Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 128 (2003) 115–118
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Short communication

RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*

Seth Redmond [a], Jamuna Vadivelu [a,b], Mark C. Field [a,*]

[a] *Wellcome Trust Laboratories for Molecular Parasitology, Department of Biological Sciences, Centre for Molecular Microbiology and Infection, Imperial College London, Exhibition Road, London SW7 2AY, UK*
[b] *Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia*

Received 20 October 2002; received in revised form 27 January 2003; accepted 28 January 2003

Keywords: Trypanosoma; RNA interference; Functional genomics; Suppression

RNA interference (RNAi), suppression or gene silencing has rapidly gained predominance as a comparatively straightforward method to achieve functional reduction in expression of selected gene products in many organisms, including those with diploid genomes. In many cases, RNAi may be phenotypically equivalent to a true knockout, but the continued presence of the gene means that the method is distinct. Following the pioneering use of RNAi as an investigative tool in *Caenorhabditis elegans*, the technique has been successfully applied to organisms ranging from kinetoplastida, fungi, green plants, planaria, dipterans, teleosts to most recently mammals, including humans [1–4]. Furthermore, the presence of RNAi across the plant, fungal and animal kingdoms, with relatively little variation in mechanism, suggests that the range of organisms suitable for RNAi may extend even further.

RNAi is based on the delivery of a double stranded RNA molecule to the target cell that is identical, at the sequence level, to the gene of interest; this dsRNA is cleaved to short fragments which base-pair with the endogenous mRNA, hence targeting this species for destruction [5]. The delivery methods are varied, but as specificity is based only upon the target DNA sequence, the possibility of cross- or co-suppression between closely related genes is a major potential cause of artefacts; these are difficult and laborious to detect by direct monitoring of nontargetted open reading frames (ORFs). Indeed, such difficulties were observed in the earliest studies of RNAi [6]. Consequently, a rigorous selection process for RNAi targets ideally includes the relatively simple but laborious process of BLAST searches against available data for the organism under study, in order to minimise the likelihood that co-suppression will occur. Despite the fact that numerous tools for similar tasks are available in the public domain, none have been adapted to serve this growing need.

The design of RNAi targets is conceptually simple, but assessing potential co-suppression can be lengthy. This is relatively straightforward for single copy genes, but the presence of a highly conserved region within a gene can cause the process to become extremely repetitive, as each putative product must be separately analysed. In this paper we describe a software solution, RNAit, to facilitate accurate design of RNAi constructs for trypanosomes. In this system, RNAi is achieved by expression of sequences corresponding to the ORF of interest from a head to head dual promoter-containing plasmid [7].

The algorithms required are available in the public domain. Hence, we used existing packages for primer design and alignment, Primer3 [8], and NCBI BLAST [9], with proprietary code for data entry, output and assessment of the target's suitability added. To make the application as widely available as possible, a web-based interface has been developed (Fig. 1). From an input DNA sequence, Primer3 generates a series of suitable primer pairs, along with details of their composition and melting temperature. RNAit then parses the first predicted PCR product, and BLASTs against available sequence data for the organism. The BLAST results are parsed using bioperl libraries [10] and the results assessed for suitability (Fig. 2). Importantly, the BLAST results are provided to the user, and are summarised by the program into 'same' and 'conflicting' (as discussed below).

Abbreviations: BLAST, basic local alignment sequence tool; CGI, common gateway interface; EST, expressed sequence tag; PCR, polymerase chain reaction; ORF, open reading frame; RNAi, RNA interference; VSG, variant surface glycoprotein

* Corresponding author. Tel.: +44-(020)-7594-5277.
E-mail address: mfield@ic.ac.uk (M.C. Field).

