Genome-wide prediction of PCR products based on termodynamic parameters

Boccia A (1), Silvestre A (2), Paolella G (1,3)

(1) CEINGE Biotecnologie Avanzate, Napoli, Italy
(2) Dep. Biochimica e Biotecnologie Mediche, Universita' 'Federico II', Napoli, Italy
(3) Dep. SAVA, Universita' del
Molise, Campobasso, Italy

Motivation

DNA amplification by polymerase chain reaction (PCR) is one of the most powerful techniques currently used in molecular biology. Nevertheless, the behaviour of the reaction is not completely predictable: non-targeted products are often amplified, particularly when complex templates, such as genomic DNA, are involved in the reaction. Mismatch tolerance is the most significant factor affecting PCR specificity, together with primer length, template size and product size limit. Computer programs are commonly used to accurately design PCR primers on the basis of a range of factors related to primer sequence and to predict the formation of non-targeted products. The specificity of primer/template interaction is tested by comparing oligo and template sequences, but the analysis is typically limited to a restricted region around the target sequence, rather than the real template, often a whole eukariotic genome. More recent programs evaluate primer/template specificity by searching for primer matches in the full genomic sequence, but they usually provide a list of potential amplimers, without attempting a quantitative analysis of amplification products. The tool presented here predicts all potential products, generated by two or more oligos on a given genome. Product yield is simulated by taking into account the stability of primer/template hybrid, calculated by using predicted thermodynamic parameters. The simulation may also be extended to cover RT-PCR by using EST DBs or predicted genes on the genome sequence.

Methods

The procedure was developed by using the PHP programming language and consists of a number of different steps. First, the target genome is searched in order to identify all similarity matches, by using BLAST (7) with modified default parameters to enhance search sensitivity. The search time is significantly reduced by always performing only one BLAST run against a single query sequence formed by concatenating all primers. Matches are then ordered on the basis of chromosome position and orientation, and only those compatible with potential amplification are retained. The stability of primer/template interaction is predicted at these potential priming sites, by calculating free energy, enthalpy, entropy and melting temperature, as described (1-6).

Results

Some commonly used melting temperature prediction methods were evaluated, and a tool to calculate primer/template stability was developed on the basis of the algorithm proposed by Allawi and Santa Lucia (1-6), with minor modifications. The tool was used as a base to build the PCR prediction program described above. The program output returns the predicted priming sites together with the termodynamic parameters. Ratherb than simply determining the expected optimum temperature, the stability of potential hybrids is calculated under the specific conditions used, including temperature, salt concentration and other parameters. The dissociation constant of the primer/template complex is used to quantify the expected yield of each amplimer, estimated as the fraction of template DNA bound to the primer given the primer concentration and the dissociation constant of the primer/template complex. A typical run is complete in 20-30s for the human genome on current hardware, independently of the number of oligos. Experimentally validated cases, where extra bands are produced, confirm that predicted results closely match the observed bands. The prediction may also be carried out for RNA amplification, either through search of EST libraries, or through analysis of transcripts expected according to exon annotation in the genome sequence. Graphical output is provided in the form of a simulated picture of a gel electrophoresis showing all the expected products and the relative predicted intensities. A web

interface is in preparation.

Contact email: boccia@ceinge.unina.it

References

1. John SantaLucia, Jr. et al., Biochemistry 1996, 35, 3555-3562
2. Allawi H.T. et al., Biochemistry,1998, 37, 9435-9444
3. Allawi H.T. et al., Biochemistry, 1998, 37, 2170-2179
4. Allawi H.T. et al., Nucleic Acids Research, 1998, 26 (11), 2694-2701
5. Allawi H.T. et al., Nucleic Acids Research, 1998, 26 (21) 4925-4934
6. Peyret N. et al., Biochemistry, 1999, 38, 3468-3477
7. Altschul, S.F. et al. 1990 J. Mol. Biol. 215:403-410