Title:

Web based computation for designing double stranded sequences to induce RNA interference

Authors:

Andrew S. Peek*, Justin J. Garretson, Yuan Lin, Jeffrey A. Manthey and Yihe Wu

Contact information:

Bioinformatics Department
Integrated DNA Technologies
1710 Commercial Park
Coralville, IA 52241

*To whom correspondence should be addressed
apeek@idtdna.com

Running Head:

Web based RNA interference design

Abstract:

Summary:

RNA interference (RNAi) is a recently observed and relatively unexplored area in molecular biology and functional genetics. The recent addition of RNAi as a tool in creating knockdowns in several diverse organisms has accentuated the possible uses for RNAi in several areas of research and therapeutics. However, the limited research history in RNAi has left the rules necessary for creating effective double stranded RNA sequences incomplete. Here we describe a computational service that allows a user to design dsRNA sequences suitable for RNAi with an extendable set of target site selection criteria.

Availability:

http://biotools.idtdna.com/rnai/

Contact:

apeek@idtdna.com

Supplementary Information:

http://biotools.idtdna.com/gateway/

Introduction:

RNA interference (RNAi) is a recently discovered natural phenomenon (Fire et al. 1998; Napoli et al. 1990) that suppresses the expression of a target gene with double stranded RNA (dsRNA). A dsRNA with sequence similarity to the target gene sequence mediates the RNAi gene silencing by several described mechanisms. There are several pathways for dsRNA to be involved in gene silencing with diverse cellular machinery, and these mechanisms may not be equally conserved among all eukaryotic organisms (Anantharaman et al. 2002). For example, the general RNAi based gene silencing by a post-transcriptional gene silencing (PTGS) via a dsRNA induced silencing complex (RISC) has been shown in worms, flies and mammals (Caplen et al. 2001; Elbashir et al. 2001a; Elbashir et al. 2001b; Fire et al. 1998), but RISC has not been identified in plants (Vaistij et al. 2002). By contrast, dsRNA induced transcriptional gene silencing (TGS) involving the target gene DNA methylation has been shown in fungi and plants (Alder et al. 2003; Schramke and Allshire 2003; Vaistij et al. 2002), but not in mammals. Furthermore, gene silencing can occur under a variety of mechanisms within a single organism (Hamilton et al. 2002).

Computationally predicting the dsRNA sequences that will effectively interact with a single mechanism or an organism's gene silencing machinery necessarily requires some knowledge of the sequence characteristics responsible for dsRNA action. Several studies of the functional requirements for synthetically produced dsRNAs and their activities have been performed (Amarzguioui et al. 2003; Braasch et al. 2003; Caplen et al. 2001; Chiu and Rana 2003; Czauderna et al. 2003; Elbashir et al. 2001a; Elbashir et al. 2001b; Hemmings-Mieszcak et al. 2003; Holen et al. 2002; Kretschmer-Kazemi Far and Sczakiel 2003; Vickers et al. 2003; Xu et al. 2003), and the characterization of naturally occurring dsRNAs and their mechanisms of action across some eukaryotic systems is also available (Kumar and Carmichael 1998; Lai et al. 2003). However, the understanding of what features make an effective dsRNA sequence is far from complete, particularly given the diversity in mechanisms of action and across eukaryotic systems. In order to facilitate molecular research into developing effective dsRNAs and understanding their mechanisms of action, we have developed a web-based application to help researchers select dsRNA sequences given a gene sequence of interest.

RNAi selection:

Given a DNA sequence of interest and a rule set of dsRNA design criteria, we have implemented a computational algorithm that will search through a DNA (or RNA) sequence, find the regions in the sequence that meet these criteria and display the resulting RNAi candidates ranked in order of their fit to the user supplied rule set. More specific to the design of dsRNA sequences that will induce gene silencing, we have limited the rules available for user modification to: length, G+C content, sequence motifs and overhanging ends. The user is able to specify a minimal and maximal as well as an optimal length for the resulting duplex sequences with a range between 18 and 26 bases. The percentage G+C content of the duplex region can also be restricted with a minimum and a maximum value.

The presence or absence of nucleotide sequence motifs can also be controlled by a sub word search and scoring algorithm. A motif (or sub word) is a sequence of DNA bases (A, C, G, T/U, M, R, W, S, Y, K, V, H, D, B, and N) that can be given one of four possible restrictions: i) must include, ii) suggested include, iii) must exclude and iv) suggested exclude. The difference between a must include and a suggested include motif is that a must include motif will be required, whereas a suggested include motif will be used in the ranking of possible dsRNA sequences depending on the weighting given to that motif, but is not absolutely required. A similar relationship exists between the must exclude and suggested exclude motifs. A second control on motifs is one of three possible location restrictions: i) distance from the 5’ end of dsRNA into the dsRNA sequence, ii) distance from 3’ end of dsRNA into the dsRNA sequence or iii) anywhere within the dsRNA sequence. The position restriction allows a fine control of preferring or avoiding individual nucleotides or motifs within the resulting dsRNA. In addition to the nucleotide sequence motif, the inclusion or exclusion restriction and position restriction, there is a weighting function available which allows the user to build a quantitative and linearly additive model of motifs, and the scoring algorithm will rank the resulting dsRNAs in order to how well they fit the model, from best fit to worst. Individual motifs can be added or removed to build a model of sequence motifs in the selection of dsRNA sequences best fitting the individual requirements of the user.

