monatshefte Chem. 125 (1994) 167–188.
[16] S. Takasaki, S. Kotani, A. Konagaya, An effective method for selecting siRNA target sequences in mammalian cells, Cell Cycle 3 (6) (2004) 790–795.
[17] S.M. Elbashir, W. Landeckel, T. Tuschl, RNA interference is mediated by 21- and 22-nucleotide RNAs, Genes Dev. 15 (2001) 188–200.
[18] M. Zuker, D.H. Mathews, D.H. Turner, Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, in: J. Barciszewski, B.F.C. Clark (Eds.), RNA Biochemistry and Biotechnology, Kluwer Academic Publishers, Dodrecht, The Netherlands, 1999, pp. 11–43.
[19] T. Holen, M. Amarzguioui, M.T. Wiiger, E. Babaie, H. Prydz, Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor, Nucl. Acids Res. 30 (2002) 1757–1766.
[20] T.A. Vickers, S. Koo, C.F. Bennett, S.T. Crooke, N.M. Dean, B.F. Baker, Efficient reduction of target RNAs by small interfering RNA and Rnase H-dependent antisense agents, J. Biol. Chem. 278 (2003) 7108–7118.
[21] Y. Xu, H.-Y. Zhang, D. Thormeyer, O. Larsson, Q. Du, J. Elmen, C. Wahlestedt, Z. Liang, Effective small interfering RNAs and phosphorothiolate antisense DNAs have different preferences for target sites in the luciferase mRNAs, Biochem. Biophys. Res Commun. 306 (2003) 712–717.
[22] S. Schubert, A. Grunweller, V.A. Erdmann, J. Kurreck, Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions, J. Mol. Biol. 348 (2005) 883–893.
[23] B.S.E. Heale, H.S. Soifer, C. Bowers, J.J. Rossi, siRNA target site secondary structure predictions using local stable substructures, Nucl. Acids Res. 33 (2005) e30.
[24] E.A. Bohula, A.J. Salisbury, M. Sohail, M.P. Playford, J. Riedemann, E.M. Southern, V.M. Macaulay, The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript, J. Biol. Chem. 278 (2003) 15991–15997.
[25] A.M. Chalk, R.E. Warfinge, P. Georgii-Hemming, E.L.L. Sonnhammer, siRNAdb: a database of siRNA sequences, Nucl. Acids Res. 33 (2005) (Database Issue), D131–D134.
[26] Y. Ren, W. Gong, Q. Xu, X. Zheng, D. Lin, Y. Wang, T. Li, siRecords: an extensive database of mammalian siRNAs with efficacy ratings, Bioinformatics 22 (2006) 1027–1028.
[27] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945) 80–83.