0166-6851/03/$ – see front matter
doi:10.1016/S0166-6851(03)00045-8

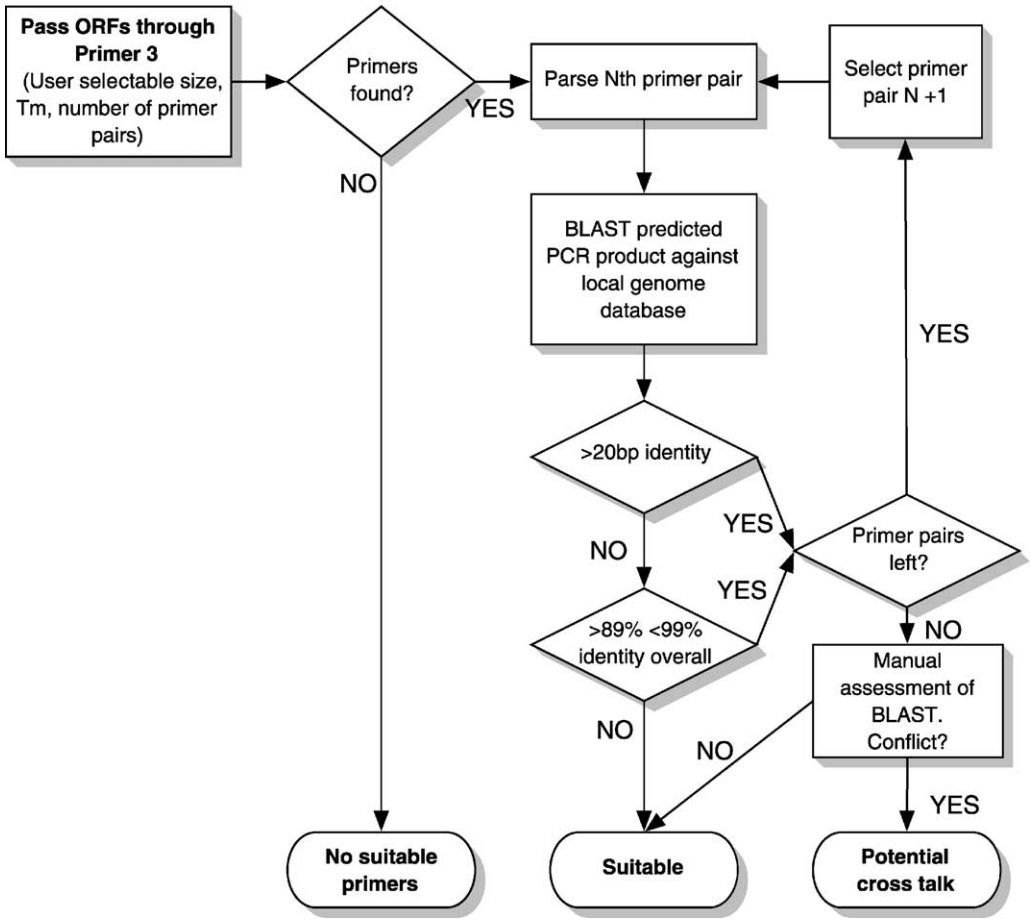
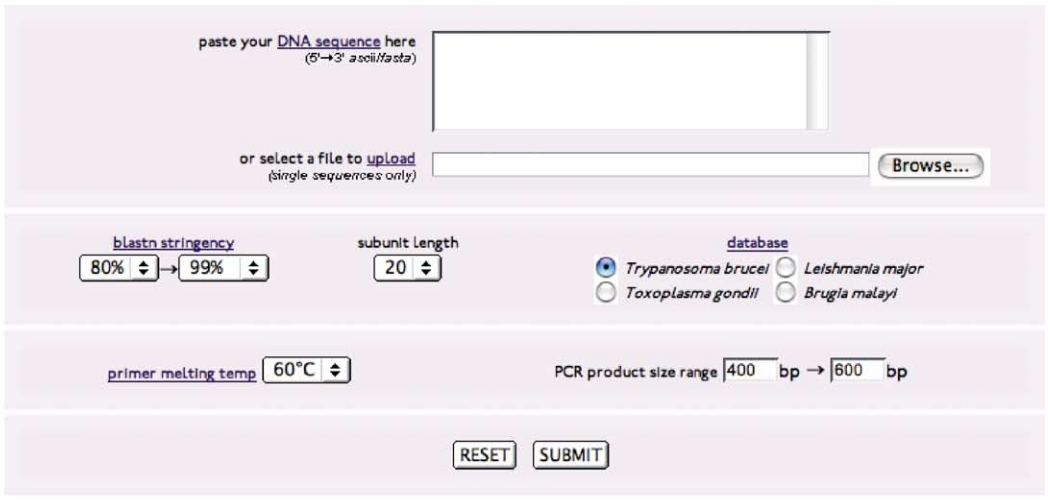


