

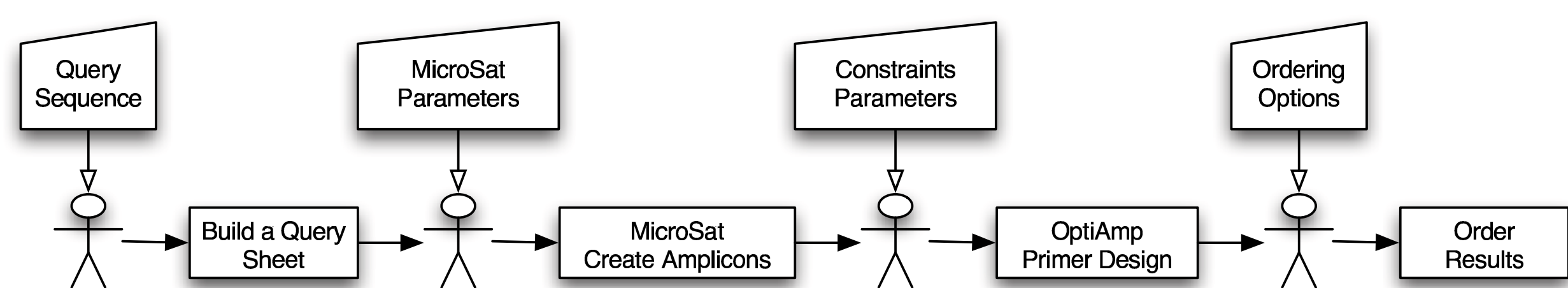
## Abstract

**OligoFactory** is a web-based set of applications for the design, on an arbitrary number of target sequences (such as ORFs), of high-quality oligonucleotides for PCR and micro-array experiments. The plugin-based architecture allows to easily embed new modules: i.e. automated amplicons determination or design methods. The user-centered web interface conforms to state-of-the-art web-standards. An XHTML form is associated with each tool to fetch parameters from the user as input. A unified presentation of results provides overviews with distribution charts and relative location bar graphs, as well as detailed features for each oligonucleotide. Input and output files conform to a common XML interchange file format to allow both automatic generation of input data, archiving, and post-processing of results. We also describe a lightweight queuing system that distributes the resource intensive calculations among multiple hosts.

The OLIGOFAKTORY groups several dedicated bioinformatic tools, embedded in a web-based framework. The dynamic and interactive web application provides consistent form-based input interface and presentation of outputs. The project is aimed at assisting researchers for a painless, rapid, automated, and reliable design.

Each oligonucleotide design application integrates a numerical approach to the minimization of undesirable interactions such as hairpins, homodimers, and heterodimers. Furthermore, each application is dedicated to a specific objective and provides great control over specific parameters and constraints.

The OLIGOFAKTORY performs the complete design process from raw query sequences data submission to ordering of designed oligos. Hence, the portal federates a collection of independent, but complementary tools that can be combined into a very large number of possible scenarios. As shown in Figure 1, users will easily produce a result sheet by using a succession of tools.



**Figure 1:** An exemple scenario of interactive primer design. Stickmans indicate user-interactive steps, namely, the edition of a web-based form. Rectangles indicate the processing of the current result sheet by a plugin module.

As shown in Figure 2, an XHTML form is associated with each tool to fetch parameters from user's input. In case of incorrect user's input, explanatory error messages will be displayed in red, in regard of the corresponding input fields. Moreover, input fields are dynamically hidden, hence only necessary parameters are displayed. Input forms are scripted with high-level building blocks that ensure validation of parameters, consistence of presentation, and automated generation of contextual help text displayed in a side bar.

For all applications, a unified presentation of result sheets (see Figure 2) provides several distribution charts to visualize relevant features of the result set. Moreover, easy-to-spot warning flags are shown in case of problems with hairpin and homodimer secondary structures and/or with specificity.