In addition to nucleotide sequence motifs in the design of dsRNA sequences, we incorporate a thermodynamic model to predict the maximal interaction between sense strand (S) and antisense strand (AS) sequences, minimizing the competing strand interactions. The thermodynamic model evaluates the energetics of the intermolecular sense strand and antisense strand interaction (S·AS), and chooses the dsRNA sequences which maximize the relative differences between the desired (S·AS) interaction, and the competing (suboptimal S·AS), (S·S), (AS·AS), (S·intramolecular) and (AS·intramolecular) interactions. The differences between the equilibrium free energies (ΔG) of formation for competing molecular species were taken as the value of

$$\delta\Delta G_{(S\cdot AS\ -\ S\cdot S)} = (\Delta G(S\cdot AS) - \ \Delta G(S\cdot S))$$

and (S·AS) pairs were scored based on their ability to simultaneously minimize the terms: $\{\delta\Delta G_{(S\cdot AS\ –\ \text{suboptimal}\ S\cdot AS)}, \delta\Delta G_{(S\cdot AS\ -\ S\cdot S)}, \delta\Delta G_{(S\cdot AS\ -\ AS\cdot AS)}, \delta\Delta G_{(S\cdot AS\ -\ S\cdot \text{intramolecular})}, \delta\Delta G_{(S\cdot AS\ -\ AS\cdot \text{intramolecular})}\}$. This model maximizes the desired (S·AS) dsRNA, while minimizing competing and potentially undesirable RNA complexes. The energetic values were calculated as previously reported (Xia et al. 1998).

Finally, the length of the 3’ overhang and the 3’ most terminal nucleotide positions of the dsRNA can be controlled in both of their nucleotide sequence and their backbone sugar chemistry. Symmetrical overhangs of between zero and three nucleotides are currently supported. Any overhanging bases can be forced to be any specific base or allowed to be the nucleotide base that was found in the target sequence. Additionally, the sugar chemistry of an overhanging base can be either ribose (RNA) or deoxyribose (DNA).

Once the target sequence, length restrictions, G+C content restrictions, motif restrictions and overhanging base parameters are configured, the user submits the rule set to the computational server with the calculate button. Evaluating the best fitting dsRNA sequences that fit the user's model is a matter of several computational steps:

i) Create all possible candidate dsRNA sequences for a particular target sequence within the length requirements
ii) Remove candidate dsRNA sequences which do not meet the G+C content restriction or must include or must exclude restrictions
iii) Evaluate the user supplied model of suggested include and suggested exclude motifs based on their weighting scheme
iv) Evaluate a thermodynamic model to maximize sense strand and antisense strand interaction and reduce intramolecular secondary and other undesirable structures between molecular species
v) Score and rank the resulting dsRNA candidates from the results of both their maximization of the motif scoring and thermodynamic models
vi) Display the *N*-best dsRNA candidates

Once these steps are complete, the results are displayed as dsRNA sequences, in rank order starting from the best fit to the rule set. The relative locations of the dsRNAs are displayed in a graphical overview of the target sequence, and more detailed information about each dsRNA candidate is given below. Additional functions can be accessed from the results display, including a detailed view of the dsRNA candidate in the target sequence and easy access to a BLAST server where sequence similarity searching can be performed to identify possible sequences that may cross-react with the dsRNA candidate.

The sequence analysis algorithms and dsRNA computational engine were written in ANSI C++ and is directed from an EJB controller written in Java running on a Linux operating system. Sequence and rule set input and display were written in VB.NET running under Microsoft Internet Information Services with the communication between Linux and Windows platforms taking place via WSDL and the Apache Axis server.

The RNA interference design server can be accessed from the URL:
http://biotools.idtdna.com/rnai/

Additional directions for this software include several areas for increasing functionality. First, a dsRNA candidate screen to allow the user to remove dsRNAs based on sequence similarity to known genomes or transcriptomes would reduce possible side effects from unintended cross-hybridization (Jackson et al. 2003; Semizarov et al. 2003). In addition to overall sequence similarity, some regions of the dsRNA sequence are more important in determining interactions between the dsRNA and the target mRNA (Chiu and Rana 2003), and incorporating these factors into the design model may improve the resulting dsRNAs. Regions within the target sequence of interest may also have some effect on the effectiveness of dsRNAs (Bohula EA 2003; Xu et al. 2003) and adding the ability to discriminate target regions may also improve dsDNA design. Supporting additional dsRNA molecule production methods is also of interest (Donze and Picard 2002; Sohail et al. 2003). Development of more robust and statistically validated models of RNAi action would improve the rules used to select dsRNA sequences, but this is presently limited by the relative lack of empirical data to derive a valid model. Finally, expanding the throughput capacity for projects involving large number of knockdown targets, and the need to design dsRNA candidates to each target, would enhance the software.