Fig. 1. Flow diagram to illustrate the computational processes within RNAit. Raw nucleotide sequence is passed to the script and then through Primer3. If no suitable primers for PCR are found the script returns the message 'no suitable primers' and exits. If primers are found which conform to specified criteria (length, melting temperature), the program enters a loop in which each primer pair is parsed; the theoretical product sequence is generated and BLASTed against a local DNA database (in this example *T. brucei* genomic and EST data). A further loop is initiated in which details of the BLAST hits are parsed; the sequences are assessed on the basis of homology, such that those which are identical, or near-identical to the sequence will be categorised as 'same'; those which are below a lower limit of homology will be 'suitable' and those which are between these two parameters, or have a continuous identical sequence of $\geq$20 bp are designated 'conflicting'. Results are summarised and printed to the screen. If no 'conflicting' reads are found the script breaks out of the primer loop and returns 'suitable for RNAi'. If the sequence is rejected the program picks the next primer pair and re-enters the blast loop. Assessments are carried out until no pairs are left, or the program finds a product without any conflicting reads. Should no suitable sequences be found, the script prints details of the sequence with the least conflicts and exits 'conflicting reads'. Manual analysis of RNAit results indicates that the default is a strict parameter set, and hence the likelihood of an erroneous assignment of 'suitable' is very low; many potential conflicts are in fact 'suitable', and the provision of the BLAST results makes assessment of conflicts rapid and accurate.

Should this first product be assessed as unsuitable for RNAi, the script moves to the next primer pair, and repeats the process until either a suitable target is found or there are no primer pairs left to analyse. Primer3 controls the suitability for PCR, and depending upon parameters selected by the user, the size of the PCR product; the second filter, suitability for RNAi, is controlled by the BLAST software together with the parser and assessment script. Primer3 also allows the user to define primer length, melting temperature and the predicted size of the product, as well as retaining the facility to ensure the selection (or avoidance) of a particular segment of DNA.

Based on prior work, co-supression can be observed when two sequences are >89% identical, or when there is a perfect match of >20 base pairs [11]. Hence, under default parameters, RNAit searches for potentially conflicting sequences at this level of similarity. Due to extensive redundancy within the database, an upper limit of 99% identity is also set, to parse BLAST results for 'self' hits. It is highly likely that 99% identity corresponds to either the same locus, or an additional locus that is functionally identical. Hence, a BLAST result returned with identity >89%, but <99% is scored as a potential conflict. Any result that is returned without falling within this range are scored as 'suitable'. A possible caveat here is that the detection of a conflict depends on the completedness of the database; with *T. brucei* this remains at less than 100% coverage and many additional parasitic genomes are far less complete. An additional complexity is the generation of false-negatives, i.e. ORFs scored as conflicting, when in fact they are suitable. Using the default parameters,

(A)



(B)

Fig. 2. Graphical user interface (GUI) for RNAit. (A) Sequence input page. Note the data entry field and the user accessible parameters for stringency, Tm, and product size. (B) data output page. The user specified parameters are returned, together with the predicted primer sequences and conditions for PCR. The predicted sequence to be used is highlighted in red, and this sequence is taken for BLAST. Results from the BLAST analysis are summarised in the same (left) and conflict (right) categories, and the entire BLAST output may be viewed by clicking on 'view blast report'. Screen shots are shown for each page. For the output page, the data returned after entering the ORF for TbRAB11 is shown.

