Deliverable D2.7

+-Funding Scheme: THEME [ICT-2007.8.0] [FET Open]

Paving the Way for Future Emerging DNA-based Technologies: Computer-Aided Design and Manufacturing of DNA libraries

Grant Agreement number: 265505

Project acronym: CADMAD

Deliverable number: D2.7

Deliverable name: Revision of algorithms for the design of primers and overlaps based on biochemical results

Contractual Date[1] of Delivery to the CEC: M36
Actual Date of Delivery to the CEC:
Author(s)[2]: Weizmann
Participant(s)[3]: Weizmann, ETHZ (Sterling)
Work Package: WP2
Security[4]: Pub
Nature[5]: R
Version[6]: 0.0
Total number of pages:

[1] As specified in Annex I

[2] i.e. name of the person(s) responsible for the preparation of the document

[3] Short name of partner(s) responsible for the deliverable

[4] The Technical Annex of the project provides a list of deliverables to be submitted, with the following classification level:
Pub - Public document; No restrictions on access; may be given freely to any interested party or published openly on the web, provided the author and source are mentioned and the content is not altered.
Rest - Restricted circulation list (including Commission Project Officer). This circulation list will be designated in agreement with the source project. May not be given to persons or bodies not listed.
Int - Internal circulation within project (and Commission Project Officer). The deliverable cannot be disclosed to any third party outside the project.

[5] **R (Report):** the deliverables consists in a document reporting the results of interest.
P (Prototype): the deliverable is actually consisting in a physical prototype**,** whose location and functionalities are described in the submitted document (however, the actual deliverable must be available for inspection and/or audit in the indicated place)
D (Demonstrator): the deliverable is a software program, a device or a physical set-up aimed to demonstrate a concept and described in the submitted document (however, the actual deliverable must be available for inspection and/or audit in the indicated place)
O (**Other):** the deliverable described in the submitted document can not be classified as one of the above (e.g. specification, tools, tests, etc.)

[6] Two digits separated by a dot:
The first digit is 0 for draft, 1 for project approved document, 2 or more for further revisions (e.g. in case of non acceptance by the Commission) requiring explicit approval by the project itself;
The second digit is a number indicating minor changes to the document not requiring an explicit approval by the project.

Abstract

Computational PCR primer design is a pivotal technology in biotechnology and for DNA editing in particular.
CADMAD partners have, during the past 2 years, developed a high throughput, microfluidic experimental system that gathered key experimental data that could, potentially, function as an improved foundation upon which computational PCR primer design builds its prediction of PCR primers.
During the past year we have written new algorithms for PCR primer design that integrate the insights from our experimental results from tens of thousands of PCR reactions. These have resulted in what we believe is the best to date PCR primer design tool.

Keywords[7]:

Computational PCR primer design, microfluidics

Introduction

a. Aim / Objectives

To develop state of the art computational primer design software for DNA editing and PCR in general

b. State of the Art

Multiple computational tools exist for PCR primer design, most of them free for use over the internet.

c. Innovation

Here, we present the first PCR primer design software that builds its predictions upon an experimentally validated database of tens of thousands of PCR reactions. Additionally, our softwares innovation is that it is focused on what our experiments identify as (by far) the most challenging problem in PCR primer design, namely template miss-priming, instead of giving a disproportional weight to factors that have little effect on PCR failure or are easy to compute, such as priemr dimer formation.