References

Alder, M.N., Dames, S., Gaudet, J. and Mango, S.E. (2003) Gene silencing in *Caenorhabditis elegans* by transitive RNA interference, *RNA*, **9**, 25-32.

Amarzguioui, M., Holen, T., Babaie, E. and Prydz, H. (2003) Tolerance for mutations and chemical modifications in a siRNA, *Nucleic Acids Research*, **31**, 589-595.

Anantharaman, V., Koonin, E.V. and Aravind, L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism, *Nucleic Acids Research*, **30**, 1427-1464.

Bohula EA, S.A., Sohail M, Playford MP, Riedemann J, Southern EM, Macaulay VM (2003) 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. Chemistry*, **278**, 15991-15997.

Braasch, D.A., Jensen, S., Liu, Y., Kaur, K., Arar, K., White, M.A. and Corey, D.R. (2003) RNA interference in mammalian cells by chemically-modified RNA, *Biochemistry*, **42**, 7967-7975.

Caplen, N.J., Parrish, S., Imani, F., Fire, A. and Morgan, R.A. (2001) Specific inhibition of gene expression by small double-stranded RNAs in vertebrate and invertebrate systems, *Proceedings of the National Academy of Sciences*, **98**, 9742-9747.

Chiu, Y.-L. and Rana, T.M. (2003) siRNA function in RNAi: a chemical modification analysis, *RNA*, **9**, 1034-1048.

Czauderna, F., Fechtner, M., Dames, S., Aygun, H., Klippel, A., Pronk, G., Giese, K. and Kaufmann, J. (2003) Structural variations and stabilizing modifications of synthetic siRNAs in mammalian cells, *Nucleic Acids Research*, **31**, 2705-2716.

Donze, O. and Picard, D. (2002) RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase, *Nucleic Acids Research*, **30**, e46.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschul, T. (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature*, **411**, 494-498.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. and Tuschul, T. (2001b) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate, *EMBO Journal*, **20**, 6877-6888.

Fagard, M., Boutet, S., Morel, J.-B., Bellini, C. and Vaucheret, H. (2000) AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi and RNA interference in animals, *Proceedings of the National Academy of Sciences*, **97**, 11650-11654.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature*, **39**, 806-811.

Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing, *EMBO Journal*, **21**, 4671-4679.

Hemmings-Mieszcak, M., Dorn, G., Natt, F.J., Hall, J. and Wishart, W.L. (2003) Independent combinatorial effect of antisense oligonucleotides and RNAi-mediated specific inhibition of the recombinant rat P2X3 receptor, *Nucleic Acids Research*, **31**, 2117-2126.

Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E. and Prydz, H. (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor, *Nucleic Acids Research*, **30**, 1757-1766.

Jackson, A.L., Bartz, S.R., Schelter, J., Koayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G. and Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi, *Nature Biotechnology*, **21**, 635-637.

Kretschmer-Kazemi Far, R. and Sczakiel, G. (2003) The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides, *Nucleic Acids Research*, **31**, 4417-4424.

Kumar, M. and Carmichael, G.G. (1998) Antisense RNA: Fuction and fate of duplex RNA in cells of higher eukaryotes, *Microbiology and Molecular Biology Reviews*, **62**, 1415-1434.

Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. (2003) Computational identification of *Drosophila* microRNA genes, *Genome Biology*, **4**, R42.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a chimeric Chalcone Synthase gene into Petunia results in reversible co-suppression of homologous gene *in trans*, *The Plant Cell*, **2**, 279-289.

Schramke, V. and Allshire, R. (2003) Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing, *Science*, **301**, 1069-1074.

Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D.N. and Fesik, S.W. (2003) Specificity of short interfering RNA determined through gene expression signatures, *Proceedings of the National Academy of Sciences*, **100**, 6347-6352.

Sohail, M., Doran, G., Riedemann, J., Macaulay, V. and Southern, E.M. (2003) A simple and cost-effective method for producing small interfering RNAs with high efficacy, *Nucleic Acids Research*, **31**, e38.

Vaistij, F.E., Jones, L. and Baulcombe, D.C. (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase, *The Plant Cell*, **14**, 857-867.

Vickers, T.A., Koo, S., Bennett, C.F., Crooke, S.T., Dean, N.M. and Baker, B.F. (2003) Efficient Reduction of Target RNAs by small interfering RNA and RNase H-dependent antisense agents, *Journal of Biological Chemistry*, **278**, 7108-7118.

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

Xu, Y., Zhang, H.-Y., Thormeyer, D., Larsson, O., Du, Q., Elmen, J., Wahlestedt, C. and Liang, Z. (2003) Effective small interfering RNAs and phophorothioate antisense DNAs have different preferences for target sites in the luciferase mRNAs, *Biochemical and Biophysical Research Communications*, **306**, 712-717.