Table 1
Results from a selection of trypanosome ORFs analysed by RNAit

Family	ORF	Predicted[a]	Obtained[b]		RNAit result[c]
			Self	Conflict	
Cyclin	cyclin 1	+	12	2	+
	cyclin 2	+	5	2	+
	cyclin 3	+	7	5	+
Procyclin	EP (ProA)	−	25	72	+
PFR	pfr a-b	−	12	21	+
	pfr c-d	−	44	18	+
Tubulin	α/β	+	63	8	+
	γ	+	1	8	+
	δ	+	7	0	+
	ϵ	+	5	0	+
	ζ	+	6	0	+
VSG	GUTat10.1	−	4	2	−
Clathrin	TbCLH	+	3	2	+
Rab	TbRAB5A	−	5	4	+
	TbRAB5B	−	1	3	+
	TbRAB11	−	7	2	+
	TbRAB11B	+	5	0	+
	TbRAB18	−	7	5	+

Complete ORFs, retrieved from the NCBI or Sanger Institute databases, were analysed for suitability for RNAi experiments using the RNAit programme. Results are classified as the number of BLAST hits returned by the programme, and are subdivided into those that obtained a "self" hit and those that returned unacceptably high identity against distinct sequences ("conflict").

[a] Expected outcome based on prior biological knowledge (i.e. known gene copy number and similarity from previous work) and direct BLAST searches. +: suitable, −: unsuitable.

[b] Outcome returned by RNAit. Self: matches that by manual inspection were clear examples of a self hit, but were not detected by the programme due to short reads, SNPs, low frequency sequence errors etcetera. Conflict: BLAST hit on a distinct sequence within the rejection threshold.

[c] RNAit result. +: suitable, −: unsuitable. The suitable category includes those ORFs annotated as 'suitable' by the automatic procedure and those included by inspection of the RNAit BLAST report. In most cases, the conflicts can be rejected on the grounds of a very short sequence only has been recognised by BLAST. Details of these analyses can be viewed at http://155.198.48.48:16080/testSeqs/.

many ORFs do apparently fail, but provision of the BLAST results facilitates rapid assessment of suitability.

RNAit was tested against the *T. brucei* genome, using a selection of ORFs including those expected to exhibit co-suppression, members of multigene families, and ORFs representing single copy genes. The results are shown in Table 1. Note that in most cases a conflicting hit was obtained, but with the sole exception of a VSG gene, manual inspection of the data indicated that the match was insignificant. Frequent short, near perfect matches were obtained, but as these were interrupted by a mismatch, these were scored as suitable. In addition, over one hundred predicted primer pairs have been successfully used to generate PCR products from ORFs on *T. brucei* chromosome I. Hence, RNAit provides a simplified method for RNAi target selection. The current implementation is for *T. brucei* only, but workers interested in other organisms are welcome to contact the authors for inclusion.

Acknowledgements

We thank Phillip Bastin (Muséum National d'Histoire Naturelle, Paris) for discussion of unpublished data on RNAi stringency. This work was supported by the Wellcome Trust (programme grant awarded to MCF, Keith Gull, Keith Matthews and David Horn) and forms part of the *T. brucei* functional genomics initiative http://www.trypanofan.org. RNAit is available over the internet at http://www.trypanofan.org. Source code (for Mac OS 10.x) is freely available by contacting the authors. The current dataset used for the analysis was obtained from the Sanger Institute ftp server and the NCBI nonredundant database on 3 September 2002; hence users should be aware that data may not always be fully up to date.

References

[1] Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc Natl Acad Sci USA 1998;95:14687–92.
[2] Wianny F, Zernicka-Goetz M. Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol 2000;2:70–5.
[3] Waterhouse PM, Graham MW, Wang MB. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci USA 1998;95:13959–64.
[4] Cogoni C, Macino G. Post-transcriptional gene silencing across kingdoms. Curr Opin Genet Dev 2000;10:638–43.
[5] Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, et al. On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 2001;107:465–76.
[6] Timmons L, Fire A. Specific interference by ingested dsRNA. Nature 1998;395:854.
[7] LaCount DJ, Donelson JE. RNA interference in African trypanosomes. Protist 2001;152:103–11.
[8] Rozen S, Skaletsky HJ, Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa, NJ: Humana Press; 2000. p. 365–86.
[9] Altschul SF, Madden TL, Schäffer AA. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
[10] http://www.bioperl.org
[11] Kohl L, Durand-Dubief M, Bastin P. Flagellum function analysed by RNA interference. Abstract 3B, 12th Woods Hole Molecular Parasitology Meeting 2001.