Several oligonucleotide design applications have been implemented in JAVA using our common multilevel optimization pipeline, namely:

- OPTIAMP: Design of primers for PCR amplifications around defined amplicons
- LOD: Long Oligo Design
- STABYEXPRESS: Generation of synonyms for primers
- siRNA: Design of siRNA based on A. Reynolds et al. 2004

Moreover, some helper applications are useful to order, detect amplicons, or to extract features from primer pairs:

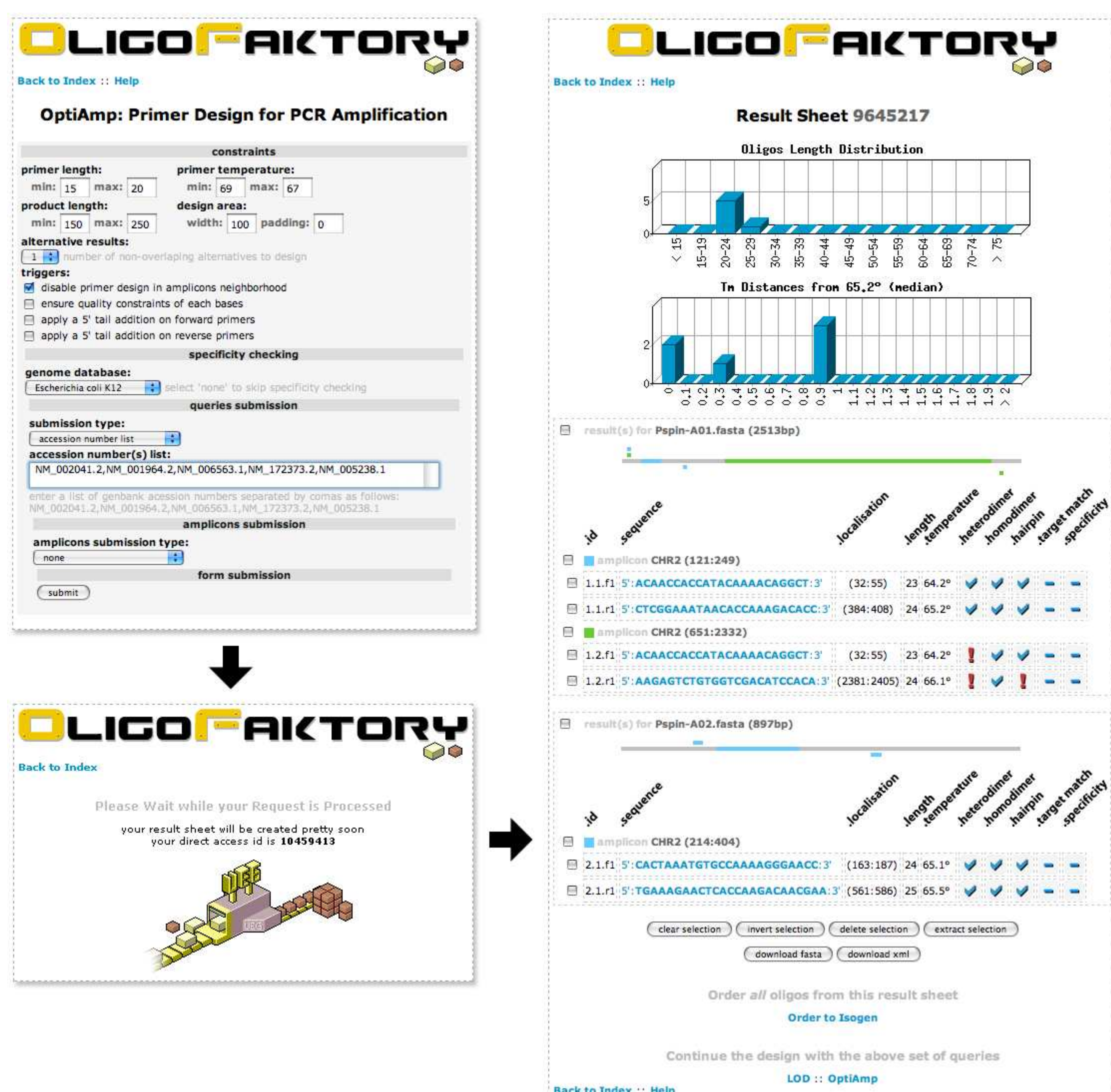
- MICROSAT: detection of (Super-)Microsatellites
- PRIMER ANALYSIS: Evaluation of primer pairs
- ORDER NOW: Ordering of results
- BUILD A QUERY SHEET: Visualization of primers and their associated amplicons

Each plugin tool reads an input parameter file and dumps results on an output file. Thus, tools work as filters, combinable through pipes. Both input and output files conform to a common XML interchange file format. Execution parameters are specific to each application and are appended via input forms.

