BIOINFORMATICS APPLICATIONS NOTE

SECISDesign: A Server to Design SECIS-Elements within the Coding Sequence

Anke Busch, Sebastian Will and Rolf Backofen

Friedrich-Schiller-University Jena, Institute of Computer Science, Chair for Bioinformatics, Ernst-Abbe-Platz 2, 07743 Jena, Germany

ABSTRACT

Summary: SECISDesign is a server for the design of SECIS-elements and arbitrary RNA-elements within the coding sequence of an mRNA. The element has to satisfy both structure and sequence constraints. At the same time, a certain amino acid similarity to the original protein is kept. The designed sequence can be used e.g. for recombinant expression of selenoproteins in *E.coli*.

Availability: The server is available at http://www.bio.inf.uni-jena.de/Software/SECISDesign/index.html

Contact: backofen@inf.uni-jena.de

MOTIVATION

Selenoproteins contain the 21th amino acid *selenocysteine*. Since selenocysteine is encoded by the stop-codon UGA, its insertion additionally depends on a specific mRNA sequence and structure downstream the UGA (called SEC Insertion Sequence, SECIS). Selenoproteins have gained much interest, since they are of fundamental importance to human health and an essential component of several major metabolic pathways, such as antioxidant defence systems, the thyroid hormone metabolism, and the immune function (for review see Brown and Arthur (2001)). For this reason, there is an enormous interest in the catalytic properties of selenoproteins, especially since a selenoprotein has greatly enhanced enzymatic activity compared to its cysteine homologue.

Detailed biochemical investigation of selenoproteins requires the production of a sufficient amount of pure protein, for which an *E.coli*-based recombinant expression system is often used. A problem arises for eukaryotic selenoproteins, since the selenocysteine insertion mechanisms differ between *E.coli* and eukaryotes. In eukaryotes, the SECIS-element is located in the 3'UTR. In contrast, the SECIS of *E.coli* must follow the UGA immediately, i.e. it is located in the protein coding sequence.

Therefore, recombinant expression of selenoproteins is complicated and rarely successful. This results in a low amount of pure protein which complicates biochemical analyses (e.g. structure determination). Furthermore, there are merely few cases of successive heterologous expression (e.g. Bar-Noy and Moskovitz (2002)), which all required a careful, hand-crafted design of the nucleotide sequence. The design of SECIS-elements is a crucial step for the expression of selenoproteins in *E.coli*.

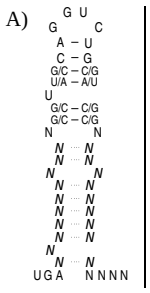
By and large, the situation is as follows. First, an eukaryotic selenoprotein cannot be expressed directly in the *E.coli* system, since it requires an appropriate SECIS-element directly after the UGA. This is not present in the eukaryotic gene. The same holds for a protein that naturally contains a cysteine which we want to replace by a selenocysteine. Second, the design of a new SECIS is likely to change the protein sequence. Therefore, one has to make a compromise between changes in the protein sequence and the efficiency of selenocysteine insertion. To ensure a high efficiency of a designed SECIS-element, we optimise its similarity to the given element and its stability. The former considers structure and sequence, the latter is assessed by the free energy of the mRNA.

THE SERVER

SECISDesign is a server for designing SECIS-elements. Our method consists of two parts: mRNA structure optimisation and an heuristic approach using inverse RNA folding, i.e. the design of RNA sequences that fold into a given structure with high probability.

The input of the algorithm devides into the structure of the SECIS-element, its nucleotide sequence, and the amino acid sequence in which we wish to insert selenocysteine. Some parts of the structure are fixed, whereas others only improve the efficiency of the element. Therefore, these features can be declared as being optional. The process of inserting a SECIS-element poses the problem of finding an mRNA sequence that contains an efficient SECIS at the right position. In addition, we also have to guarantee that the encoded amino acid sequence has a high similarity to the original one (concerning BLOSUM62 or PAM250). Often, some amino acids must not be changed in order to preserve the biological function of the protein. Therefore, the user can specify such fixed positions.