2. Implementation

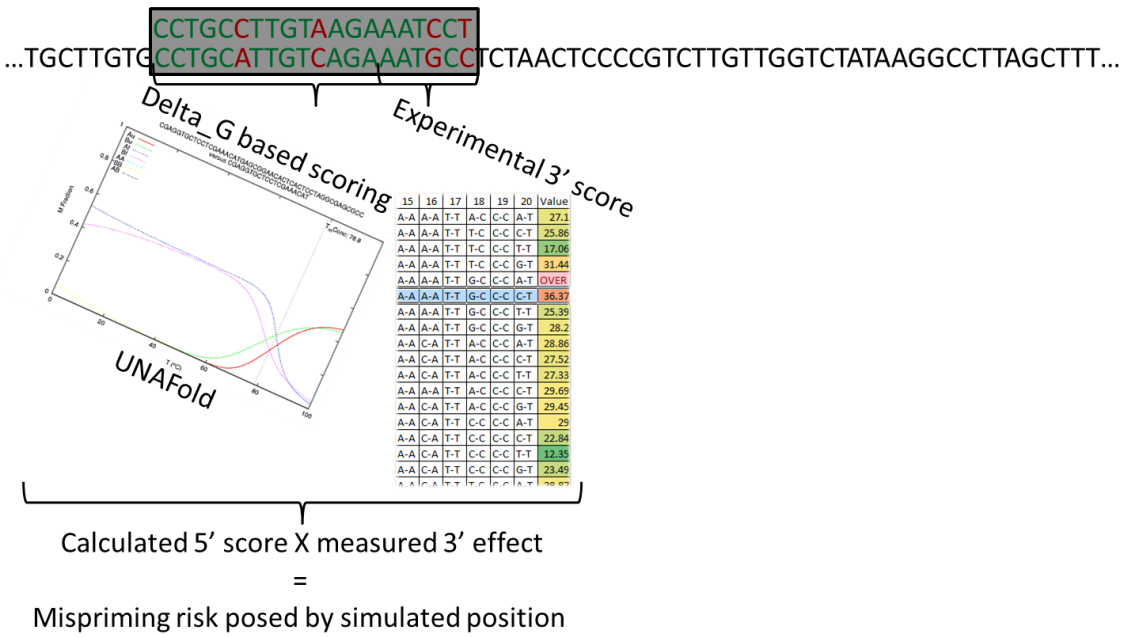
We have implemented a novel algorithm for PCR primer evaluation. It consists of 2 major elements: (1) the evaluation of the miss-priming of the 3' terminus of a primer, which we detected to have a major effect on miss-priming and (2) the effect of the remaining 5' portion of the primer, which we detected to have an important role in miss-priming, albeit less dramatic than that of the 3' end.
The evaluation of the 3' end we have microfluidically, exhaustively searched the entire sequence space of possible mutations. We can, therefore, simply lookup for the relevant microfluidic results when evaluating a potential miss-priming event at the 3' end of a new primer. When evaluating a potential miss-priming event at the 5' end of a new primer use a related, but different strategy, which involves a combination of experimental results (from our microfluidic experiments) and computational prediction of the interaction energy of the miss-priming complex. This is required since the sequence space of the entire 5' portion of a primer is impossible to cover completely (as we