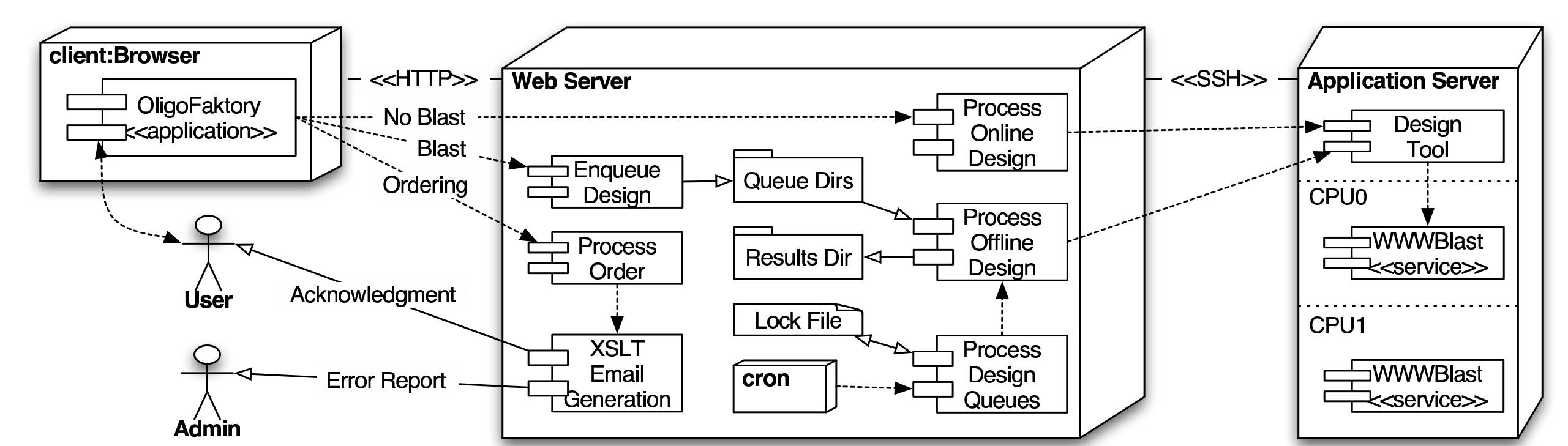
**A tiny queuing system for efficient primer design** Multiple users can submit multiple jobs through the OLIGOFAKTORY web-portal and some design jobs may require a similarity search across a nucleotide BLAST database. This IO-bound process requires to limit the number of such queries executed in parallel on the application server. We set up a lightweight queuing system (Figure 3) to ensure that, at any time, the executed jobs do not require more than the available IO resources. This queuing system has the advantages of:

- it only relies on shell scripts and the *crontab* service
- it uses e-mail notification
- it supports distant execution of jobs through SSH

A remote user submits a job to the system through the web-portal. If the design requires specificity checking, it is added to a job queue and processed by `process_design_quers.sh` and `process_offline_design.sh`. Otherwise, the job is processed online using `process_online_design.sh`.



**Figure 2:** This figure shows the *input form*, the online design waiting screen and the *output form*. The OPTIAMP module designed primer pairs around the amplicons (colored bars) of two sequences (gray bars) provided by the user



**Figure 3:** Deployment diagram of the queuing system behind the OligoFactory web-portal.  $\dashrightarrow$  represent dependencies, whereas  $\rightarrow$  indicate read and/or write processes.

The queuing system actively watches for putative job requests and, before starting the design, locks the system to limit resources usages. We have chosen to distribute the computation charge among several independent and dedicated servers. In our application, a web server is the entry-point to job submissions and manages the queues and results directories. In order to ensure interactive performances to any visitor of our web server, we wanted to reserve the resources of this server exclusively for dynamic web page generation. It will also be in charge of the mailing capabilities. Resource-consuming tasks (primer design and BLAST specificity checks) are launched on a third party server, parameters and results are stored on the web server.

The OLIGOFAKTORY web-portal is available at <http://ueg.ulb.ac.be/oligofactory>. A User's Guide explaining the available tools with their input parameters, and presenting the result sheet structure is available on the web-portal.