During the first step of SECISDesign, a dynamic programming approach designs an mRNA which is optimal with regard to two aspects. First, it can fold into the target structure and has maximal similarity to a given SECIS-



B)	Amino Acid Sequences (starting after pos. of Sec)	Sim. (BLOSUM62)	mRNA-Sequences and -Structures	Prob.
mouse MsrB	IFSSSLKFVPKGKE			
MsrB_BaNo	IFSTVAGLHPKGKE	35	AUAUUUAGCACGGUUGCAGGUCUGCACCCUAAAGGCAAAGAA ....((.((((....))))))..............	0.01
MsrB_mfe	IFSSLPGLVPKGTE	37	AUUUUCUCUUCGCUACCAGGUCUGGUGCCAAAAGGAACAGAA ..(((((.((.((.((((....)))))).)).))))).....	0.23
MsrB_mfe_nc	IFSSLPGLVPQGAE	33	AUCUUCUCGUCGCUACCAGGUCUGGUGCCACAAGGAGCCGAA ..(((((.((.((.((((....)))))).)).))))).....	0.75

Fig. 1. A) SECIS FdhF of *E.coli* with optional pairs (given in italic type) B) design of FdhF in MsrB: **MsrB_BaNo**: sequence given in Bar-Noy and Moskovitz (2002), **MsrB_mfe**: using mfe, FLS, **MsrB_mfe_nc**: using mfe→nc, FLS (default values for other parameters)

sequence. Second, it encodes an amino acid sequence that is maximally similar to the original one. Besides, it preserves fixed positions. Even if the nucleotide and amino acid conditions are contradictory, a solution will be found by allowing insertions and deletions on the amino acid level. For more details see Backofen and Busch (2004).

In a second step, SECISDesign tries to optimise the free energy of the target structure to increase the folding probability while the constraints from the first step are retained. To this end, an inverse RNA folding is initialised with the sequence resulting from the first step and done by one of the following local search approaches: adaptive walk (AW), full local search (FLS), and stochastic local search (SLS). They differ in the way to accept or decline a newly found solution as a replacement of the previous one. This is known as *search criterion*. During this search, a further nucleic mutation will be accepted, if it has a foldability that is allowed concerning the search criterion and keep a certain similarity on the amino acid level. One can choose from the following methods to evaluate the RNA foldability: minimising the distance of the minimum free energy structure of the designed sequence and the target structure (mfe-mode), maximising the probability of the sequence folding into the desired structure (pf-mode) (Hofacker et al. (1994)), and minimising the average number of incorrect paired nucleotides (nc-mode) (Dirks et al. (2004)). To evaluate the energy and the foldings, we use the Vienna RNA Package (Hofacker et al. (1994)). In most cases, the best results were found by using FLS and a combination of the mfe-mode and either pf- or nc-mode.

Finally, we get a sequence which folds into the SECIS-structure with high probability and guarantees a given minimum similarity to both the typical SECIS-sequence and the original amino acid sequence.

Usage. In order to obtain an mRNA sequence coding for a selenocysteine containing protein, the SECISDesign server needs the amino acid sequence and the position, where the selenocysteine should be inserted. In addition, the user can give information about positions whose allowed changes are restricted to certain amino acids. Moreover, the target element can be selected from a list of SECIS-elements of *E.coli*. Although the server is taylored for SECIS, the method is not restricted to the motifs of the list. The user can define new and even unrelated elements. Finally, one can select the search strategy, the evaluation of foldability, and further parameters which are used during the second step. To facilitate the use of SECISDesign, we set recommended values by default. Since the computation usually takes only about one minute, the user can test several parameters to find a solution that fits his requirements best.

RESULTS

Bar-Noy and Moskovitz (2002) have studied the mammalian methionine sulfoxide reductase B (MsrB). In order to express this selenoproteine in *E.coli*, they designed a SECIS by hand that gives an amino acid similarity score of 35 (BLOSUM62) but has an unpaired lower part of the stem (Fig. 1). The application of SECISDesign leads to better results concerning the quality of the SECIS-element (since the bases of the lower part of the stem are paired), the probability of folding into the target structure, and often, also concerning the similarity score (Fig. 1).

REFERENCES

Backofen, R. and A. Busch (2004). Computational design of new and recombinant selenoproteins. In *Proc. of the 15th Annual Symposium on Combinatorial Pattern Matching (CPM2004)*.

Bar-Noy, S. and J. Moskovitz (2002). Mouse methionine sulfoxide reductase B: effect of selenocysteine incorporation on its activity and expression of the seleno-containing enzyme in bacterial and mammalian cells. *Biochem Biophys Res Comm 297*(4), 956–61.

Brown, K. M. and J. R. Arthur (2001). Selenium, selenoproteins and human health: a review. *Public Health Nutr 4*(2B), 593–9.

Dirks, R. M., M. Lin, E. Winfree, and N. A. Pierce (2004). Paradigms for computational nucleic acid design. *Nucleic Acids Research 32*(4), 1392–403.

Hofacker, I. L., W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte Chemie 125*, 167–188.