[7] Keywords that would serve as search label for information retrieval

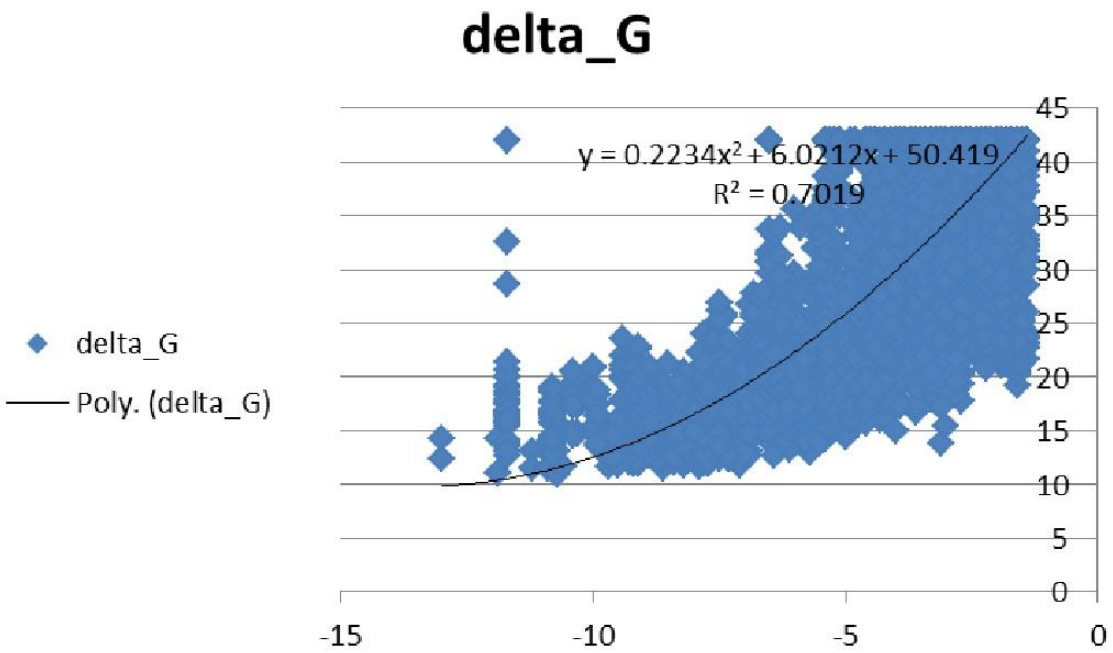
did in the terminal 6bp of the 3'), even using microfluidics. Nevertheless, we utilize the 5' microfluidic data that we do have (tens of thousands of reactions) to improve its prediction, with good results.

3. Results

Most methods to assemble DNA sequences use PCR to amplify the concatenated products and even for the assembly itself. We consider this PCR step as the most likely point of failure and therefore have concentrated our efforts on minimizing its use, predicting its outcome and preventing possible errors. PCR failures can be divided into two major causes, primer dimers and mispriming. Existing state of the art (e.g. primer3 software) was found to be too selective for primer design under the very tight restrictions posed by DNA assembly requirements. We therefore performed large-scale experiments using the latest Fluidigm microfluidics platform, designed to 1) quantify the impact of primer-dimers, and 2) quantify and investigate the basis of mispriming events. Our computational analysis concluded that primer dimers, in contrast to template miss-priming, are not a common cause for PCR failures. We therefore produced a new method for in-silico PCR validation that, given a template and primer candidate, scans the template's length for alternative priming sites that will jeopardize the proper, designed PCR process.



The figure above describes the rational of our new primer miss-priming evaluation algorithm based on microfluidic primer testing results. In short, a candidate primer is viewed as a sliding window along the entire template on which it might miss-prime. For every position, the resulting primer-template sequence is scored by a function that takes into account two scores: (1) a (microfluidic) 3'-mutation-specific, thermodynamically calculated (delta G) score for the 5' portion of the primer-template complex using every 5' combination of mutation over every combination of 3' mutations, according to the microfluidic results and (2) a microfluidic experimentally validated score for the terminal 6bp on the 3' terminus of the complex (utilizing the results from the exhaustive screening of every possible 3' terminus using microfluidics). In the first part of our new algorithm, when evaluating the 3' portion of a potential primer-template miss-priming complex our algorithm searches our experimental lookup table for the identical microfluidic PCR result and uses it as an experimentally validated prediction of its outcome.



The figure above describes the second part of our new algorithm for primer design which deals with the evaluation of the 5' component of the primer. The Ct of Each 3' mismatch group was evaluated with many different 5' mismatch groups (tens of thousands of carefully planned PCR reactions performed experimentally using microfluidics). The trend-lines above (polynomial) describe how the 5' portion delta G affects the experimentally measured Ct value. When evaluating the 5' portion of a primer with a specific 3', our algorithm can search for similar experimental results from our microfluidic experiments (above) and evaluate the primer more accurately according to experimental results.

4. Conclusions

We have completed a project that included (1) the careful planning and execution of tens of thousands of reactions in a microfluidic platform, (2) computational analysis of its results and (3) revision of our PCR primer planning software in light of the results. The work will be prepared for publication within a few months.